UMinho | 2023

Luís Miguel Ferreira Rosa **Smart Human Mobility in Smart Cities**

Universidade do Minho
Escola de Engenharia

Luís Miguel Ferreira Rosa

Smart Human Mobility in Smart Cities

October, 2023

Universidade do Minho
Escola de Engenharia

Luís Miguel Ferreira Rosa

Smart Human Mobility in Smart Cities

Doctoral Thesis

Doctorate in Informatics

Work supervised by

Cesar Analide Rodrigues

Fábio André Souto da Silva

October, 2023

Acknowledgements

Writing this acknowledgments section feels like one of the toughest parts of my entire PhD journey – a paradox, right? The journey so far has not been easy at all. It might sound weird, but these four years have been a wild mix of triumphs and setbacks, highs and lows.

I would like to express my profound gratitude to my supervisor, Professor Cesar Analide, and my co-supervisor, Professor Fábio Silva, for their unwavering support, guidance, and invaluable teachings, which laid the crucial foundations for the successful completion of my PhD. Equally significant are my heartfelt thanks to the members of the 'ISLAB family' at the ALGORITMI Center, particularly Professor Paulo Novais, Professor Dalila Durães, Pedro, António, Bruno, Dalila, Ricardo, João, Benjamin, Francisco, and many others who have been an integral part of the laboratory. Their mentorship and readiness to assist have been invaluable throughout this journey.

To my family. They've been my lifeline during those late-night study sessions after long workdays. I am deeply grateful to my father, Francisco, who encouraged me to embark on this adventure, even when I was initially scared witless. My mother, Marinha, was always there with cups of coffee and sandwiches so I wouldn't pass out in front of the computer. My brother, Francisco, proudly celebrating his brother's achievement from afar, was always ready to offer his assistance whenever needed. And to my sister-in-law Patrícia, her sister Claúdia, and her parents Mrs. Fátima and Mr. Hélder, as well as, our new family member and my godson, 'Miguelito'.

I extend my gratitude to my godparents, Uncle Alfredo and Aunt Adília, and my cousins, Ana and Anabela, for their unwavering support and kind words. Special thanks to Aunt 'São' and my cousin Carla, with whom I've shared countless Christmases, and now, look, the once rebellious kid has got a PhD. Time really does heal, doesn't it?

To those who may no longer be with us but are undoubtedly watching from above. I pay tribute to my paternal and maternal grandparents, whom I have never had the privilege to meet (except for my maternal grandfather), but who, I am confident, are observing their grandchild forging his own path with fearlessness and optimism. I remember Mr. Sérvio, my Aunt São's father-in-law, who was the life of our childhood Christmas dinners for many years. My Uncle Antonio, Mr. Sérvio's son and Aunt São's husband, may have been sparce with words but had a unique sense of humour. We've lost some great people along the way.

To my friends – ’Marquito’, ’Italiano’, ’Huguinho’, ’Limiano’, Luísa, Miguel, Vítor, Pedro, Paula, ’Bifes’, ’Chiquinho’, ’Fred’, Gabriel, and Catarina – your presence has been a cornerstone of support throughout this journey. You showed me that life isn’t all work and study; there’s time for fun and hanging out. Having a small, tight-knit crew of friends is the real deal.

I would like to extend my sincere appreciation to Johnson Controls Company. My special thanks go to Isabel Casaca, Nuno Caleira, and João Sousa (my ‘boss’ at the time), who, during my tenure, were instrumental in creating an environment within the company that allowed me to embark on this academic endeavour without imposing constraints or barriers. I am grateful to Tiago Lagarto, who, despite becoming the ’Dono Disto Tudo’ [Owner of Everything] within the company, always had some wise advice. To my Project Manager, Hugo Ribeiro, I am thankful for his patience, words of encouragement, and genuine concern.

Lastly, to Xpand IT company. Ricardo Pires (my current Project Manager) as well as to all the colleagues, directors and staff of the company, many thanks for the support!

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the Universidade do Minho.

Braga, 23 de Outubro de 2023

(Location) (Date)

Assinado por: **LUÍS MIGUEL FERREIRA ROSA**
Num. de Identificação: 13571274
Data: 2023.10.26 00:03:58+01'00'

Luís Miguel Ferreira Rosa

(Luís Miguel Ferreira Rosa)

Abstract

Smart Human Mobility in Smart Cities

In our modern world, technology is present everywhere in our lives, from our community to our workplaces, and even our homes. In fact, our decisions are increasingly influenced by technological gadgets. Governments also see them as valuable allies in improving the quality of life in cities, providing solutions for the management and sustainability of human mobility. As urban populations continue to grow, smart cities must find smarter ways to utilize existing resources.

Smart cities are generating real-time aggregate data, and the Internet of Things has contributed to a strong connection between thousands of devices. This integration allows cities to connect disparate utility, infrastructure, and public service grids. To analyse the vast amounts of data generated, Artificial Intelligence empowers decision-makers to understand their surroundings to help them make humanized decisions, and take efficient actions in order to maximize the chances of successfully accomplishing a desired task or goal.

Although current solutions offer tools to tackle mobility challenges and address urban mobility problems (traffic congestion, transport systems, parking, and environmental issues), there is a lack of a comprehensive platform that adapts the service to the diverse mobility needs of citizens. Others (public and private sectors) are investing in infrastructures—physical and digital—to support innovative mobility solutions but, despite being filled with many functionalities, many fall short in effectively leveraging the principles of *big data* and *machine learning* to improve human mobility.

To address this challenge, the main objective of this doctoral work is to propose an architecture that integrates Internet of Things, Artificial Intelligence, Cloud technology, and smart city planning to optimize people's mobility. The preliminary results of this architecture, named *WalkingStreet*, offer a set of personalized recommendations using intelligent interpretability tools for individual's daily lives, and potential risks to individuals or to the community. Therefore, this document serves as a guide for the area of human mobility in smart cities.

Keywords: Artificial Intelligence, Crowdsensing, Human Mobility, Smart Cities

Resumo

Mobilidade Humana Inteligente nas Cidades Inteligentes

Atualmente, a tecnologia está presente em toda parte, na nossa comunidade, no nosso local de trabalho ou mesmo nas nossas casas. Na verdade, as nossas decisões são cada vez mais suportadas pela tecnologia. Por outro lado, as autoridades governamentais vêem-na como aliada para melhorar a qualidade de vida dos seus cidadãos nas cidades. Esta realidade, juntamente com o aumento da população na área urbana, significa que as cidades precisam de se tornar mais inteligentes na gestão dos seus recursos.

As cidades inteligentes geram dados em tempo real e a Internet das Coisas contribui para uma forte conexão entre milhares dispositivos. Esta fusão está a ajudar as cidades a conectarem diferentes infraestruturas e serviços públicos. Para analisar os dados gerados, a Inteligência Artificial capacita os agentes a entender os ambientes ao redor para ajudá-los a aplicar decisões humanizadas e, em seguida, ações eficazes para maximizar as hipóteses de realizar com êxito a tarefa ou objetivo.

Embora as soluções atuais ofereçam ferramentas para enfrentar os desafios da mobilidade e resolver problemas de mobilidade urbana (ou seja, sistemas de transporte, estacionamento ou questões ambientais), não possuem uma plataforma que adapte o serviço à mobilidade do individuo. Outros, seja do setor público ou privado, constroem infraestruturas- físicas e digitais -para suportar soluções inovadoras, mas, apesar de estarem repletos de funcionalidades, apresentam uma visão geral da aplicabilidade dos princípios de *big data* e *aprendizagem automática* para melhorar a mobilidade humana.

Como meio para resolver estes desafios, o principal objetivo deste trabalho de doutoramento é propor uma arquitetura que combine Internet das Coisas, tecnologia Cloud, Inteligência Artificial e planeamento de cidade inteligentes para ajudar a mobilidade das pessoas. Os resultados preliminares desta arquitetura, chamada *WalkingStreet*, fornecem um conjunto de recomendações personalizadas usando ferramentas inteligentes de interpretabilidade sobre a mobilidade do indivíduo ou comunidade. Portanto, este documento apresenta-se como um guia para a área de mobilidade humana nas cidades inteligentes.

Palavras-chave: Cidades Inteligentes, Inteligência Artificial, Mobilidade Humana, Sensorização

Contents

List of Figures

List of Tables

Listings

Acronyms

AI Artificial Intelligence

AMQP Advanced Message Queuing Protocol

API Application Programming Interface

ARIMA Autoregressive Integrated Moving Average

ARIMAX Autoregressive Integrated Moving Average with Explanatory Variable

ASCII American Standard Code for Information Interchange

BaaS Backend-as-a-Service

BLSTM Bidirectional LSTM

CART Classification and Regression Trees

CCTV Closed-circuit Television

CDRs Call Data Records

CNN Convolutional Neural Network

CNN-LSTM Convolutional Neural Network-Long Short Term memory

CoAP Constrained Application Protocol

CSV Comma-Separated Values

DARPA Defense Advanced Research Projects Agency

DL Deep Learning

DM Data Mining

DNN Deep Neural Network

DT Decision Tree

EC European Commission

ESP Extended Service Providers

EU European Union

FaaS Functions as a Service

GCP Google Cloud Platform

GKE Google Kubernetes Engine

GNSS Global Navigation Satellite System

GPS Global Positioning System

gRPC Remote Procedure Call

GUI Graphical User Interface

HCI Human Computer Interaction

HTTP Hypertext Transfer Protocol

ICT Information and Communication Technology

IoT Internet of Things

ISLab Synthetic Intelligence group

IT Information Technology

JSON JavaScript Object Notation

KDD Knowledge Discovery in Databases

LIME Local Interpretable Model-Agnostic Explanation

LSTM Long Short-Term Memory

MaaS Mobility as a Service

MAE Mean Absolute Error

ML Machine Learning

MQTT Message Queuing Telemetry Transport

MSE Mean Squared Error

MTA Metropolitan Transportation Authority

MVC Model-View-Controller

MVT Model-View-Template

NN Neural Network

NYC New York City

NYPD New York Police Department

O-D Origin-Destination

ODBC Open Database Connectivity

OLE DB Object Linking and Embedding Database

PhD Doctor of Philosophy

PoC Proof of Concept

POI Points of Interest

POJO Plain Old Java Object

Pub Publish

REST Representational State Transfer

RMSE Root Mean Square Error

S2 aaS Sensing as a Service

SARIMA Seasonal Autoregressive Integrated Moving Average

SARIMAX Seasonal Autoregressive Integrated Moving Average with Exogenous factors

SC Smart City

SHAP Shapley Additive Explanations

SLSTM Stacked LSTM

SM Statistical Method

SODA Socrata Open Data Application Programming Interface

SoQL Socrata Query Language

SP Service Provider

Sub Subscribe

TCP Transmission Control Protocol

UI User Interface

URL Uniform Resource Locator

XAI Explainable Artificial Intelligence

XML Extensible Markup Language

YAML Yet Another Markup Language

Chapter 1

Introduction

The development of technology, specifically Artificial Intelligence (AI), has allowed us to design solutions for the most diverse areas. Since its appearance, its capabilities have been molded to the most different needs of humanity. An example of this is its applicability by governmental and non-governmental authorities in support of their decisions. On the other hand, human mobility is increasingly both a challenge and an opportunity for cities. This type of mobility will be a central topic in the next chapters.

In this chapter, we begin to highlight some examples of how the potential of AI, when combined with well-implemented Data Mining (DM) policies within Big Data infrastructures, and transparently supported by Explainable Artificial Intelligence (XAI) models, has long served as an ally in improving the lives of citizens in cities. Investment in these tools and adequate infrastructure make Smart City (SC).

In particular, we will present the percentage of internet users in the world who access the internet on device handles, along with a description of the gradual attractiveness of cities. The easy access to this type of infrastructure in areas with a higher concentration of people is a crucial aspect that will be explored in this work. It is, therefore, important to understand these phenomena to demonstrate that cities are no longer just places that concentrate people, but have also transformed into rich spaces that produce the most vital raw material for this work: data.

In this chapter, we utilize various definitions and characteristics of human mobility as presented by researchers and authors, who offer different perspectives on the subject. Some of them delve into the interactions between these properties that contribute to accelerating the digital transformation of cities. As cities progress towards digitization, the elements constituting them take on new roles, and human mobility is a significant phenomenon in this context.

Finally, we carefully select the most appropriate method and research strategy for studying on human mobility in a technological environment such as SC. The enhance the quality of the work, we include a comprehensive plan of activities and outline the main objectives to be developed and achieved over the next 5 years. This chapter ends with a description of the organizational structure of this document.

1.1 Background and Motivation

The world is witnessing continuous technological development, and each day brings forth revolutions in internet, mobile and machine-to-machine technologies. This revolution, coupled with the rapid expansion of cities, has posed challenges governance and authorities to make better decisions for their citizens. Historically speaking, cities have been regarded as the pinnacle of economic and sociocultural achievements in human civilization, and the location of non-primary economic activities [49]. Thus, we need to talk about a *new urban agenda* in the *age of information* and terms such as SC, Internet of Things (IoT) or AI have been widely cited in the scientific community as they illustrate the convergence of these two phenomena.

In recent years, there has been a significant surge in the volume of literature dedicated to SC and understanding their implications. For example, Alexander Pando—Forbes Council—said *"implementing smarter cities is a necessary step to prepare for future urban growth. The population will only increase in the foreseeable future, and technology provides the opportunity for cities to flourish"* [154]. However, this type of cities need the integration of various resources, such as data storage, information services, and algorithms. These resources may be distributed across different units within the same organization, other external organizations, or even sourced from the vast expanse of the Internet.

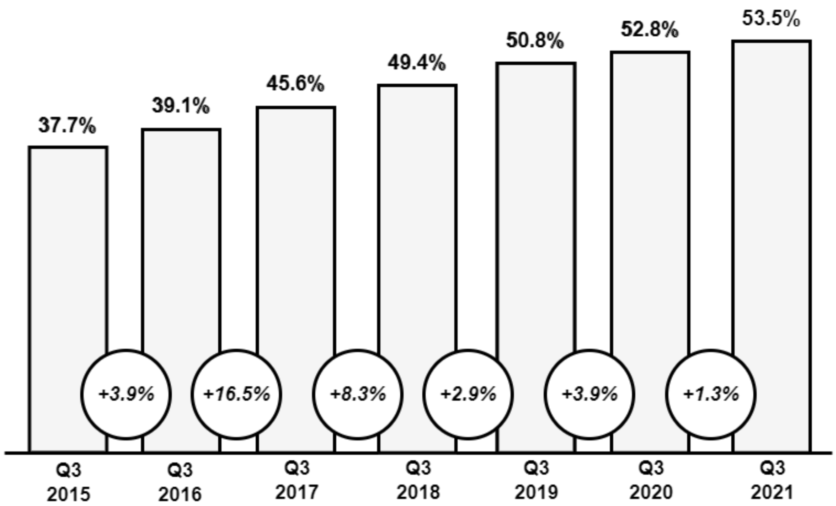

Figure 1: Daily time that internet users aged 16 to 64 spend using the internet on mobile phones as a percentage of total daily internet time (adapted from [55]).

Figure 1 provides valuable insights into the interconnection between technologies and individuals/organizations. This digital transformation empowers citizens to engage with their ecosystems, ultimately cutting costs and enhancing the human mobility within the city. Furthermore, forward-thinking governments leverage this information to fuel invention and solutions to optimize their cities. Nonetheless, the increasing volume of information generated in IoT poses significant challenges concerning its representation, storage, searchability, and interconnectedness. As a consequence, all this is going to change the way in which big data is handled.

Big data is not a new concept or idea. With recent developments in technologies such as sensors and cloud computing, along with the generation of data from various sources (sensors, humans, applications), organizations tend to store data for extended periods due to the availability of inexpensive storage and processing capabilities [157]. Once big data is stored, a number of challenges arise, especially concerning data processing and analysis. Thus, big data has become an increasingly significant challenge for the industry.

Developments in the world of AI with Machine Learning (ML) approaches have given businesses a way to enhance the analysis of big data [26]. Big data encompasses the type of data which can be supplied into the analytical system so that a ML model could "learn"to improve the accuracy of its predictions. Moreover, when combined with predictive analytics, big data enables pattern recognition, data mining and knowledge discovery.

One of the key ML techniques—and the main approach we will work with—is Deep Learning (DL). Through it, computers will learn how to think for themselves, in a way which will resemble how we, as humans, learn during our early years. Although there is ongoing debate surrounding DL and whether it is the future of AI or not, it is clear that the potential for future applications is massive, given the rapid evolution of these technologies in a relatively short period of time. In one way or another, technologies are being widely used, as they are applied to various problems across different domains. Major companies such as Microsoft, Amazon, Google and Facebook are investing millions into DL research, making it an increasingly significant part of our everyday lives. It is expected that ML will become a fundamental component of most software applications in the near future.

1.2 Human Mobility

In historical terms, cities have been regarded as the highest forms of economic and sociocultural achievements in human civilization [157]. However, this fact is challenging if we analyze the growth of the urban population. Today, approximately 359 million Europeans—72% of the total European Union (EU) population—live in cities, towns, and suburbs, with 26 cities having more than 1 million inhabitants. Additionally, only 7% of the EU population lives in cities with over 5 million inhabitants compared to 25% in the USA. Projections indicate that by 2050, more than 6.3B people—60% of the global population—will be living in cities [32]. Actually, according to the United Nations, in mid-2023 approximately 4.6 of the more than 8 billion people worldwide lived in towns or cities. This represents 57% of the global population.

Simultaneously, the gentrification[1] of cities has led the authorities to adopt laws and policies to address issues related to planned and sustainable development and efficient management of cities. One of the primary objectives is, for example, to increase the efficiency of public transportation, accessibility and overall mobility of people. Notwithstanding, Figure 2 reveals more details about the population living in urban areas, categorized by region.

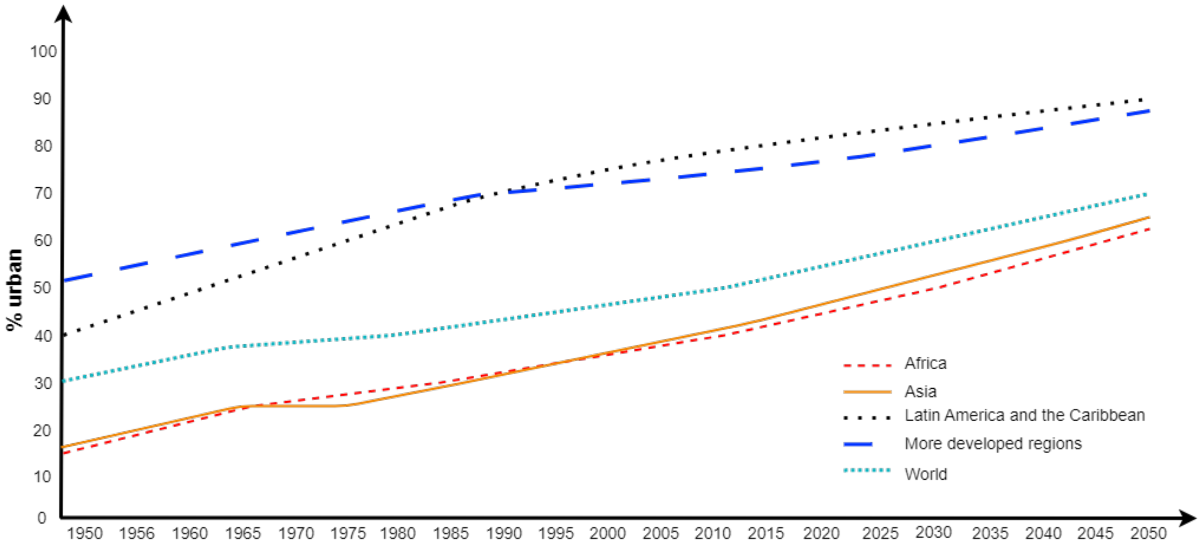

Figure 2: Percentage of population living in urban areas by region (adapted from [127]).

In the same way, a SC serves as a high-speed communication hub equipped with a strong and modern infrastructure making use of Information and Communication Technology (ICT) to establish real-time connections with other cities all over the world [58]. Quite simply, SC uses IoT devices (e.g., connected sensors, lights and meters) to collect and analyze data, and then use this data to improve environmental efficiency and an intelligent use of resources. Thus, the IoT enables the city's infrastructure to be intelligent.

The aforementioned topics (i.e., SC and IoT) pose several challenges and they are the reason why we provide a new approach to address human mobility within cities. Our approach involves selecting a human mobility prediction model and examining the role and potential of sensing devices as a crucial component of smart mobility management. To achieve this, we intend to conduct a set of distribution analyses in order to gain a deeper understanding of human mobility behaviour. By describing the overall human mobility pattern, it is possible to compare the outcomes from many locations, cities or even countries, thereby enabling the analysis of potential differences in trends related to human displacements.

With rapid growth economic came equally rapid industrialization and urbanization. In the process of urgently ensuring the city's competitiveness, a car-oriented transportation system was established. However, the automobile-centred lifestyle created too many problems, including traffic jams and environmental pollution. Walking emerged as the most basic and ecological means of transportation for humans, promoting

[1] result of wealthier people moving into a specific area

a healthier lifestyle without contributing to pollution. Moreover, walking became the most basic foundation for the development of Compact Cities and Traffic-Oriented Development, aiming to solve urban problems caused by cars.

In fact, the transition from automobile-centred cities to pedestrian-friendly cities has been getting more attention worldwide. People's outings mainly occur on roads, and they perceive the urban landscape predominantly through their sight. Visual cognition, together with several other factors like the street's physical environment, generates a response that addresses the surrounding and can eventually be experienced as satisfaction.

In order for inhabitants to feel satisfied with their outings, the landscaping plan must consider what the pedestrians see. While, previous studies have explored the relationship between pedestrians and the outing environment, there have been few studies that analyze the outing itself as an activity and its connection with the surrounding environment. The reason for this limited research is the labor-intensive nature of such a project, and the scope of analysis has been restricted to field investigations or limited research studies. Therefore, if the goal is to do the necessary research and create a truly pedestrian-friendly city, and the researchers must overcome this limitation.

1.2.1 Human Mobility Definitions

Human Mobility has been defined and analyzed from distinct points of view, drawing insights from diverse literature and methods. This diversity is partly due to the complex nature of human mobility, which has significantly evolved between the late twentieth century and the early twenty-first century. Several factors, such as the process of globalization, political changes, technological advancements, and economic shifts, have influenced and shaped human mobility patterns. These drivers co-exist with a number of conditions and enabling factors which will determine where and how individuals make theirs choices to move.

The topic of human mobility has been gaining interest among various authors, leading to different definitions influenced by external factors. In simple terms, Anqi Weang, a Postdoctoral Research Associate in the Functional Membrane and Energy Materials group at Imperial College London, defines: *"Human mobility is the movement of human beings in space and time, reflects the spatial-temporal characteristics of human behaviour"* [211]. Researchers Robert Kölbl and Martin Kozek from Austria further elaborate on this definition by stating that *"Human mobility modelling is currently described mainly through socio-economic variables, such as travel time, travel costs, income and car-ownership."*. From these examples, it becomes clear that mobility has an enormous impact on human societies, and an accurate quantitative description of human mobility is of fundamental importance to understand the processes related to human movement and their impact on the community and the environment [18].

With big data analytics, human mobility research can be used to facilitate the development of SC across multiple disciplines, such as smart traffic management, smart urban planning, smart health, smart safety, smart commerce. A framework was established to link international academic research with city-level management policies, and this framework was applied to the case of Hong Kong. Literature regarding human mobility research using big data was reviewed, which contributed to (1) discovering the spatial-temporal phenomenon, (2) identifying differences in human behaviour or spatial attributes, (3) explaining the dynamics of mobility, and (4) applying findings to city management. The research findings were further examined for their application in SC development through email inquiries to various governmental departments in Hong Kong. As data analytics continues to evolve and practical value improves, and as data from multiple sectors is utilized, opportunities to achieving smarter cities from a policymaking perspective are highlighted.

From a social perspective, the Russian-born American sociologist and political activist, Pitirim Sorokin, argues that the two societies aren't the same in terms of movement [139]. Moreover, the speed of mobility can change from one time period to the next, depending on how developed the society is. Social mobility is influenced by various factors within the society and the workforce, allowing individuals to reach new roles which offer them better living standard and greater rewards. In essence, people compete and cooperate with others in society to move up the social mobility ladder.

In technological terms, the advancement of sensors and mobile devices has enhanced the potential of individual locations, creating important geospatial data. On the side of data analytics, the practical value of data-driven research is to be discovered, whereas, city development is witnessing a constant emergence of smarter applications in multiple domains and disciplines. In summary, with big data analytics, human mobility research can be used to facilitate SC development in multiple domains, such as smart traffic management, smart urban planning, smart health services, smart safety measures, smart commerce solutions and smart human mobility initiatives.

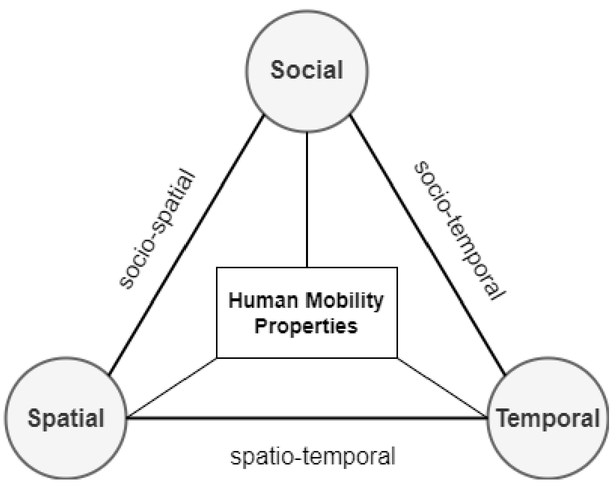

Figure 3: Properties of Human Mobility (adapted from [106]).

As previously mentioned, human mobility, the movement of human beings in space and time properties, is also significantly influenced by the social aspect. Figure 3 illustrates three ways in which these characteristics interact: socio-spatial, socio-temporal and spatio-temporal. These interactions are typically analyzed as distinct processes, through different methods and literature. This is in part due to differences in the types of interactions which are described as follows:

- **Socio-Spatial:** This interaction includes the social contexts variable, including factors like home, cluster of neighbouring homes, neighbourhoods, and cities in the spatial context, as well as social contexts such as friends, work and leisure activities. Additionally, geospatial covariates, such as the locations where individuals carry out their everyday life, are taken into consideration. It also helps reveal broad patterns of how populations move within and beyond different spaces.

- **Spatio-Temporal:** This interaction combines the spatial and temporal attributes to explain human movement patterns. Typically, it involves examining mobility patterns that include shorter, within-county trips, followed by longer domestic travel between counties and international travels corresponding to particular seasons or periods.

- **Socio-Temporal:** This interaction predicts the mobility of pedestrians by modelling two dimensions: social and time. Social dimension considers how which each individual's actions can impact others. Time dimension explores how past states of individuals may influence future states. In other words, this interaction explores how social signals and social context influence time perception, altering subjective duration and making an event seem "out of sync".

In this work, we focus on a specific set of metrics (see Section 2.3.3) that summarise the most relevant social, temporal, and spatial aspects of human movements, and we evaluate them against the realism of the goals we consider in this thesis. Specifically, we provide analytical and simulation evidence to demonstrate the capability of reproducing the accuracy of these metrics using sample mobility traces.

1.2.2 Human Mobility, City and Technology

As it was mentioned, human mobility offers a better standard of living and greater rewards. This means that human mobility has a relevant paper in the SC context. People are the heart of this type of city. The positive impact that any given technology put in place has on the local population - and the improvements it can make to their everyday life - is what really makes a city *smart*. Therefore, human mobility is the backbone of a SC.

1.2.2.1 Human Mobility: a key element of Smart City model

Human mobility plays a significant role in shaping our daily lives and our urban environment. The interconnection between various aspects of human mobility, cities, and the infrastructures they provide is crucial for understanding human movement patterns, and opportunistic content forwarding. Several studies putting human mobility as a key element of the SC models have been developed based on Figure 4. The model of SC has contributed to the acceleration of the cities' digital transformation. The integration of human mobility data into SC models has been applied to research or build new projects, considering the vast amount of data generated when we are connected day-to-day to online environments. In other words, this strong interaction between *Human-Technology-Environment* helps driving cities to become "smarter". In addition, digital channels have become more robust and improved, prompting local authorities and governments to rely on IoT infrastructures to support this progress. However, it is essential to have in mind the right balance between human mobility and fostering a deeper human connection with technology and digital channels. However, we highlight the following sub-interactions:

- **Human-Technology Interaction:** In this interaction, the human has a new role in existing and emerging technologies [178]. Human practices and social innovations are closely intertwined with these modern technologies. On the other hand, this dynamic also presents several issues and challenges considering the human role in all areas of contemporary ICT infused societies. Therefore, we are encouraged to explore the interdisciplinary nature of human-technology interaction from multiple and equally valid perspectives. This work intends to provide a platform for interdisciplinary dialogue about how humans and societies both affect and are affected by the vast array of technology available today.

- **Human-Environment Interaction:** In this element of the SC model, the technology is beyond the organizational context; user interfaces have already surpassed the desktop metaphor, and interactions have advanced from keyboard-and-pointer to touch-and-gesturing interfaces [197]. These increasing interactions between people and collections of devices result in the creation of smart environments, presenting new challenges in providing appropriate interaction affordances. The implications of increased interactivity must be discussed, as it brings about novel implications and challenges. In addition, this topic becomes a challenge that should not be overlooked in view of the other prominent issues related to user control, transparency, ethics and privacy.

- **Environment-Technology Interaction:** This interaction helps researchers accurately predict the future urban context. Based on data collected through technology infrastructures available within a city, it becomes possible to anticipate forthcoming events and situations which will impact the future urban environment [66]. The integration of environmental and human mobility observations is an important step in addressing challenges such as deploying smart spatial planning policies to mitigate the density of people in urban areas. However, this approach may not always be able to efficiently compete against other technologies deeply embedded in societies by lock-in mechanisms, e.g., learning by doing and using, scale economies, subsidies and network externalities [77]. Despite these challenges, the integration of human-environment interaction creates a context in which new technologies can develop in order to achieve desired transitions, without causing profound disruptions to the current system.

Still, Figure 4 shows that interactions between humans, the environment, and technology vary across spatial-temporal scales, with the context being a key component at the common intersection of these domains. These scales are characterized by two features: the "grain"representing the finer spatial resolution; the "resolution"referring to the granularity used in measurements; the grey square indicates the size of the total study area. In the SC context, the grain scale pertains to the level of details and precision in the SC architecture, providing more interaction details for a given architecture. On the other hand, the grey square refers to the abstract description considering a sub-interaction of the system of interactions. Variations in interaction at different scales can be used, for example, to understand the level of space and resources available for the benefit of citizens. It also allows the SC architects to understand in detail interactions and validate the associated components' interaction. From a practical standpoint, managing the complex phenomena involved in these same interactions can be achieved by decomposing a problem into a set of single-scale architectures that exchange information across the scales.

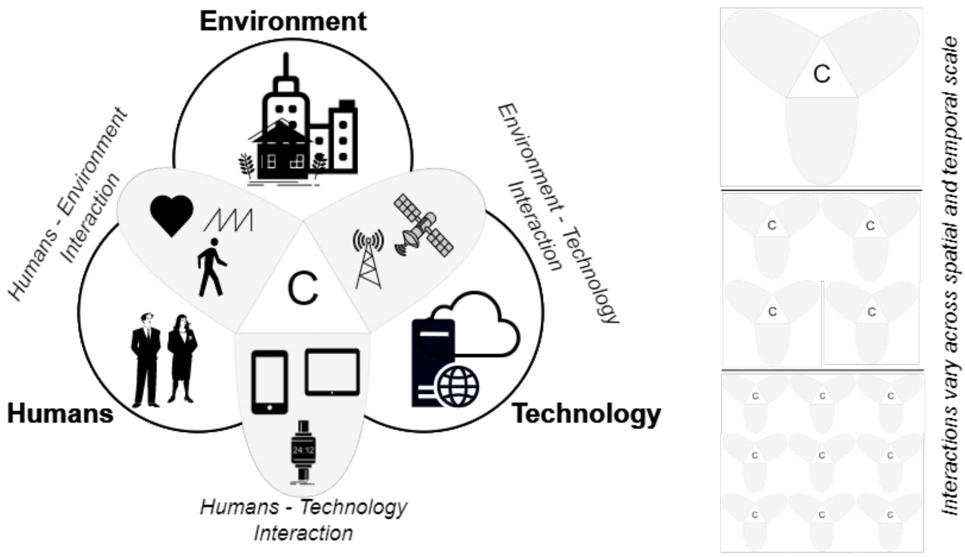

Figure 4: Model of Smart City interactions between humans, environment, and technology (adapted from [179]).

1.2.2.2 A new role of Human Mobility

With the advancements of the web of things, human mobility reflects people's travel patterns and preferences, which are especially crucial for urban applications such as urban planning and business location selection. This means that human mobility provides people with benefits as they are motivated by different factors in society and work to pursue new roles which offer them a better standard of living and greater rewards. Therefore, human mobility has assumed a new role, presenting challenges to cities on five levels.

First, the gentrification of cities. Gentrification is the process of changing the character of a neighbourhood through the influx of more affluent residents and businesses [224]. The increasing centralization of human mobility congests and further strains ageing infrastructure. However, the fact that too many people using a finite amount of space at a given time does not necessarily signifies a problem. With this thesis, we show that it can represent an opportunity.

Second, extracting and transferring knowledge from data. We investigate the problem of generating mobility data for a new target city by transferring knowledge from mobility data and multi-source data of the source cities. This knowledge can be used for prediction tasks, making it a breakthrough in solving human mobility problems within cities. Indeed, using advanced ML algorithms can extract valuable knowledge from experimental or computational data on human mobility-related problems.

Third, personal devices. They play an important role in human mobility as they connect everything physical to digital. With the adoption of personal mobile devices, cities and governments across the world partner with enterprises to create more navigable experiences, enabled by location data and location-based services for people. The real-time location data available is previously analyzed to provide people with the right mode of transportation for the right distance, time, and cost, allowing for a "smoothing"of demand.

Fourth, the scalability of technological infrastructures. Human mobility has always had a great influence on the spreading of Information Technology (IT) infrastructure ideas. However, the modelling of spreading processes faces challenges due to the categorical, temporal, and spatial characteristics of these infrastructures. In the context of categorization, according to their practical application scenarios, IoT devices can be classified into multiple categories, integrating all of them into the same service can be a challenge. Moreover, their temporal characteristics vary with the occurrence/movement of smartphones and IoT devices. For example, we find that daily diurnal patterns exist for IoT devices, similar to humans, and IoT devices move more frequently than smartphones. Still, their active periods are not the same. From the spatial characteristics of IT infrastructures, we can explore important location properties between IoT devices and human interactions.

Finally, anonymized data and privacy concerns. This process consists of the following main steps: anonymization and aggregation, extrapolation (adjusting numbers to be representative of the full population), and finally spatial and temporal aggregation. The data generated from personal mobile devices can uncover important insights into the flow of pedestrians throughout a city and privacy issues must be considered in human mobility research.

1.3 Research Hypothesis

Nowadays, society has presented the scientific community with the challenge of finding solutions that leverage technology to address human mobility in cities. Within this context, this project will explore several areas and paradigms, including SC, IoT, Human Mobility and AI, among others. With this in mind, the broad goal of this work is to expand technologies and mobility services to assist both public and private stakeholders in making well-informed decisions regarding human mobility within a SC. In the same way, another contribution of this research is to foster increased discussion on human mobility issues among key stakeholders.

Once we have completed a literature review, and before formulating a specific hypothesis, it is essential to formulate potential discussion topics. What are the requirements, limitations and challenges of current human mobility in the context of developing cities? How can ML principles applied in big data be aligned with the management of human mobility? Which DM tools can be adapted to research human mobility patterns effectively? How to solve the discrepancy between real issues and the way authorities have addressed human mobility in their cities? What is the nature of the collaboration between the multiple stakeholders (i.e., community and local authorities) and what role is played by each? How can the human mobility service be useful for various applications in modern days? Based on these reformulated questions, the specific research hypothesis can be framed as follows:

"A Mobility as a Service (MaaS) architecture that uses Cloud tools to store large amounts of data and Explainable Artificial Intelligence (XAI) techniques to explain AI models improve human mobility and capture real models for better plans of cities"

However, to be more organized and clearer, we reformulate these preliminary questions according to the objectives that we intend to achieve with the work. Therefore, we present the following Research Questions:

Research Question 1 - Can sensorization techniques such as crowdsensing and computing, taking into consideration the monitoring of a large number of individuals within a city, be facilitated from a MaaS architecture?

With the fast proliferation of handled devices, humans are not only data consumers, but also data producers with their objective or subjective sensing needs. This computing paradigm, along with an increasing number of mobile phone users, promotes participatory sensor networks. However, the integration and study of the data collected by this wide range of devices are challenging, and the application of adoption concepts and sensing techniques is fundamental to ensure that the intended result remains not distorted. This research question will be answered through example scenarios and specifications of Open-Data Application Programming Interface (API) services available for local authorities of SC.

Research Question 2 - Can we conceive and define an innovative MaaS architecture that allows gathering, addressing, and monitoring data from several open-source platforms located in SC environments?

Collecting big data from heterogeneous Open-Data API services is essential in order to efficiently process and analyze it, which is a crucial aspect of becoming our work a truly data-driven study. All service data should be integrated under one roof (also called a "data hub"). This research question can be demonstrated by developing a centralized data infrastructure that synchronizes multi-services API, collects processes, and stores data.

Research Question 3 - Can a MaaS architecture that uses ML algorithms to discover and describe human mobility patterns in urban spaces help the authorities in better city planning?

ML is a multidisciplinary field which can be proposed for modelling and predicting human mobility. In its turn, multi-source heterogeneous data provide a new driving force for exploring urban human mobility patterns from a quantitative and microscopic perspective. The studies of human mobility modelling and prediction will play a vital role in a series of applications within cities. The validation of this research question can be obtained through the applicability of DL algorithms, characterizing human mobility patterns from individual, collective and hybrid levels.

Research Question 4 - Is planning and execution of a Proof of Concept (PoC) important to understand if the proposed goals can be reached?

This PoC assesses the ability of ML regression to predict the presence of human mobility in New York City. Through the analysis of a dataset generated from free public data published by New York City agencies and other partners, we answer this research question. Therefore, deep learning is a promising method for analyzing the mobility patterns of pedestrians and urban residents, extracting and embedding patterns from complex large-scale data.

Research Question 5 - Does the participation of the community foster the development Decision-Support Services to support decision-making by the authorities that govern the SC?

To integrate different open-source data levels and convert captured data into knowledge we develop a Decision-Support for the design of smart human mobility policies in the case of emerging cities. The main goal is to improve the livability of crowded pedestrian spaces through the provision of decision support for the management of pedestrian flows. The first objective is to operate the *WalkingStreet* service in the living lab, which will demonstrate we can manage crowded pedestrian spaces. Therefore, a pilot is organized in the Department of Informatics, University of Minho (Braga) which features state-of-the-art crowd capturing, monitoring systems, and analysis techniques.

The methodology to answer these questions is *Action Research*. We aim to demonstrate through collaborative communicative processes and small scientific studies to reveal new information with respect to the reliability of the human mobility applications. This theory, along with a continuous improvements approach of practical projects, supports decision-makers in cities with valuable human mobility information. In addition, each one of these goals focuses on Computer Science and Data Science areas, but we also conducted our research taking other fields into consideration. Therefore, the research could be enriched with solutions coming from other areas, such as Psychology, Sociology, and Mobility.

1.4 Research Strategies and Methods

In this step, we outline the research procedure that encompasses various stages, starting from general assumptions to detailed methods for data collection, analysis, and interpretation. The plan will involve making several decisions, one of which involves selecting the appropriate approach to address the research problem. To aid in this decision-making process, we bring to the study procedures of inquiry (called research designs), and specific research methods for data collection, analysis, and interpretation. The selection of a research approach will be guided by the nature of the research problem or issue being investigated. For conducting rigorous research and taking informed action the approach to be used is *Action Research*.

The choice of using *Action Research* instead of the Scientific Method is based on its ability to complement the features of the work. This research approach incorporates both quantitative and qualitative information. When conducting qualitative research, the goal is to gain a deep understanding of the topic, issue, or problem from an individual perspective. This will involving an in-depth literature review on the hot topics mentioned in Chapter 2 and collecting data from the community in order to get valid findings. On the other hand, the quantitative approach involves systematically investigating phenomena by gathering quantifiable data. In this stage, we will categorize and conceptualize the information. However, we believe that there is potential to further enhance the proposed research methodology. Therefore, we will develop new prototypes and conduct experiments in a laboratory setting, via computer and simulated experiments. To increase community engagement, we will also conduct field tests in a real environment with remote monitoring. In both research methods, the PoC will serve as research-support solutions to provide insights, data, and strategies for developing a concept.

Based on the analysis of SC experiments, Action Research is also characterized as participative. For instance, in [35], the authors extend the concept of action research to participatory action research (or collaborative action research). In this work, the members of the organization under study should actively participate throughout the research process, instead of merely being subjects, and there is an intention to take action. Another distinguishing property, that differentiates it from *Research Methodology*, is its open-mindedness. In [36], the researchers adopted an action research-driven progressive and exploratory approach based on six digital rights-related factors. These properties are some of the reasons why Action Research should be chosen over a traditional research method. Additionally, during the initial context analysis, some questions arise early on, as shown in Section 1.3.

1.4.1 Proof of Concept

In the action-research approach, the PoC is considered the most effective way of meeting requirements for human mobility research. This process allows researchers to explore and experiment with innovative ideas and experiences in the development phase, ensuring that solutions are suitable for their intended purposes. Most importantly, it tests the technical feasibility and practicability of a product idea through the design concept and functionality of prototypes. Over the last decades, with the increasing adoption of mobile devices, the emergence of mobile ubiquitous technologies, and the growth of the IoT infrastructures, the study of human mobility in cities has attracted more and more attention from academic entities. Several researchers have used the features of PoC to map the huge availability of technology in cities and its interactions with human mobility. In its turn, PoC is commonly applied in multi-scale procedures, involving the extraction and analysis of crowdsensing data, converting it into meaningful information through models and algorithms, and finally sharing the insights with the community.

My research plan intends to follow up this kind of approach. The planned activities will focus on identifying the opportunities and potential of open-source data API design, as well as describing key data analysis techniques to gain an advanced understanding of the movement patterns of pedestrians. The main objective of my research is to develop a platform design that enables the execution and calculation of in a Machine Learning environment. This platform will facilitate the interaction and contribution of individuals and the community.

1.4.2 Data Collection

Data collection is also another research method to be employed during the research plan [83, 141]. Its purpose is to collect human mobility data related to the phenomenon under investigation, using multiple devices available within cities (from individual mobile phones to interactive devices accessible to the general community). Regardless of the kind of data, when collecting data, a prior evaluation should be conducted to validate which attributes are most relevant for each data collection, thereby enhancing accuracy and expanding the scope of data analysis.

Subsequently, after selecting the most important attributes, a system should be built to integrate and aggregate the generated data from multiple sources [13, 140]. The data integration process involves gathering datasets from several API services and gaining insights to determine their future business operations. Following this, the same system conducts data aggregation, wherein the primary goal is to combine these multiple datasets. Through data aggregation, we can determine the origin of the datasets and gain insights from a summarized version of the data. Therefore, the quality of data collection depends on the accuracy and thoroughness of these processes.

1.4.3 Experimental Case

The experimental case is another interesting research method with the purpose of demonstrating the impact of one single factor on another, but there is always a risk that other factors might come into play and invalidate the results of the experiment. For example, it can investigate how weather conditions influence human mobility in the city centre. Additionally, it promotes individual or community-level contributions in the experiments.

To ensure precise measurements and thorough control over the various factors influencing experiments, laboratory experiments are conducted in the field of Computer Science, where human mobility behaviour of both individuals and communities is studied in a controlled environment. Notwithstanding, this kind of environment may not fully represent the complexities of human mobility and behaviour that arise from human interactions across space over time. Moreover, it allows exploration of scenarios with intervention strategies associated with complex human movement dynamics and evaluates the effectiveness of interventions in a local context. Our experiments aim to provide a theoretical and practical contribution to the evolving role of human mobility within cities. These developments intend to prove that, through a set of open-source public services, we can understand behavioural geography, complex systems, and human spatial movements. What is proposed in this research methodology is to explore human behaviours and dynamics and how their direct and indirect participation can acquire a new role within the cities.

1.5 Research Plan

The proposed work serves as an interface between several scientific disciplines, such as SC, IoT, AI, Human Mobility and Big Data. The results of the intersection of scientific fields are interesting for numerous stakeholders, ranging from the general public to municipalities and governments. Over the course of the next three years of doctoral studies, we aim to accomplish our goals through a series of key tasks.

The primary objective of this Doctor of Philosophy (PhD) is to design an architecture (called *WalkingStreet*) that places the mobility of the individual (pedestrian) at the centre of the solution, while seamlessly integrating into a SC. However, to realize this vision, we must outline a serious of concrete steps to be undertaken throughout the PhD journey. Keeping this in mind, we present a Gantt chart outlining the project timeline consisting of seven main tasks.

A Gantt chart is a common tool used in the project management field and provides a preliminary time schedule for the completion of the doctoral degree program. The schedule indicates the plans to complete the educational component, the field and laboratory work, the data analysis, and the writing and submission of scientific articles and presentations. To summarize, our Gantt chart shows what has to be done (the activities) and when (the schedule). Therefore, we enter the data to create a chart represented in Figure 5.

Task #1 - Research of the State of the Art

The initial phase of the research will involve conducting a thorough review of existing literature in the following areas: SC, IoT, Big Data, Sensor Fusion, Cloud Computing and AI. This review will be based on the title, abstract and the introduction of articles in the Web of Science database. Additionally, the list of keywords used in the search included combinations of these areas. Based on the highly relevant contents of the abstract and citation, we have gathered multiple papers that introduce various architectures aimed ate improving the accuracy of deep learning and smart cities. The literature review process was

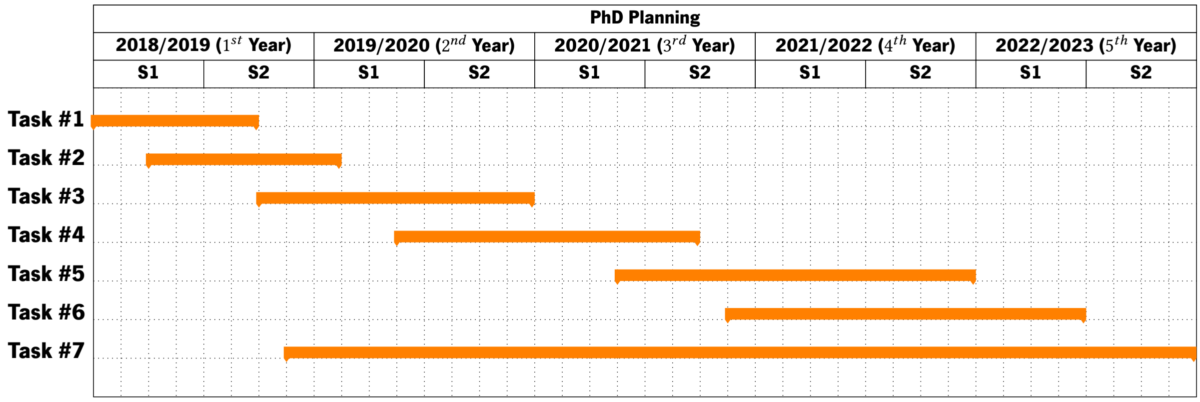

Figure 5: Schedule of activities in semesters.

conducted by consulting several websites, namely the institutional repository of the University of Minho, Elsevier Journal Finder, Google Scholar, Scimago Journal & Country Rank, Semantic Scholar, and Scopus, as well as scientific databases such as ACM Digital Library and Springer Link. Following this process, we have selected relevant works for full reading and analysis to gain a comprehensive understanding of the current state of research in these areas.

Duration: September 2018 - April 2019

Task #2 - Experiments with Tools and other Resources

Smart human mobility can be characterized by different adjectives, such as technological and interconnected, but also sustainable, comfortable, attractive and safe. Our focus will be on demonstrating how innovative IoT technologies can address mobility issues in smart cities. An extensive analysis will be performed to evaluate the performance of mobility prediction algorithms. We will assess the influence of different data features on the effectiveness of these algorithms. Citizen participation is essential not only for developing efficient livable cities, but also for sourcing valuable data from diverse contexts. Basically, the more data we collect, the better our understanding of smart cities.

Duration: February 2019 - September 2019

Task #3 - Integration of open-data multi-sources with Machine Learning algorithms

To ensure better human mobility we need to understand mobility behaviour. We want to promote more benefits in human mobility, which is a particularly demanding task in urban areas, where people frequently interact and share common spaces. By analyzing the social information database from municipal departments and studying the daily behaviour patterns of urban residents, we aim to identify the reasons that motivate citizens to visit city centers. The spatial distribution and the urban inner space structure can provide decisive support for authorities in providing social management content based on urban management. Therefore, ML and DM of human activities will play a pivotal role in our research.

Duration: June 2019 - July 2020

Task #4 - Design and computing of Human Mobility metrics

In this task, we focus on preparing our prototype for a high level of optimization and sophistication. Based on the type of involvement from users, we apply sensorization techniques which can be categorized into opportunistic crowdsensing, where data is automatically sensed, collected, and shared without user intervention and, in some cases, even without the user's explicit knowledge, and participatory crowdsensing, where users voluntarily contribute information. Regarding the last classification type, we believe that by utilizing both approaches, we can adapt our decisions and solutions to different populations within SC. It is important to recognize that mobility within a SC is not the same for every citizen (e.g., old people or people with reduced mobility). This is the main reason to track human activity: to better understand its impact on urban space. For a SC to work it needs people. The data generated by citizens provide valuable feedback that is essential to both diagnose problems and find appropriate solutions for sustainable urban development.

Duration: May 2020 - June 2021

Task #5 - Decision Support System based on Participatory Sensing

Research from Cisco has found that 60% of IoT initiatives face challenges and stall at the PoC stage, with only 26% of businesses considering their IoT initiative to be a complete success [47]. While these statistics might be concerning for some in the industry, they present an opportunity for academics and researchers. In particular, in the context of SC initiatives, which incorporate multiple stakeholders, utility providers, and citizens themselves, it is important to incorporate a pilot phase into any IoT project.

With the promise of improved mobility for the population, multiple stakeholders must work collaboratively to achieve common goals and benefits. This collaborative community approach ensures that large-scale operations are credible, well-planned, and effectively structured to deliver the desired outcomes and benefits to the citizens, such as quantifying the number of people in a particular area. The primary goal in this phase is to provide services that enhance urban living by making it safer, more affordable and more mobility-friendly. Understanding repetitive behaviour patterns is especially useful for offering real-time recommendations tailored to individual users. MaaS initiatives should be focused first and foremost on the citizen experience.

Duration: March 2021 - July 2022

Task #6 - Human Mobility Explanation System Design, Conception and Implementation

In this phase of the PhD programme, our main focus is on incorporating explainability as a means of understanding human mobility behaviours. To achieve this, we propose a set of XAI models for viewing and interpreting the outcomes of the implemented AI system. Explainability is a crucial aspect of our proposed architecture, especially when dealing with complex black box models that may be difficult to interpret in the context of human mobility phenomena. By using explainable AI models, we aim to create a proactive pipeline that can effectively identify and analyze human mobility patterns and behaviours. Basically, it is a task-specific approach to human mobility analysis. However, these algorithms are specifically designed for human mobility analysis but are also versatile enough to be applied to various AI system working with diverse human mobility datasets.

Duration: July 2021 - January 2023

Task #7 - Thesis Writing and Scientific Dissemination

In conclusion, our work plan involves several key steps: we are currently focused on writing the thesis, and presenting the results of our research through workshops and targeted presentations. We are also actively publishing articles in peer-reviewed conferences and journals. These publications not only support our research but, also disseminate our work. Additionally, as part of our research journey, we actively participate in scientific events and collaborate with fellow researchers. Ultimately, after completing the thesis and obtaining authorisation from the academic committee of the PhD programme we must deposit the thesis in an academic repository to ensure community and public access.

Duration: July 2019 - July 2023

1.6 Synthesis

The introduction section of our dissertation has defined the scope and focus of our research, providing a preliminarily overview of the area of our dissertation. We have also highlighted the background and motivation behind our research, offering insights into the main topics of interest. Moreover, our work aims to meet the expectations set forth in the introduction, and, therefore, we have formulated research hypotheses and questions which will guide our investigations throughout the study.

Through comprehensive research strategies and methods, we address these research questions. But, first of all, based on a thorough literature review, we analyze the advancements and drawbacks of each methodology. By combining different approaches, we aim to enhance the quality of our research outcomes.

After, considering these combinations, we identify and describe the PoC. In this step, we pay special attention to the kind of services associated with capturing and aggregating data. Then, we employ an experimental case to gain deeper insights to the AI models that benefit from the collected data, allowing us to validate and verify the consistency and reliability of our results concerning the XAI models we employ.

Finally, we design a plan to describe the objectives and significance of our proposed research, as well as how it will be conducted. The aim of this research plan is to employ effective strategies and methodologies to extract and analyze outcomes, including the impact that human mobility results will exert on the relevant research fields. Therefore, through a meticulous planning of activities, we antecipate discussing potential problems and alternative strategies to achieve our research objectives.

1.7 Document Structure

This document is structured into seven chapters, each serving a specific purpose. The first chapter addresses the context of the research and outlines the main objectives. The background and motivation section highlights the potential of massive device adoption and increased human mobility in an urban area as a unique opportunity to elevate cities into an unprecedented technological level. To ensure effective research planning, we carefully consider the planning activities, including specific dates for each task. This helps us maintain a clear idea of our progress and provides a structured framework for the dissertation. For each task, we define clear objectives, describe the contents, identify the main deliverables and foresee potential difficulties and challenges. A detailed description is provided in the form of a Gantt's diagram.

During our study of the state of the art and literature review, we have identified significant and recent contributions in various areas, namely the conceptual, methodological and application levels. Some topics such as SC, IoT, Big Data, Sensorization, DM and AI, XAI are mentioned. We also discuss the new role of cities equipped with advanced information infrastructures. Numerous experiments have been conducted, resulting in in scientific contributions in terms of human mobility metrics and a detailed understanding of human behaviour phenomena. However, these promising results can be improved with the support of advanced storage tools and the use of forecasting models. To ensure clarity and transparency regarding the outcome of these sophisticated AI techniques for end users, some projects explain the human mobility results obtained from "smart"context via XAI models.

In the upcoming chapter, "Related Work and Projects", we delve into several SC projects involving other hot topics such as city planning, citizen-centred services, MaaS and XAI. The fundamental concepts and principles of smart human mobility align closely with those of SC technologies and infrastructure, presenting a large number of business opportunities and vast potential for growth. This means that urban environments planned on a citizen-centred approach invest in innovation to enhance services, reduce costs, and improve communication and interaction between local authorities and citizens.

In the ”Innovation and Research”chapter, we explore the human mobility infrastructures. Moreover, sensing the environment around us and objects populating this environment became synonymous with the concept of pervasive or ubiquitous computing. Participatory sensing has emerged as a valuable tool, providing valuable metrics and community patterns that shed light on the complex and diverse nature of human behaviour in our daily activities. Controlling and understanding human activities pose a big challenge. It essentially recognises different activities at the same time, overlaps with others or is performed in parallel by multiple citizens. Finally, preliminary results highlight the potential role of IoT in a SC for implementing ML techniques in predicting human mobility. These findings not only validate the future developments outlined in the research plan but also emphasize the importance of IoT in shaping the future of urban mobility.

The *WalkingStreet* platform is covered in our work as we aim to develop human mobility solutions in the age of SC. This platform comprises many different components, each playing a crucial role in the project plan and implementation plan. The project plan provides an architecture overview, encompassing a solution from infrastructure to application levels. These multiple levels propose decentralized, central and cloud modules. While the decentralized approach offers flexibility, it requires careful attention to infrastructure challenges, especially in extreme scenarios involving data capturing, storage and processing. The implementation plan, a key part of the ”Design and Implementation”chapter, delves into different processes within each module, such as data collection, communication, processing, management, and monitoring of human mobility data. Streamlining these processes not only optimizes the implementation plan but also contributes to cost savings that can be allocated to address other challenges faced by the platform. In the closing sections of this chapter, we emphasize the importance of acknowledging the challenges and issues that may arise in the building of this system.

The discussion of the research outcomes has provided valuable insights into human mobility, particularly, understanding that the management of human mobility is influenced by citizen participation. However, it is essential to interpret the results with caution, considering the limitations of the current research. This chapter provides a reflection on the research process, acknowledging and discussing. the limitations and potential consequences of the design, as well as the implications of the interpretation of the results.

Finally, the last chapter of this document serves as a comprehensive summary of the work accomplished and highlights the main conclusions to be drawn. Additionally, the chapter discusses potential avenues for future work.

Chapter 2

Review of the State of the Art

As we mentioned, cities are spaces where the phenomenon of human mobility can contribute to the optimization of resources, provide services and policies that improve the quality of life of those who live or visit them. Since the aftermath of Second World War, cities have undergone significant transformations, and technology has played an important role in reshaping urban dynamics. No less interesting, citizens are no longer passive observers, but active contributors to the well-being of their communities. In other words, technological support and infrastructure have empowered citizens to participate actively in the context of human mobility research.

In this chapter, we delve into how cities have become a hub of technological opportunities today. We conduct a thorough bibliographic review to understand the rapid transformation of cities and their journey towards becoming Smart City (SC). We also explore key indicators that evaluate smart cities and delve into the human mobility sensing process. Understanding the infrastructure for information acquisition and data transportation becomes crucial as they form the foundation for the Human Mobility as a Service Model, our proposed solution, which, like other SC projects already developed, aims to captivate and engage the participation of citizens.

This inclusion of the individual in a project of this size can be explained from different levels of participation. The level of citizen involvement can vary, and some argue that they are the starting point for building human mobility solutions in smart cities. However, other researchers reinforce that it is essential to acknowledge that different stages must be fulfilled in the process to create a mobility service that meets citizens' needs. In this context, this chapter discusses the properties of data storage and analysis for human mobility, and highlights specific elements of Artificial Intelligence (AI) that best align with the requirements of this type of project. Additionally, we introduce methods of explainability to ensure that the results of the solution can be easily understood by humans.

2.1 City: a technology Hub of opportunities

The world population is growing fast. Cities offer a large concentration of resources and facilities, making them attractive for people migrating from rural areas [157]. However, these emerging patterns of urbanization require innovative approaches, policies, and strategies [209]. The increasing population puts a strain on essential resources such as energy, water, public service personnel, education, and more. To better address these challenges, cities have to adapt and adopt smarter approaches to efficiently use existing resources. Thus, a smart urban development, with community-led development practices, can be considered a SC.

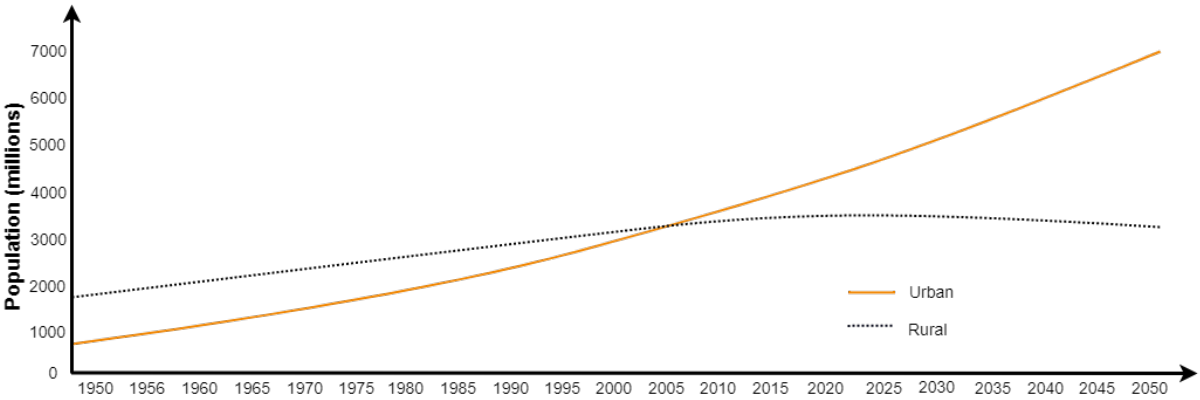

Figure 6: The world's rural and urban population, 1950–2050 (adapted from [143]).

In Figure 6, the first decade of the twenty-first century has seen the world become 'urban' [97, 204]. The Economist highlighted this transformation in 2007, stating that *"Wisely or not, Homo sapiens have become Homo Urbanus"* [81]. As this trend continues, cities must adapt and become more efficient accommodate the growing population. There are some reasons that explain how SC has become a global phenomenon. Quite simply, smart cities use Internet of Things (IoT) devices such as connected sensors, lights and meters to collect and analyze data. Consequently, this allows cities to enhance their environmental efficiency and make intelligent use of resources. By incorporating IoT technology, city infrastructures become more intelligent and capable of optimizing various aspects of urban life.

As technological Hubs, cities have been implementing smart management solutions to address diverse challenges. For instance, street lighting poses several issues, including lights being left on during the day, leading to a waste of electricity. Inadequate/poor fault management mechanisms also result in delayed repairs for faulty lights. Moreover, non-uniform light intensity levels contribute to energy wastage, while unchecked power theft and leaks from lines mean revenue loss for street lighting suppliers [98]. Waste management is another challenge modern cities have to deal with [68, 131]. It consists of different processes such as collection, transport, processing, disposal, management and monitoring of waste materials. These processes consume considerable resources in terms of money, time and labour. Therefore, optimising waste management processes helps to save money that can be used to address other challenges that SC need to tackle. Additionally, the efficient control of drivers and fleet tracking is also an important issue. Waste truck drivers require navigation systems to fulfil their tasks efficiently and effectively. In its turn, this work aims to introduce a new type of smart city, accompanied by new definitions and technologies, as an alternative solution to address the challenges of human mobility within urban spaces.

2.1.1 Sensing urban spaces

The rapid expansion and impact of the IoT on urban areas call for further research to explore the capabilities needed to adopt IoT technologies in organizations. In this survey, we aim to provide a broad overview of the IoT concept, its implementation, applications, and the challenges in poses. As depicted in Figure 7, SC and IoT, originating from different backgrounds, are converging towards a common goal. We propose the Sensing as a Service (S2 aaS) as a bridge between these two concepts.

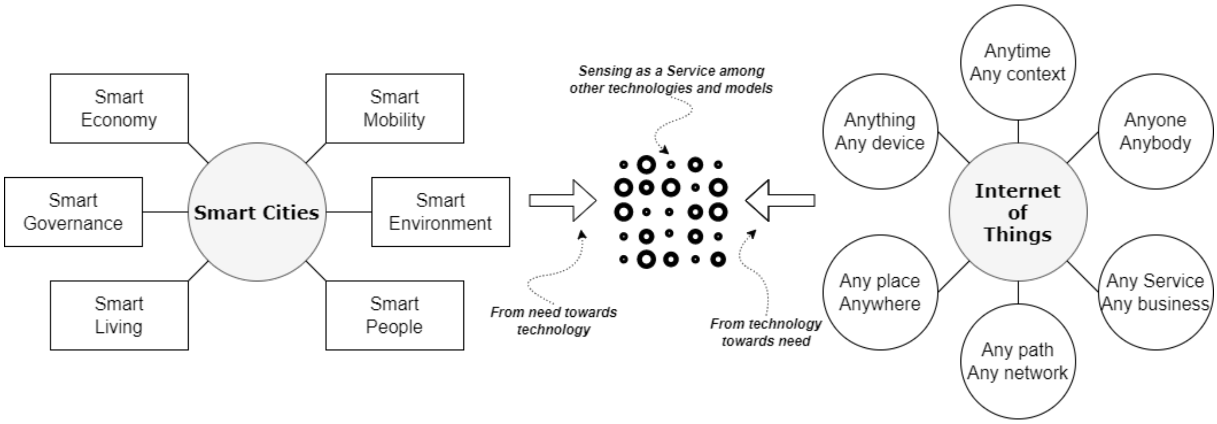

Figure 7: Relationship between Sensing as a Service, Smart City and Internet of Things (adapted from [157]).

Although all SC projects supported by the IoT have been designed for different contexts, a large part of them utilizes sensors and devices. As per industry reports, sensors are witnessing significant growth, leading to a huge volume of data, trace logs, service information, and their respective relationships on the Internet. According to Internet World statistics [135], there are currently 1.7 billion internet users, compared to the world population of more than 6.7 billion people. Approximately 40% of the world population is connected to the internet across the globe, which is only the tip of the iceberg.

2.1.2 Human Beings: a new data-source

Migration to urban areas has brought significant transformations in socio-spatial human mobility (e.g gentrification of cities). Coupled with the wide availability of devices, this trend has an impact on the big data universe, compelling authorities to quickly update their current processes, tools, and technology to handle massive data volumes and extract valuable insights from big data. As evidence of this, it is estimated that around 4.4 trillion GB of data will be generated through the IoT by the year 2020 [161].

Human mobility data represents and enormous amount of data generated from human interactions with different personal devices. One of the main motivations for collecting and analysing this is to understand human mobility patterns, such as identifying areas with the highest frequency of pedestrians or factors that cause greater concentration of people. Additionally, this data can be shared within a community, where everyone collaborates to shape the future of human mobility. Once stored, aggregated, and properly utilized, human mobility data facilitates the competition around innovative, environmentally sustainable, and user- friendly mobility concepts by giving all users equal and transparent access to relevant data. Within the human mobility data space, all individuals enjoy unique opportunities to benefit from the added-value potential of their data. In the modern world, well-analyzed data holds immense value and is integral to the functionality and capabilities of devices that have garnered global attention.

2.1.3 Smart City Indicators

Today, a city functions as a high-speed communication hub, equipped with a strong and modern Information Technology (IT) infrastructure which connects it with other cities all over the world in real-time. Caragliu et al., in their study titled "Smart Cities in Europe", argue that a city can be deemed smart if it invests in human and social capital, and a modern communication infrastructure, fostering sustainable economic growth and a high quality of life. This is achieved through the wise management of natural resources and participatory governance [37]. Similarly, Hartley defines a smart city as *"a city connecting the physical infrastructure, the IT infrastructure and social infrastructure, and the business infrastructure to leverage the collective intelligence of the city"* [85]. Meanwhile, R. H. Hollands says that *"territories with a high capacity for learning and innovation, which is built into the creativity of their population, their institutions of knowledge production, and their digital infrastructure for communication"* [90]. However, a literature survey shows that there is no singular clear cut definition of a SC. Instead, one can find many alternatives, resulting in an ambiguous meaning for the concept.

SMART ECONOMY (Competitiveness)	SMART PEOPLE (Social and Human capital)	SMART MOBILITY (Transport and ICT)
Innovative spirit Entrepreneurship Productivity Flexibility of labour market	Level of qualification Social and ethnic plurality Creativity Participation in public life	Local accessibility Availability of ICT-infrastructure Sustainable innovative Safe transport system
SMART ENVIRONMENT (Natural resources)	SMART LIVING (Quality of life)	SMART GOVERNANCE (Participation)
Attractiveness of natural conditions Pollution Environmental protection Sustainable resource management	Health conditions Individual safety Housing quality Education facilities and social cohesion	Participation in decision-making Public and social services Transparent governance Political strategies and perspectives

Table 1: Dimensions of a Smart City and related aspects of urban life (adapted from [125]).

As we can see in Table 1, a SC is characterized by excelling in six aspects of urban life, each represented by a different block. These characteristics are achieved through the smart combination of endowments and activities of self-decisive, independent, and informed citizens [119, 125]. People's participation is a crucial factor, directly or indirectly influencing all six dimensions.

These interconnected components rely on data collection and an Information and Communication Technology (ICT) infrastructure integrated into the city's hard infrastructure to deliver smart services to city actors, while effective governance is necessary for orchestrating these subsystems and ensure the success of the smart city mission. In other words, the term "smart"includes multiple features as technological advancement, interconnectedness, but also sustainability, comfort, attractiveness, and safety [28]. Aiming at technological innovation to improve the management of urban processes and the quality of life for citizens, it is a model of a city in which governments are investing to achieve balanced urban development. In this thesis planning, we intend to go beyond the conceptual framework of SC and delve into the study of digital projects created for smart urban development.

2.2 Infrastructure for Information Acquisition

The urbanization infrastructure trend in the information age has led to improvements in urban living, promoting social inclusion, economic development, and environmental sustainability [41]. As cities grow, it becomes pertinent to think through what constitutes and how to strengthen the various building blocks of a SC. Basically, the growth of cities has put pressure on industries and the scientific community to develop a smart city infrastructure. Figure 8 is an example of a SC infrastructure.

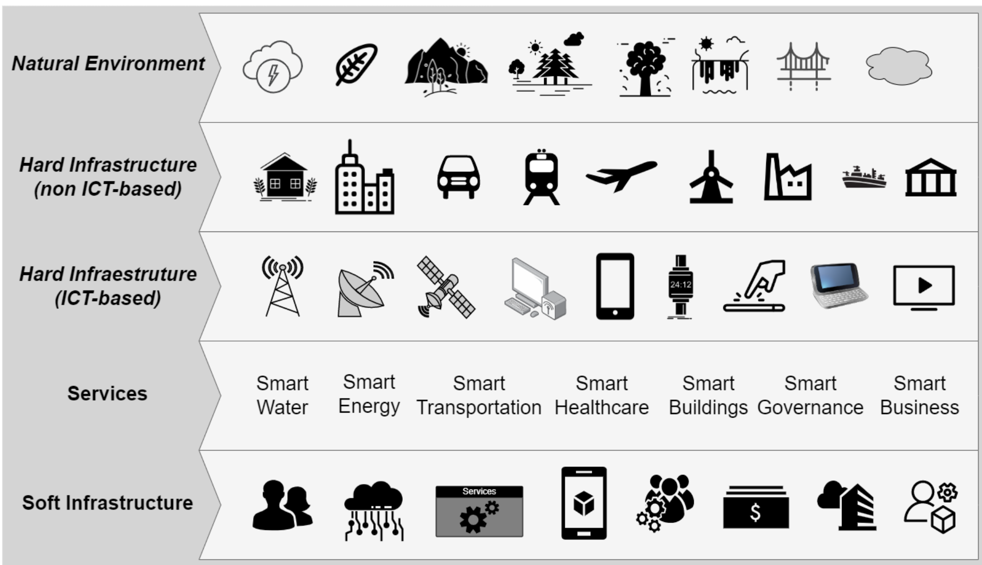

Figure 8: Multi-tier Smart City meta-architecture (adapted from [101]).

The term "infrastructure"encompasses several technological aspects, which range from information structure to technology delivery or ICT management. It also involves defining the relationships, views, assumptions, and rationale behind physical structures such as systems or buildings [16]. in other words, since SC rely on ICT for innovation production and development, the smart city architecture can be designed in meta-architecture, frameworks, and patterns.

The above process steps result in the definition of the smart city meta-architecture. It incorporates the *Natural Environment layer*, respecting all the environmental features of the city's location. In the next tier, *Hard Infrastructure (Non-ICT-based) layer* establishes a complete and sustainable infrastructure for SC. The *Hard Infrastructure (ICT-based) layer* connects the digital infrastructure, physical infrastructure, and the natural environment through the IoT. The *Smart Services layer* is delivered through both the hard and soft infrastructure components:

- Smart Human Mobility: human mobility management, adoption of a set of policies to reduce the impact of human mobility phenomenon;
- Smart Transportation: parking management, intelligent transportation, traffic management;
- Smart Government: typical administrative procedures or service co-design platforms;
- Smart Economy: enterprise resource planning, customer relationship management functions, online procurement systems or e-banking systems;
- Smart Safety and Emergency: accident management, crime prevention, public space monitoring, climate effects changes, alerting and emergencies;
- Smart Health: telemedicine, telecare and health record management;
- Smart Tourism: authorities want to offer a variety of smart city services should apply this architecture, city guides, location-based services, marketplaces, and content sharing;
- Smart Education: distance learning, digital content, digital libraries, ICT-based learning and ICT-literacy;
- Smart Buildings: building performance optimization and remote monitoring and control;
- Smart Waste Management: monitoring, city waste management, emission control, and recycling with the use of ICT;
- Smart Energy: artificial lighting, smart grids and energy efficiency's management;
- Smart Water: quality measurement, water management, and remote billing;
- Smart Living: things (e.g., doors, lamps, etc) based on home to gather information and monitor the user environment.

Finally, the *Soft Infrastructure layer* is responsible for providing new services to the city based on effective and advanced technology. Standardized frameworks are essential to support the integration of ICT services provides the SC with an adaptive capability. This adaptability is achieved by offering flexible control for the configuration of relevant data sources and context-driven filtering of relevant events to support decision-making. The system considers usage patterns and users profiles to suggest optimal configurations for smart city applications.

The SC infrastructure places significant emphasis on role of ICT components, which play individual roles within the system (i.e., communication, authentication, data repositories, etc.), as well as on the ways that all these components interact to create a comprehensive intelligent city scope [99]. Moreover, by integrating the physical, institutional, and digital aspects of the city, it enables cities and communities to find the best combination of the different intelligent city components based on their needs and objectives. In this sense, if authorities want to offer a variety of SC services, they should apply this architecture to achieve their goals effectively.

2.2.1 Multi-source Environment

Understanding human mobility behaviour allows for better urban planning. By treating mobility as a service, data on human activities is collected, enabling better estimation of demands and provision of adequate services. Thus, in our work, we employ a combination of two sensorization environment processes to capture human mobility data: location-oriented and location-enabled.

The location-oriented data collection process involves capturing human movements as they pass through a predefined location [93]. It includes several input/output points dispersed geographically to cover the target area and the target part of the network (e.g., train stations, etc.). The most straightforward way is to manually note the number of pedestrians passing a predefined location within a specific time interval. As an alternative, advanced infrastructures like LinkNYC and PLASMV allow the New York, in USA, or Aveiro, Portugal, authorities to communicate with residents and visitors in real-time [95, 152]. Content can be targeted by neighbourhood, proximity to Points of Interest (POI), or other custom attributes, and can be responsive to the surrounding environment and time of day. The different types of data recorders and sensors placed on or under the traffic network surface benefit the data collection.

On the other hand, location-enabled devices rely on the Global Navigation Satellite System (GNSS). to track movement in different locations. A constellation of satellites, wearable devices (e.g., smartwatches) and smartphones provide global coverage transmitting positioning and timing data [89, 105, 107]. This type of system also allows small electronic receivers to determine their location (longitude, latitude, and altitude) and calculate the current local time to high precision, which allows time synchronization. Additionally, these technologies simplify daily activities for individuals. While location-enabled devices offer a comprehensive approach to track mobility behaviour, respondents must carry the handheld GNSS devices continuously for accurate data collection as forgetting it would result in unreported gaps in trip data. Furthermore, to evaluate the success rates of such data collection methods, respondents often needed to manually note their trips, requiring significant effort and discipline.

2.2.2 Sensor Networks Architecture

To build interconnected and interoperable smart objects, the adoption of standard communication protocols is essential. These protocols contribute significantly to the majority of the aspects of sensor network architecture. The fundamental concepts and ideals of this architecture are closely aligned with those of SC technologies and infrastructure, presenting numerous business opportunities and potential for extensive growth. Additionally, this integration enables the management of the networks through remote monitoring, empowering IoT innovation to enhance services, reduce costs, and improve communication and interaction within smart cities. In the following sections, we review the IoT architecture and a service model specifically designed to support acquisition of human mobility information.

2.2.2.1 IoT Architecture

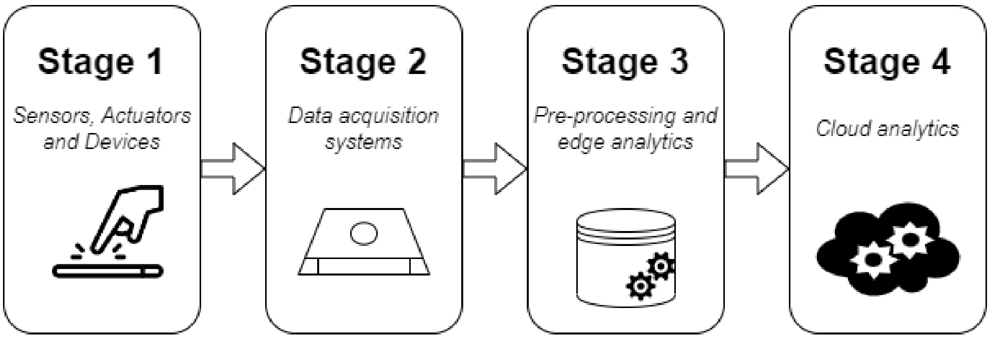

Figure 9: Four stages of Internet of Things architecture (adapted from [25]).

The IoT is on the verge of revolutionize the way we interact with "things". It goes beyond Internet-connected consumer devices and encompasses the technology to create autonomous systems that sense and respond to real-world without human intervention. However, the IoT architecture can vary depending on the specific solution being developed. Therefore, we need to develop a process flow for a definite framework over which an IoT architecture varies from solution to solution, based on the type of solution which we intend to build. Therefore, we need to develop a process flow for a definite framework over which an IoT solution is built. In Figure 9, we present a general IoT architecture comprising 4 stages:

- *Stage 1 (Sensors, Actuators or Devices):* A thing, in the context of IoT, should be equipped with sensors and actuators that enable them to emit, accept, and process signals;

- *Stage 2 (Data Acquisition Systems):* The data collected by the sensors starts is initially the analogue form, and it needs to be aggregated and converted into digital streams for further processing. Data acquisition systems perform these functions of aggregation and conversion;
- *Stage 3 (Pre-Processing, Edge Analytics):* Once IoT data has been digitized and aggregated, it may require further processing before it enters the data center, and this is where Edge Analytics comes into play;
- *Stage 4 (Cloud Analytics):* Data that requires more in-depth processing, and where immediate feedback is not necessary, gets forwarded to a physical data center or cloud-based systems. In these more powerful IT systems, the data is analyzed, managed, and securely stored.

The IoT architecture involves designing and integrating these various elements to create a robust service delivery network, which can cater to future needs. Just like the architecture of our buildings, factories, roads, bridges, towns and commercial hubs, the architecture serves as the backbone and should be carefully crafted to ensure functionality, scalability, availability, maintainability, and other essential criteria are met.

2.2.2.2 Human Mobility as a Service model

In the context of sensor architecture, the timely acquisition, transmission, and processing of data play a crucial role in today's world. The sensing as a service model aims to aggregate measures and information collected by sensors in a specific environment (e.g., a city) and share them within a larger context (e.g., a neighbourhood). This data, measurements, and information can be used to derive valuable knowledge, such as pattern analysis, to better understand how an environment operates [159].

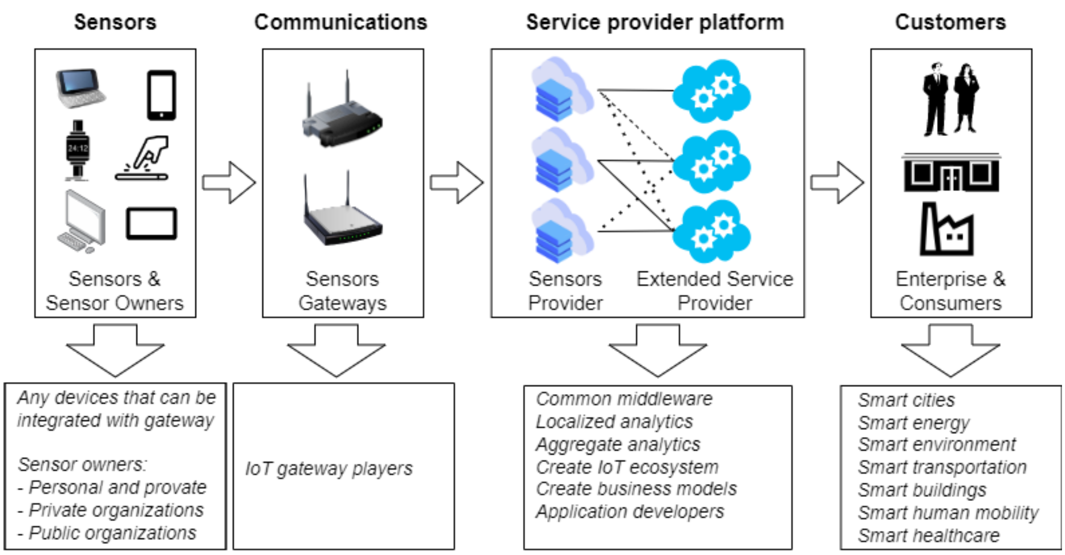

Figure 10: Sensing as a service model (adapted from [1]).

An important feature of this service model is encouraging users and owners of sensor networks to agree to share data within a larger community. By accessing and utilizing data from various sensors in different administrative domains (e.g., homes, companies, public administration and government, and social networks), this model facilitates valuable insights and cooperation among diverse sensor networks. This service demonstrates how large and independent sensor networks can collaborate effectively to serve the needs of larger communities. In addition, the open data policy adopted by these administrative entities has made petabytes of remote sensing data, with different spatial, spectral, and temporal resolutions, accessible. To effectively manage and analyze these vast amounts of data, a novel service provider platform, namely the cloud computing model where a third-party provider delivers hardware and software tools, offers highly scalable and networking services as a viable option.

Once we can categorise this service a ownership of the data, we can also use it to become "IOT-as-a-Service"that provides a free or chargeable service to any person to view or manipulate the data. Figure 10 summarises our discussion so far. The Sensing as a Service (S2 aaS) model proposes four main barriers:

- *Sensors and sensor owners/Sensor gateway* capture the information about, for example, user behaviour or user preferences more accurately;
- *Service Provider (SP)* detect available sensors, communicate with sensor owners, and get permission to publish the sensors in the cloud;
- *Extended Service Providers (ESP)* the most intelligent among all the four layers which embed intelligence into the entire service model. Each SP has access to the sensors which are registered with it. And ESP communicate with multiple SP regarding sensor data acquisition on behalf of the sensor data consumer;
- *European Commission (EC)* need to register themselves and obtain a valid digital certificate from an activity in order to access consumer sensor data. Sensor consumers can register their interests with both SP and ESP. Sensor consumers define what kind of sensor data they want and how much are they willing to offer.

To overcome these barriers, the S2 aaS model is a visionary business model that facilitates data exchange (i.e., trading) between data owners and data consumers. In this world, data owners (who own IoT solutions) are rewarded (e.g., money, loyalty points, gifts, vouchers, bitcoin, actionable advice, etc.) for sharing (i.e., trading) their personal data collected by IoT products [156]. On the other hand, companies, as data consumers, gain a better understanding of their customers, data owners. As a result, companies will be able to optimize their business operations by reducing costs and develop new products and services to meet individual customer needs. Data consumers may recover their data acquisition costs through business process optimization and increased customer (i.e., data owners) satisfaction. Data consumers can include governments or entities. Thus, the S2 aaS model offers convenience and efficiency, making it a compelling reason for adopting an IoT solution.

2.2.3 Human Mobility Tracking

Human mobility tracking involves objects equipped with sensors, actuators, and middleware communicating to serve a meaningful purpose. The actuators and sensors enable interaction with the physical world [205]. Various types of sensors can be used in multiple ways to measure the same environmental aspects. In addition, sensor boards can use be customized for IoT applications or to build specific boards for different uses, allowing the addition of any sensors we want. These technologies collect substantial data, which becomes valuable and useful once stored, organized, and processed. However, as discussed in the following sections, human mobility tracking is not just a single technology but a combination of interconnected technologies.

2.2.3.1 Networking Protocol

SC has the capacity to provide unprecedented data about urban activities, transforming the way we can understand, manage, and study cities. Access to such information about their city, enhances citizen engagement [132]. Here are some examples:

- *Sensing Pedestrians:* using sensors to count people in high streets and city centres;
- *Sensing Bicycles:* monitoring and promoting healthy travelling, while assessing the availability of green or pollution-free areas in a city centre;
- *Counting on Public Transport:* beyond city streets; buses, trains, and trams can also benefit from data collection;
- *Sensing Passengers on Buses:* enables revenue protection by –reconciling tickets bought with passenger numbers -and supports effective fleet bus management with services around the city centre;
- *Sensing the users of Trains and Trams:* monitors the number of people arriving on platforms and provides vital information on the use of routes by time of day.

As already stated, sensors have been ubiquitous in many industries long before someone even heard the term IoT. Today, with the rise of smart buildings, Industry 4.0 projects in smart factories, and IoT-driven smart city initiatives, sensors have become even more essential, with advancements in their capabilities and applications.

Just like sensors, actuators are transducers, and just like them, they have been in use for quite some time and even before the emergence of IoT. Actuators receive a signal and initiate actions to interact with and influence an environment. In a sense, an actuator can be seen as the opposite of a sensor, and it is importance is comparable. However, while most companies focus on acquiring and analyzing data, far fewer use data as triggers to make something happen in the physical world, where significant value can be found. This application extends beyond the scope of automation and is also relevant in consumer IoT applications. Some examples of actuators include:

- *Google Assistant:* a virtual personal assistant that supports both text and voice input and adapts to user conversations regardless of the input method;
- *Google Home, Home Mini, and Home Max:* Wi-Fi speakers which double up as smart home control hubs and personal assistants, allowing users to playback entertainment throughout their home, manage everyday tasks and ask Google things;
- *Apple HomeKit:* enables users to set up their iOS Devices to configure, communicate with, and control smart-home appliances.

In the scope, actuators are primarily about turning something on or off by applying force. However, actuators have plenty of applications, including in industrial fields or robotics, where they are used for grippers. They also play a role in smart consumer devices, where they work in conjunction with sensors to improve experiences, for instance enhancing sleep quality.

At times, other pieces of technology are needed such as middleware. Middleware is a software that serves as a bridge between operating systems or databases and applications, especially on a network. It facilitates the connection and management of autonomous IoT components. A middleware devoted to sensor and actuator resources management, like S2 aaS Cloud [69], enables the establishment of higher-level services, by addressing challenges such as dynamic resources provisioning, sensor virtualization, load balancing and multitenancy mechanisms [24].

S2 aaS Cloud, as a sensing-as-a-service middleware, offers sensor consumers a web application, hosted on a server instance, through which they can create virtual sensors to access sensor data that is placed in a cloud database. Moreover, when. It provides sensor consumers with a web application, hosted on a server instance, through which they can create virtual sensors to access sensor data that is placed in a cloud database. Moreover, when S2 aaS Cloud handles sensing requests through, its load balancer, which selects the server instance with the smallest outstanding request queue.

2.2.3.2 Human-Device Interaction

This type of tracking allows human interaction with small devices like smartphones, smartwatches, or any other hand-held device, opening up possibilities for sensing human mobility behaviour. Carrying a smartphone has become a habit, and is therefore considered less of a burden, reducing the risk of non-reported trips. In addition, the use of sensors provide a solid foundation for required data interpretations (e.g., using the accelerometer to determine travel mode) and improves location precision. There are three ways to sense mobile phone data:

1. *Call detail record and network signalization data* are standardized data collected by mobile network operators for billing purposes. Such data includes records of all user-initiated activities such as calls, internet usage, and data services, and provides spatial and temporal parameters [29, 226]. Moreover, the telecom operator's base station points allow for the high-precision collection of users to move from one area covered to another. Therefore, the frequency of data collection varies based on the device, network, and user activity.

2. *"Passive"tracking* involves using dedicated applications that run as GNSS-based data loggers in the background on smartphones. It is employed to investigate individual mobility patterns [33, 116], analyze speed [94], monitor traffic [88] and enable large-scale sensing of human behaviour for smart city-oriented applications [123]. Data collected provides higher spatial and temporal resolution information on the context of travel activities, allowing for the extraction of time-relevant insights.

3. *"Active"and/or "interactive tracking"* utilizes interactive mobile applications where respondents can report additional trip data at the start of the trip or transport mode. It is also used as ground truth for the development of supervised machine learning models in order to replace parts of traditional travel surveys [146]. Data collected provides higher spatial and temporal resolution information on the context of travel activities and, consequently, extraction of time-relevant insights.

Crowd-oriented sensing offers multiple advantages compared to traditional data collection approaches. It reduces the number of unreported trips that were common in travel diaries and surveys, as users tend to postpone completing them. Thus, the use of devices can support more balanced sensing of human mobility behaviour.

2.2.4 Participatory Sensing

In recent times, there has been a significant contribution of sensing by individuals, groups of people, or social networks to gather sensory information. This paradigm involves the use of evermore-capable mobile phones and cloud services to create a collaborative sensor network. The network's purpose is to collect, share, and analyze data, enabling the extraction of vital community information and constructing a spatiotemporal assessment of the phenomenon of interest [145, 164]. In essence, given the widespread use of smartphones and the mass of people in urban areas, participatory sensing reaches an extraordinary level of granularity both in time and space. This allows for a deeper understanding of human mobility phenomena occurring in the city. However, it is vital to understand the specific role of each level of granularity in participatory sensing.

2.2.4.1 Individual Data Level

At an individual level, participatory sensing fosters community engagement by enabling participants to document their neighbourhood's strengths and needs. It also promotes the distribution and decentralization of collected data so that participants can access, analyze, and make observations about the data. Moreover, this paradigm is complemented by management, mapping, visualization, and social networking through a web-based data platform [166]. This web-based platform combines data exploitation and analysis with a community-driven approach, facilitating the mapping of specialist knowledge and skills alongside relevant use cases from non-specialists.

2.2.4.2 Community Data Level

This level of granularity in participatory sensing is probably extensive, given the potential number of participants. It is environmentally focused on a group of personal devices that offers valuable learning opportunities about the human mobility context. Moreover, this approach empowers communities to actively engagement with the ordinary and everyday, enabling individuals to critically and conscientiously act on the submitted and collected data in a campaign to influence policymakers and effect changes in local and personal environments [87]. Through web-based resources, remote policy decision can be made by configuring, aggregating, comparing, and interpreting data obtained through this community-level campaign. Thus, by putting sensors into people's hands, participants can understand the significance of their individual data-sensing contributions at a community level.

2.2.4.3 Social Network Data Level

The rapid advancement of mobile Internet and social network services has led to the integration of online interactions with physical elements, facilitating new possibilities [153]. For example, a group of citizens can capture geo-tagged images while moving about town, and when uploaded and displayed in a community context, these images can promote neighbourhood identity and local services. Moreover, the same images can be used to prioritize maintenance services. Pedestrians could use similar techniques to document and select scenic and shaded walking routes away from roadways and combine their shared data with POI.

2.3 Human Mobility Data Analysis

As was the case for SC and IoT, there have been many attempts to define human mobility data, resulting in varying degrees of popularity and citation. However, none of these proposals has deterred authors of Big Data-related works from proposing new definitions or building upon existing ones. In 2010, Apache Hadoop defined big data as *"...datasets, which could not be captured, managed, and processed by general* *computers within an acceptable scope"*[42]. The International Data Corporation later defined *"big data technologies as a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis"* [74].

Big data is not a new concept, initially limited to a few organizations such as Google, Yahoo, Microsoft, and the European Organization for Nuclear Research. Nevertheless, advancements in technologies such as sensors, computer hardware, and Cloud storage, have increased storage and processing capabilities while rapidly reducing costs. As a result, many sources (sensors, humans, applications) have started generating data. Organizations now have the ability to store this data for a long time due to the affordability of storage and processing capabilities. Once stored, the role of big data in IoT is to process this large amount of data in real-time, followings three sequential steps as shown in Figure 11.

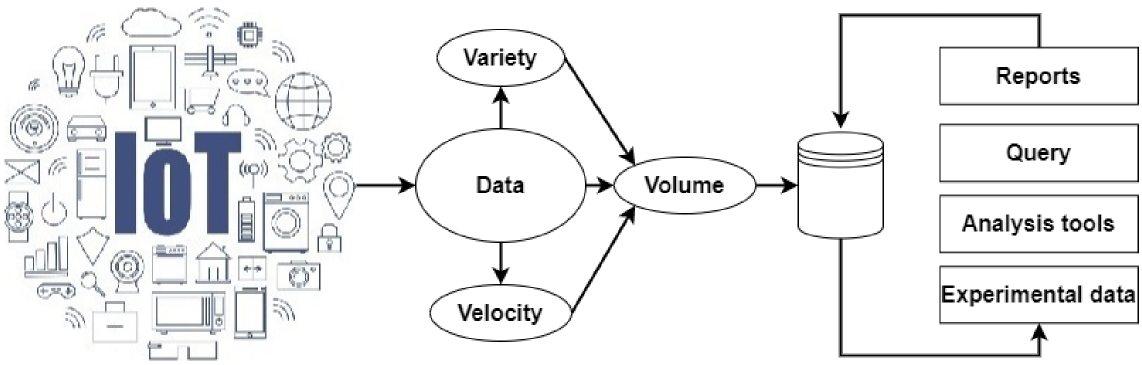

Figure 11: Internet of Things big data processing (adapted from [208]).

A large volume of unstructured data is generated by IoT devices and collected within the big data system. This IoT generated big data is heavily influenced by the 9V factors discussed in Section 2.3.1 2.3.1, namely volume, velocity, and variety. In the big data system, which operates as a shared distributed database, a huge amount of data is stored in big data files. To analyze this stored data, we would use analytic tools like the *KNIME Platform*. As IoT unstructured data is collected via the internet, big data for the IoT need lightning-fast analysis with large queries to gain rapid insights from the data and make quick decisions. Hence the need for big data in IoT is compelling.

2.3.1 Characteristics

Various academics have out forth different definitions of Big Data, Del Mauro's proposal defines it as *"the Information assets characterized by such a High Volume, Velocity, and Variety to require specific Technology and Analytical Methods for its transformation into Value"* [56]. Ed Dumbill, the program chair at the O'Reilly Strata Conference, describes Big Data as *"...data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of our database architectures. To gain value from this data, we must choose an alternative way to process it"*

[60]. Finally, Doug Laney introduces the widely spread idea of three vectors [57], stating that: *"Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation"*.

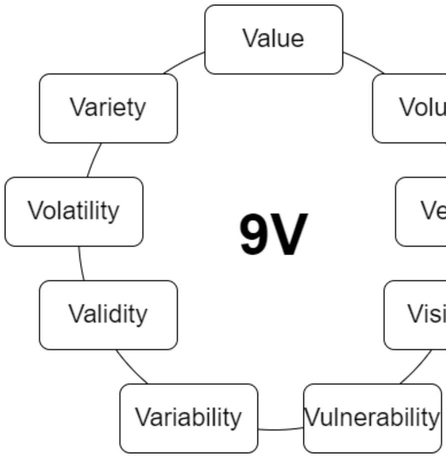

Figure 12: Big data: 9V characteristics (adapted from [51]).

As mentioned in Section 2.3, there is no clear definition for 'Big Data'. Big data, as its name indicates, represents massive amounts of data, and in addition to volume, it exhibits characteristics of variety, velocity and veracity. In Figure 12, three characteristics are used to define big data, also known as 9V's [71, 137]:

- *Volume* relates to the size of data, often measured in terabytes, petabytes, zettabytes, etc. The data is so large and complex that conventional data processing methods are inadequate for storing and analyzing it.
- *Variety* refers to the diversity of data types and data sources. In addition, different sources will produce big data such as sensors, devices, social networks, the web, mobile phones, etc, and can include weblogs, Radio-Frequency Identification sensor readings, unstructured social networking data, streamed video, and audio. Approximately 80 percent of the data in the world today is unstructured, meaning it lacks a predefined data model or organization.
- *Velocity* refers to the speed at which data is generated, analyzed, and reprocessed. Today this is mostly possible within a fraction of a second, known as real-time. For example, every 60 seconds, there are 72 hours of footage uploaded to YouTube, 216,000 Instagram posts, and 204 million emails sent.
- *Veracity* relates to the unreliability associated with data sources. For instance, sentiment analysis using social media data (e.g., Twitter, Facebook) is subject to uncertainty. It is essential to distinguish reliable data from uncertain and imprecise data and manage the uncertainty associated with the data.

- *Variability* introduced as an additional dimension by SAS, deals with inconsistency in the velocity of big data flow, which is referred to as variability. Data generated from various sources and locations may have complexities in management ranging from transactional data to big data and different semantics.
- *Value* data in its raw form is unusable; it needs to be analyzed to extract high value. For example, logs from a website cannot be used in their initial form to obtain business value. It must be analyzed to predict customer behaviour.
- *Vulnerability* is the evaluation of security vulnerabilities against cyber-attacks targeting large data volumes and providing solutions.
- *Validity* concerns the accuracy rates in predictions made by data science applications.
- *Volatility* addresses how long a large amount of data is kept in databases, data loss, and sudden manipulation of data with different characters.

2.3.2 Cloud Storage Environment

Cloud computing is logically related to Big Data. The word ”cloud”is used as a metaphor to represent the internet. Cloud technology means an internet-based technology used to store and manage business data [191]. Unlike a physical data centre consisting of hardware and pieces of machinery for data storage, all data is stored and accessible through the internet. Cloud technology involves the process of cloud computing, where various applications can be used to manage and coordinate data on the internet. There are advantages and disadvantages to this technology.

- **The disadvantages**
  - *Security:* When all data and networks are stored and run in the cloud, it is difficult to constantly oversee its safety and functioning. With data hackers always on the lookout for vulnerabilities, ensuring data security by the cloud provider becomes essential, making users skeptical of its adoption.
  - *Lock-in:* Although cloud service providers promise flexibility and integration, switching cloud services is something that hasn't yet completely evolved. Organizations may find it difficult to migrate their services from one vendor to another. Hosting and integrating current cloud applications on another platform can lead to interoperability and support issues. For instance, applications developed on the Microsoft Development Framework (.Net) might not function properly on the Linux platform.

- *Reliability:* With a managed service platform, cloud computing offers greater reliability and consistency compared to in-house IT infrastructure. An organization can benefit from a vast pool of redundant IT resources, along with a quick failover mechanism that seamlessly transitions hosted applications and services to any of the available servers, in case of severe failure. Furthermore, the cost of owning a cloud service is different from customizing the service to suit our business needs and when enough capital has been put into the former, the next thing is making our data accessible to the public.
- *Lack of Control:* With cloud infrastructure entirely owned, managed, and monitored by the service provider, customers have limited control over the back-end infrastructure. They can only control and manage the applications, data, and services operated on top of the cloud, without direct access to key administrative tasks such as server shell access, updating, and firmware management.

- **The advantages**
  - *Scale and Cost:* While it is true that maintaining our business in the cloud can be expensive, cloud providers offer customers cost-scalable options like one-time payment and pay-as-you-go, providing flexibility in paying for services based on actual usage and budget constraints.
  - *Encapsulated, Change, Management:* Cloud computing provides enhanced and simplified IT management and maintenance capabilities through central administration of resources, vendor-managed infrastructure, and Service-level Agreement backed agreements. backed agreements. This eliminates the need for in-house IT infrastructure updates and maintenance.
  - *Next Generation Architectures:* Cloud computing harnesses ever-increasing computing resources, giving businesses a competitive edge over competitors, by reducing the time required for IT procurement and enabling quicker market entry for newer applications and services.
  - *Choice and Agility:* While this may not directly address the challenge of transitioning between cloud providers, it highlights the inherent scalable nature of the cloud. User can customize their cloud platform to suit their requirements. When additional bandwidth is needed, a cloud-based service can rapidly provide it without the complexities associated with physical IT infrastructure.

Despite the inherent pros and cons of every technology, cloud computing remains a trusted solution for the growing business demands. In line with this, our proposed Mobility as a Service (MaaS) architecture, *WalkingStreet*, integrates Cloud Computing service as a fundamental component. Within the architecture, this Cloud service plays a central role in centralizing and storing information from other dispersed upstream components, which handle high data volumes. The Cloud computing module also encompasses elements like a frontend platform, a backend platform or servers, a network service and a cloud-based delivery service. In the frontend, there is a client-side ReactJS application interacting with the Cloud environment, while the backend relies on the resources provided by the cloud computing service, including a server, data storage, security mechanisms and provider's control. For more in-depth information on Cloud Storage Environment, refer to Section 5.2.4.

2.3.3 Exploratory and Metrics Methods

Until now, we have talked about characteristics data and storage environments data. However, in the context of human mobility data, it is crucial to explore and experiment with advanced techniques metrics, and exploratory data analysis. Moreover, this exploration can lead to data-driven decisions that will enable stakeholders, including local authorities and other public entities, to enhance city management. To put all of that into perspective, we present a host of important analytical methods and techniques which form the foundation of data understanding. However, before delving into these methods, let us first define the properties of the human mobility data.

2.3.3.1 Characteristics of Human Mobility data

There are several properties in human mobility data, encompassing static points and movement trips, which are crucial for geospatial analytics. Static point data allows the analysis of metrics or events concerning entities that remain stationary, like average prep time at a restaurant. Aggregation of data points is used for events occurring at a certain place and time, such as aggregating orders delivered onto hexagonal grids within a city. However, to understand the metrics between two events during a trip, movement analysis is very useful. Thus, human mobility refers to people's movement in both spatial and temporal dimensions [128, 196].

Movement analysis is primarily done on ping data, which helps to understand how citizens and assets move on the ground. This includes analyzing the origin points, where trips begin, such as residential areas or transport hubs. Similarly, destination points represent where the trips end, typically corresponding to users' homes during evenings. Moreover, paths taken by individuals are also an important aspect of human mobility analysis, identifying the most common routes taken by citizens. Distance and Time also serve as properties, defining the average distance of trips and how it varies over time [220]. For example, a lot of trips to the beach are long-distance and amore frequent on weekends. Lastly, flow analysis (or network analysis) helps understand how the city moves at different times of the day, revealing patterns like the flow of pedestrians towards the city centre. In summary, the characteristics of human mobility data encompass temporal, spatial, and social aspects, aligning with the proprieties defined in Section 1.2.1.

2.3.3.2 Metric Definitions

In this section, we highlight important evaluation metrics that characterize human mobility behaviour. In recent years, they have provided a great opportunity to study individual and community mobility phenomena, leveraging multi-source spatiotemporal-social trajectory datasets. With that in mind, it should be noted that different metrics tell different stories. Its effectiveness depends on how they are selected, combined, and used to inform actions. The individual mobility models focus on characterizing the mobility patterns of individuals and introduce the following features:

- **Displacement:** This feature quantifies the distance covered by an individual along a path, capturing human movement dynamics in both short-term and long-term paths. These paths can be a set of locations for a pedestrian. Although the Euclidean formula is more appropriate for short distances, the Haversine formula is preferred as it considers the curvature of the Earth [43, 177]. This formula is shown in Equation 1.

$$d = 2R \times \arcsin \sqrt{\sin^2 \times \frac{\phi_2 - \phi_1}{2} + \cos\phi_1 \times \cos\phi_2 \times \sin^2 \times \frac{\lambda_2 - \lambda_1}{2}} \tag{1}$$

  where R is radius of earth, ϕ is latitude and λ is longitude. The displacement between two consecutive points will be calculated for each user at a time interval.

- **Duration:** This metric calculates the time taken between the departure time or origin v_i and the end time or destination v_j in a Human Mobility network. It reflects the duration of time spent in visiting different locations, such as determining an individual's home and workplace. By analyzing duration, we capture the basic characteristics of human mobility [219].

- **Entropy:** This metric quantifies the level of randomness or disorder in probability distributions. It assesses how evenly the probability mass is spread across all possible outcomes when considering an individual's movement among N possible locations [212]. Entropy is calculated using Equation 2.

$$H(X) = -\sum_{x} p(x) \times \log_2 p(x) \tag{2}$$

  If X can take the values $x_1, ..., x_n$ and $p(x)$ is the probability associated with those values given $x \in X$. The logarithm of base 2 is the usual choice but any other base can be used since $Hb(X) = \log_b(a)[Ha(X)]$. Moreover, by convention, the entropy $H(0) = 0$ since $\lim_{x\to 0} p \times \log_2(p) = 0$. This magnitude determines the average or expected information of an event measured as the mathematical expectation $E[log(1p(X))]$. It is also based on the probability of occurrence of each symbol, indicating that entropy is not a function of the values of the series themselves but rather a function of their probabilities.

- **Interval:** Interval refers to the elapsed time between two consecutive trips, which is different from duration [227]. For taxi operations, taxi drivers cruise for a new passenger after passengers get off. So, the interval metric measures the duration of the non-occupied state for taxis and indirectly reflects travel demands. In other words, it indicates the degree of taxi usage. Obviously, a taxi with shorter intervals between trips will make more money as it can pick up more passengers.

- **Radius of Gyration:** This radius calculates the accumulated distances from the centre of mass of a user's trajectories, indicating their spatial coverage [182]. The resulting value is directly proportional to the travel distance in relation to a centre of mass and provides insights into an individual movement patterns. For this computation, the function only requires one parameter: points (p), since the centre of mass (p_c), is optional. The turning radius (r_g) metric is calculated as shown in Equations 3 and 4.

$$r_g = \sqrt{\frac{1}{N} \sum_{i \in L} n_i \times (r_i - r_{cm})^2} \tag{3}$$

  where N is the total number of visits (or time spent) a particular individual made to all his/her visited locations L, n_i are the visits (time spent) at a particular location; L is the set of visited locations, r_i is the GPS position of the location recorded as latitude; and longitude and r_{cm} is the centre of mass of the trajectories defined as:

$$r_{cm} = \frac{1}{N} \sum_{i=1}^{N} r_i \tag{4}$$

We also provide a summary of the subject's development and review the state-of-the-art mobility models, including the community mobility model (also known as the population mobility model). In its turn, this type of model mainly focuses on the mobility patterns of groups of people between two regions in urban areas. It can predict the distribution of migratory flow at some time in the future based on the population of regions.

- **Predictability:** Context plays a vital role in human mobility, and an individual's predictability can provide insights into their context preference. Estimating an individual's predictability helps in understanding their behaviour, such as whether knowing that a person goes shopping will often inform us about their overall predictability in other activities [202]. Basically, this context preference is represented by analyzing the relative frequency of an individual staying at places from each category. A Machine Learning (ML) tool can then be used to determine an individual's predictability based on their context preference.

- **Gravity Model:** This model assumes that the number of travelers between two locations (flow) increases with the populations of the locations while decreasing with the distance between them [190]. Given its ability to generate spatial flows and traffic, similar to Newton's gravity law, this model can estimate demand between two geographical locations i and j, the number of trips F_{ij}, based on their population sizes (P_i and P_j) and the distance between them, d_{ij} by using Equation 5 [91].

$$F_{ij} = k\frac{P_i^n P_j^m}{f_\gamma(d_{ij})} \tag{5}$$

- **Radiation Model:** In this model, the total human mobility traffic going from P_i (i.e., source population) to P_j (i.e., destination population) depends not only on the populations in both wards but also on the number of opportunities between P_i and P_j, represented by s_ij, which is measured as the total population within a circle of radius π_{ij} [181]. The radiation model is shown in Equation 6.

$$\pi_{ij} = \frac{P_i P_j}{\left(P_i + S_{ij}\right)\left(P_i + P_j + S_{ij}\right)} \tag{6}$$

- **Intervening opportunity:** In this approach, the destination zones j are ranked according to their distance to i (e.g $j = 1$ for the nearest zone, $j = 2$ for the second nearest zone, etc..) and $j = J$ for the zone furthest away. The probability π_{ij} that someone from zone i makes a trip to zone j is then estimated by Equation 7 [115, 206].

$$\pi_{ij} = \frac{\exp\left(-\alpha D_{j-1}^{\text{cum}}\right) - \exp\left(-\alpha D_j^{\text{cum}}\right)}{1 - \exp\left(-\alpha D_J^{\text{cum}}\right)} \tag{7}$$

  Where D_j^{cum} is the cumulative attraction of intervening opportunities between origin i and destination zone j, including those in j, and α is a positive scale parameter. The cumulative attraction can be calculated by iteration: $D_j^{\text{cum}} = D_{j-1}^{\text{cum}} + D_j$. For the nearest zone, j = 1, the cumulative attraction in j is equal to the attraction in j, D_j , and the attraction in $j1$ is equal to 0. The numerator is then equal to $1 - \exp\left(-\alpha D_j\right)$. From this, it follows that the denominator should be included to satisfy D_j^{cum}.

These human mobility metrics serve as valuable quantitative tools used to assess the quality and impact of human mobility outputs. However, each metric only tells a part of the story and each metric also has its own scope and limitations, making it essential to consider multiple metrics in combination to obtain a comprehensive understanding.

2.4 Smart Human Mobility

The increased prevalence of IoT sensors present both opportunities and potential risks in terms of infringing citizen sovereignty [21]. On one hand, IoT sensors can be used to identify and address certain issues in a community, allowing residents to have a say in what problems are prioritized for resolution. But even then, there is also a risk of infringing on citizen sovereignty if the deployment of IoT sensors is biased, focusing only on certain areas or relying too heavily on historical data from specific locations. Biases in the data can lead to unequal representation and may impact the effectiveness of the approach.

On other hand, the possibilities and problems of human mobility data mean developers are facing the challenge of bridging the gap between technology and the humanities [163]. As technology becomes more sophisticated, designer hold a key role in customizing such concepts for mass use. Additionally, as the focus shifts from technological solutions to considering the citizen's experiences, the impact of different political and cultural systems in various countries will become more pronounced. The old adage that "all politics is local"will be reinforced.

2.4.1 Constraints Address

Human mobility data indeed presents several challenges that span theoretical, technical, technological and practical aspects. Classify this data is not straightforward due to its multidimensional nature, comprising a combination of many aspects to reveal the whole picture. Moreover, the term "human mobility data"it is a buzz phrase widely used in different contexts and continues to evolve over. One of the primary challenge lies in data collection. It is important to capture the context in which the data has been generated and to filter out the noise during pre-processing and compressing. Data pre-processing is a complex and time-consuming, especially when dealing with big volumes of unstructured and structured data originating from a large number of sources. Handling such data continuously requires innovative technologies and architectures, designed to efficiently extract value from very large diverse datasets, by enabling high-velocity capture, discovery, and/or analysis.

Different data aggregation and representation strategies may be needed for different data analysis tasks. To effectively work with that data, a streamlined approach is essential, combining powerful analytics tools with efficient data movement from its source to an analytics platform quickly. Both big data and IoT technologies are continuously evolving to meet the changing requirements of the information technology field. However, several requirements and challenges must be addressed to capture and use human mobility data effectively [158, 199]. These challenges are as follows:

- *Privacy Issues:* Concerns about personal privacy have escalated due to the proliferation of online platforms: the internet is booming with social networks, e-commerce, forums and blogs. Because of privacy issues, people are apprehensive about their personal information being collected and used in unethical ways that can potentially cause them a lot of trouble.

- *Security issues:* Data security is a major concern, especially as businesses own sensitive information about their employees and customers, including their social security number, birthday, payroll, etc. Instances of data breaches and hacking have raised questions about the proper collection and protection of such data. There have been a lot of cases where hackers have accessed and stolen customer data from big corporations such as Ford Motor Credit Company and Sony.

- *Mining different kinds of knowledge databases:* Data Mining (DM) should encompass a wide range of tasks, including data characterization, discrimination, association, classification, clustering, trend and deviation analysis, and similarity analysis.

- *Interactive mining of knowledge at multiple levels of abstraction:* The DM process should be interactive, allowing users to refine their search for patterns based on returned results.

- *Incorporation of background knowledge:* Utilizing background knowledge can guide the discovery process and enable patterns to be expressed concisely and at different levels of abstraction.

- *Efficiency and scalability of DM algorithms:* To effectively extract information from a huge amount of data in databases, DM algorithms must be efficient and scalable.

- *Parallel, distributed, and incremental mining algorithms:* Given the huge size of many databases, the wide distribution of data, and the computational complexity of some DM methods, algorithms that can be processed in parallel are essential for handling the computational complexity of DM methods.

Some of the challenges have been addressed in recent DM research and development, but some are still at a research stage.

2.4.2 Data Mining

Data Mining (DM) is a fundamental technology that forms the basis for AI. It is a process to extract implicit information and knowledge which is potentially useful but not known in advance from vast, incomplete, noisy, fuzzy, and random data [5]. The key distinction between DM and data analysis (as discussed in Section 2.3) is that DM aims to uncover information and discover knowledge without relying on clear assumptions.

This technology, also referred to as Knowledge Discovery in Databases (KDD), employs various methodologies to analyze data from different dimensions and perspectives. It seeks to reveal previously unknown hidden patterns, classify and group data, and summarize identified relationships. Figure 13 shows DM as a crucial step in an interactive knowledge discovery process.

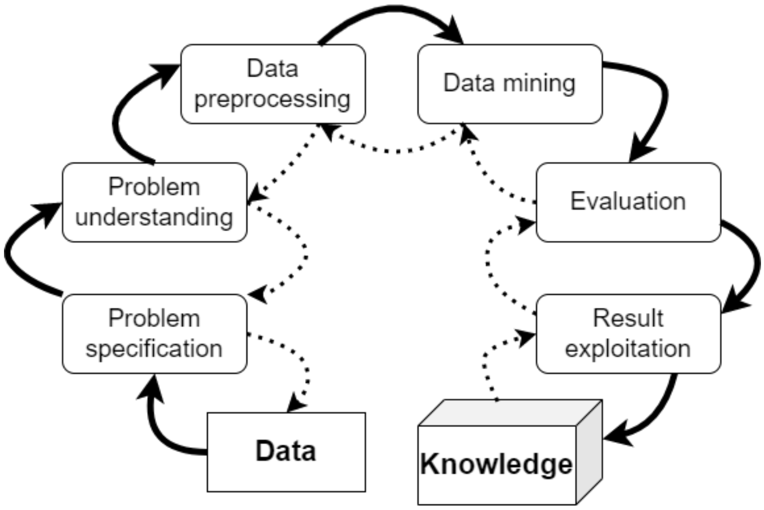

Figure 13: Knowledge discovery from data process (adapted from [75]).

It envolves the non-trivial extraction of implicit, previously unknown, and potentially valuable information from data stored in databases. While DM and KDD are often used interchangeably, data mining is actually just one step within the broader knowledge discovery process. In DM, large datasets are analyzed to identify patterns, and it relies significantly on statistical methods. The tasks are twofold:

- Creating predictive models to predict unknown or future values of the same or other features;
- Creating descriptive patterns that are interesting and interpretable to humans, providing insights into the data.

This process uses automated data analysis techniques to uncover previously undetected relationships among data items and is commonly applied to data stored in data warehouses. In this context, clustering and regression are two of the major DM techniques used in the development of this thesis.

2.4.2.1 Clustering

Clustering is the process of identifying similar classes of objects [59, 92]. By using clustering techniques, we can group data objects into clusters based on their similarity, revealing overall distribution patterns and correlations among data attributes. They can be used as a pre-processing approach for attribute subset selection and classification. The different types of clustering methods are:

- **Partitioning Method:** Suppose we are given a database of n objects and the partitioning method constructs k partition of data. Each partition will represent a cluster and $k \leq n$. It means that it will classify the data into k groups, satisfying the following requirements:

  - Each group contains at least one object.
  - Each object must belong to exactly one group.

  For a given number of partitions (say k), the partitioning method will create an initial partitioning. Then it uses the interactive relocation technique to improve the partitioning by moving objects between groups.

- **Hierarchical Methods:** This method creates a hierarchical decomposition of the given set of data objects. We can classify hierarchical methods on the basis of how the hierarchical decomposition is formed:
  - *Agglomerative Approach:* Also known as the bottom-up approach, it starts with each object forming a separate group and the repeatedly merges objects or groups that are close to one another until all of the groups are merged into one or until the termination condition holds.
  - *Divisive Approach:* Also known as the top-down approach, it starts with all objects in the same cluster and then iteratively splits clusters into smaller ones until each object is in its own cluster or the termination condition holds. This method is rigid, i.e., once a merging or splitting is done, it can never be undone.
- **Model-based Method:** This method hypothesizes a model for each cluster to find the best fit of data for the given model. It locates clusters by clustering the density function, which reflects the spatial distribution of data points. This method can automatically determine the number of clusters based on standard statistics, considering outliers or noise, leading to robust clustering.
- **Constraint-based Method:** In this method, clustering is performed by incorporating user or application-oriented constraints. A constraint refers to the user expectation or the desired properties of clustering results and provide an interactive way of communicating with the clustering process. Constraints can be specified by the user or the application requirement.

2.4.2.2 Regression

The regression technique can be adapted for prediction. Regression analysis helps model the relationship between one or more independent variables and dependent variables [7, 65, 180]. In DM, independent variables are attributes already known, while the response variables are the ones we want to predict. Unfortunately, certain real-world problems are not easy to predict. For instance, sales volumes, stock prices, and product failure rates may be difficult to predict due to the complex interactions of multiple predictor variables. In such cases, more complex techniques (e.g., logistic regression, Decision Tree (DT), or Neural Network (NN)) may be necessary to forecast future values accurately. The same model types can often be used for both regression and classification. The types of regression methods are:

- **Non-linear Regression:** The simplest way of modelling a non-linear relationship is to transform both the forecast variable y and/or the predictor variable x before estimating a regression model. This transformation creates a non-linear functional form, although the model itself remains linear in its parameters. A $log - log$ functional form is specified as:

$$\log y = \beta_0 + \beta_1 \log x + \varepsilon. \tag{8}$$

In Equation 8, the slope β_1 can be interpreted as an elasticity: β_1 is the average percentage change in y resulting from a 1% increase in x. Other useful forms can also be specified. The log-linear form is specified by only transforming the forecast variable and the linear-log form is obtained by transforming the predictor.

One of the simplest specifications is to make f piecewise linear (Equation 9). That is, we introduce points where the slope of f can change, called knots. This can be achieved by letting $x_{1,t} = x$ and introducing variable $x_{2,t}$.

$$x_{2,t} = (x - c)_+ = \begin{cases} 0 & x < c \\ (x - c) & x \geq c \end{cases} \tag{9}$$

The notation $(x - c)_+$ means the value $x - c$ if it is positive and 0 otherwise. This forces the slope to bend at point c. Additional bends can be included in the relationship by adding further variables of the above form. An example of this follows by considering $x = t$ and fitting a piecewise linear trend to a time series.

Piecewise linear relationships constructed in this way are a special case of regression splines. In general, a linear regression spline is obtained using Equation 10.

$$x_1 = x \quad x_2 = (x - c_1)_+ \quad \ldots \quad x_k = (x - c_{k-1})_+ \tag{10}$$

where $c_1, \ldots, c_{k-1}$ are the knots (the points at which the line can bend). Selecting the number and position of knots $(k - 1)$ can be challenging and somewhat arbitrary, though some software offers automatic knot selection algorithms but are not yet widely used.

For smoother results, piecewise cubics can be used instead of piecewise lines, ensuring continuity and smoothness (so that there are no sudden changes of direction, as we see with piecewise linear splines). In general (Equation 11), a cubic regression spline is written as:

$$x_1 = x \quad x_2 = x^2 \quad x_3 = x^3 \quad x_4 = (x - c_1)_+ \quad \ldots \quad x_k = (x - c_{k-3})_+ \tag{11}$$

Cubic splines usually provide a better fit to the data, though forecasting y becomes less reliable when x goes beyond the range of historical data.

- **Linear Regression:** In simple linear regression, the goal is to predict scores on one variable based on the scores of a second variable. The variable being predicted is called the criterion variable (Y), while the variable used for prediction is called the predictor variable(X). When there is only one predictor variable, this method is know as simple regression. In Figure 14, we can see that there is a positive relationship between X and Y.

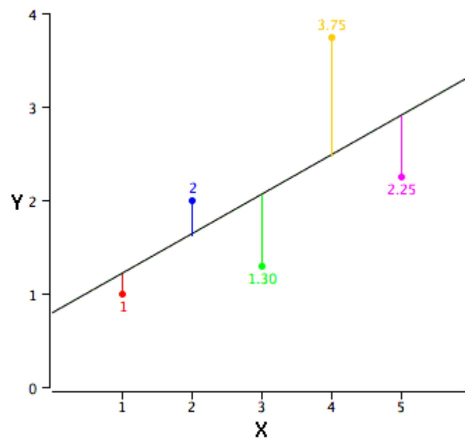

Figure 14: A linear regression scatter plot of the example data.

The essence of linear regression involves finding the optimal straight line that best fits the data points. This line is referred to as the regression line. The black diagonal line represents this regression line, offering predicted Y scores for each conceivable X value. The vertical lines drawn from the data points to the regression line represent the errors of prediction. The error of prediction for a specific point is calculated as the difference between the point's value and the predicted value (the value on the line). Notably, the red point is very near the regression line, implying a small prediction error. By contrast, the yellow point is much higher than the regression line, indicating a large prediction error.

As demonstrated in Table 2, the predicted values (Y') and the errors of prediction $(Y - Y')$ are showcased. For example, the first point has a Y value of 1.00 and a predicted Y (called Y') of 1.21, resulting in an error of prediction of -0.21.

It's worth noting that the term "best-fitting line"requires clarification. Typically, the most commonly-used criterion for determining the best-fitting line is the line which minimizes the sum of squared prediction errors. This criterion was employed to derive the line depicted in Figure 14. The final column in Table 2 shows the squared errors of prediction. The sum of the squared errors of prediction is lower for this specific regression line than for any other alternative line. Equation 12 presents the formula for a regression line.

X	Y	Y'	$Y - Y'$	$(Y - Y')^2$
1	1	1.210	-0.210	0.044
2	2	1.635	0.365	0.133
3	1.3	2.060	-0.760	0.578
4	3.75	2.485	1.265	1.600
5	2.25	2.910	-0.660	0.436

Table 2: A table of the errors of prediction.

$$Y' = bX + A \tag{12}$$

where Y' is the predicted score, b is the slope of the line, and A is the Y-intercept. The equation for the line showcased in Figure 14 is:

$$Y' = 0.425X + 0.785 \tag{13}$$

For X = 1,

$$Y' = (0.425)(1) + 0.785 = 1.21. \tag{14}$$

- **Multivariate Non-linear Regression:** The term multivariate non-linear regression refers to non-linear regression involving two or more predictors (x_1, x_2, ..., x_n). When multiple predictors are involved, the non-linear relationship can no longer be visualized with a two-dimensional space [53].

- **Multivariate Linear Regression:** In the context of sample data, we measured variables Y, X_1, X_2 and X_3. Y is the continuous response variable (often called "dependent"), while $X_1, X_2, ..., X_p$ is the predictor variable (commonly referred to as "independent") [8]. Usually, the key questions revolve around predicting Y on the basis of the $X's$ and determining the "independent"influence of wind speed. These inquiries can be addressed using multiple linear regression analysis. In the multiple linear regression model (Equation 15), Y follows a normal distribution with a mean:

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p + \sigma(Y) \tag{15}$$

The model parameters β_0 + β_1 + ... + β_p and σ must be estimated from data where β_0 corresponds to the intercept, $\beta_1 ... \beta_p$ are the regression coefficients and σ = σ_{res} represents the residual standard deviation.

In addition to the regression model, although not used in this project, NN also performs classification tasks involving datasets that exhibit non-linear and intricate relationships. However, for these models to outperform conventional machine learning models such as SVM's, decision trees, logistic regression, and naive Bayes, a lot more data is needed. Fortunately, with the current abundance of datasets on the magnitude of thousands and even millions of data points and the advancements in parallel computing, training NN for regression models has generally achieved better results in smart city projects compared to other model types [10, 188].

2.4.2.3 Types of Data Mining Tools

Undoubtedly, the adage "data is money"rings true in today's world. Along with the transition to an app-based world comes the exponential growth of data. However, a significant portion of this data is unstructured, needing a systematic approach to extract useful information from data and transform it into a comprehensible and actionable format. This is where DM comes into the picture. With plenty of tools available, using AI, and other advanced techniques, the process of extracting insights from data becomes achievable.

The selection of a DM system depends on several features, including data types, system issues, data sources, functions and methodologies, query language, and graphical user interface. Each of these features plays a vital role in determining the suitability of the system. The data types feature corresponds to the system's ability to handle different data formats, such as formatted text, record-based data, and relational data. It's essential to ensure compatibility with specific data formats like American Standard Code for Information Interchange (ASCII) text, relational database or data warehouse. Then, in system issues, we must consider the compatibility of a system with different operating systems. One system may run on only one operating system or on several. It is also worth considering systems with web-based interfaces that support Extensible Markup Language (XML) data as input. Data sources refer to the data formats in which systems will operate. While some systems might only work on ASCII text files, others can handle multiple relational sources. Compatibility with Open Database Connectivity (ODBC) connections or Object Linking and Embedding Database (OLE DB) for ODBC connections is also important. Moreover, functions and methodologies are features that provide only one function such as classification, while some provide a broader array of capabilities. These could include concept description, discovery-driven Online Analytical Processing analysis, association mining, linkage analysis, statistical analysis, classification, prediction, clustering, outlier analysis, and similarity search. Finally, query language and the graphical user interface are a Graphical User Interface (GUI), important to promote user-guided, interactive DM model. Unlike relational database systems, DM systems often lack a shared underlying query language.

There are numerous DM tools available in the market, but choosing the most suitable one is not simple [20, 214]. Moreover, Data Scientists employ a variety of techniques for different types of tasks, such as cleaning, organizing, structuring, analyzing, and visualizing data. In this topic, we present a series of factors that should be considered before investing in any proprietary solution. Table 3 lists the market's top seven DM tools:

Tool Name	Type	Language	Features
KNIME	- Enterprise Reporting - Business Intelligence - Data Mining	Java	• Scalability and high extensibility. • Well-defined API for plugin extensions. • Import/export of workflows. • Intuitive GUI.
Weka	- Machine Learning	Java	• Classification and regression. • Three graphical user interfaces. • Poor documentation.
Rapid Miner	- Statistical Analysis - Data Mining - Predictive Analytics	Language Independent	• Data handling and aggregation functions. • File operators to operate. • Intuitive GUI.
Orange	- Machine Learning - Data Mining - Data Visualization	Python and C	• Visual programming and data analytics. • Large toolbox and scripting interface. • Extendable documentation.
R Language	- Statistical Computing	C and R	• Clustering and time series analysis. • Parallel computing. • Intuitive GUI
Keel	- Machine Learning	Java	• Classification and cluster discovery. • Regression and association discovery. • Intuitive GUI.
DataMelt	- Data Mining	Java	• Statistics and numeric computations. • Linear regression and neural networks. • Used in Windows and Android platforms. • Intuitive GUI.

Table 3: Data Mining tools comparative analysis.

Comparing the features of previous analytic platforms, *KNIME* provides the IT infrastructure for easy and secure deployment in the present work. It also doesn't require coding effort, making it ideal for the proposed PoC that is built by NN and DL architectures. Covering the main data wrangling and ML techniques, a key feature of the *KNIME Analytics Platform*is visual programming. This visual programming involves a workflow of dedicated nodes, where each node implements a specific task. With this workflow, there's no need to write code.

In extreme contexts where *KNIME Analytics Platform* and its extensions cannot reach, there are integrations with other scripting and programming languages, such as Python, R, Java, and JavaScript. Its flexible nodes can be combined within a component, and consequently, the component has the views (for example, a selection and visualization in one chart) of the contained nodes and connects them. Another important integration is the *KNIME Deep Learning – TensorFlow Integration*. This extension allows the conversion of Keras models into TensorFlow models. This enables the reading, writing, creation, training, and execution of DL without writing code.

2.4.2.4 Deep Learning

ML technique is currently the major application of AI. It uses algorithms to glean insights from data. Moreover, it gives systems the ability to learn from experience autonomously, without being explicit programming. In other words, the goal is to devise algorithms capable of automatic learning, without human intervention or assistance [183].

In ML, algorithms receive input values and use historical data to predict corresponding outputs. This is particularly relevant when analyzing urban human mobility datasets extracted from relevant locations in a Smart City, utilizing access points (e.g., bluetooth, wi-Fi direct access points, cellular tower or information LED screens), Global Positioning System (GPS) information by individual devices, and aggregated GPS points recorded by devices such as smartwatches or smartphones. This data should be processed to extract several variables, aiming to find human mobility patterns in the data. These patterns can inform decisions about security within specific urban areas. The search for these patterns involves ML methods and they are, often categorized as Supervised Learning algorithms. These algorithms require input and desired output data for training and provide feedback on the accuracy of predictions during algorithm training. They identify which variables or features, the model should analyze to generate predictions. Once training is complete, the algorithm will apply learned insights to new data.

Deep Learning (DL), a subset of ML, diverges from traditional shallow ML in its capabilities, allowing computers to solve a host of complex problems that couldn't otherwise be tackled. In 1950, Alan Turing said *"While successes have been achieved in modelling biological neural systems, there are still no solutions to the complex problem of modelling intuition, consciousness, and emotion - which form integral parts of human intelligence"*. Based on the structure and functionality of biological neural networks, DL emulates the workings of the human brain. Neurons in a biological neural network are connected via axons, and DL replicates this communication process [2].

Urban human mobility can already use deep learning to build data-based models and figure out the expected performance of a given mobility network. The studies of human mobility modelling and prediction play a vital role in a series of applications such as urban planning, epidemic control, location-based services, traffic forecasting systems, intelligent transportation management and, more recently, various mobile and network applications.

DL models incorporate scheduling optimization engines, suggesting more efficient schedules automatically. For instance, they can predict the likelihood of a given trip adhering to a timetable. If buses are unlikely to stick to the timetable, optimization can improve on-time performance. In the same way, optimization models enhance the accuracy and efficiency of citywide human mobility predictions. Moreover, a spatial-temporal mobility event prediction framework based on NN is a general and effective method for representing mobility events. This approach trains deep networks to derive detailed results for citywide prediction scenarios. In Section 4.2.3.1, we outline the definitions and differences of models employed during our Doctor of Philosophy (PhD) program.

2.4.3 Information Platform Diffusion

As evident in Section 2.4.2.4, ML and its role in augmented analytics have an effect on the visualization of human mobility data. There are several ways to integrate ML into data visualization tools, resulting in enhanced dynamic visualizations with real-time analytics, and expediting the data discovery process. These set algorithms of AI are designed to automatically improve analysis by processing information and accommodating continuous data streams. Through visualizations, one gains a comprehensive understanding of activities at every stage of the production chain, including the impact of new factors on existing human mobility data [17, 72]. More importantly, combining these techniques and data visualization we establish baseline metrics for performance, enabling swift responses to deviations from those standards. This approach facilitates fast and effective reactions. However, achieving the required level of granularity necessitates careful consideration of how data is offered via open-data Application Programming Interface (API). Therefore, aggregating data from multi-sources, coupled with real-time publishing and retrieval of data for users through simple Hypertext Transfer Protocol (HTTP), presents a tremendous challenge. Meeting the demand for data visualization (using a typical dashboard), by billions of global users requires the integration of several API services. In other words, a robust API infrastructure paired with a user-friendly platform, is necessary to unlock the full potential of human mobility data. These infrastructures exert considerable influence over operations.

2.4.3.1 Open-data Service API

An open-data API (or public data API) is an application programming interface for public entities and made available on the internet. This service, accessible through the network, offers unrestricted access to consumers worldwide. While some may consider the terms ”open”and ”public”APIs interchangeably, differences exist. Open API is freely shared, whereas public API may impose certains limitations in terms of sharing assets. In this case, the proprietary organization is sharing its own API and back-end data publicly and wants to retain control over its application. Alternatively, an entity might opt to release a suite of API to encourage third-party developers in vertical industries, encouraging them to discover and share novel applications, wether for academic or industrial purposes. For example, in New York, any citizen can access publicly shared data from city agencies and other partner organizations free of charges through NYC Open Data [151]. However, from the perspective of the data publisher, some challenges arise [14, 96]:

- **Management:** Once an API becomes publicly available, it is difficult for an organization to control who uses the API and how they use it. As a result, API management must be taken seriously; otherwise, it can lead to customer dissatisfaction and potential issues. For example, in the case of New York agencies, the careful decommissioning of older API is crucial to uphold consumer satisfaction and protect the organization's reputation;

- **Security:** Despite being open and universally accessible, an open API often employs security measures to regulate access and safeguard data transmission. This can involve access restrictions, data encryption, and other security protocols. For example, New York organizations that offer public data API services should prioritize encryption for data transmission and address privacy concerns for users accessing data;

- **Expand user base:** The ability to expand a user base without the cost of niche industry software development; a chance to create revenue streams by licensing new programs; and the option for an organization to retain ownership of its proprietary source code.

In a world where systems are increasingly interconnected, several publishers (ranging from community-driven organizations to public authorities) are recognizing that offering open API to the public is a strategy for widespread adoption of the services. This enables the public to access those services in new and creative ways [78, 134]. However, from the consumers perspective, certain concerns arise:

- **Reducing dependencies:** Reduced dependencies between development teams and certain application components. This can lead to less time spent fixing coding errors and the ability to still use preferred development tools alongside an open API;

- **Drawbacks of open APIs:** An open API should be treated like any other customer-facing product because the company's reputation can be significantly impacted based on how the open API is received by developers;

- **Potential software bugs:** They may exhibit software bugs that result in performance issues, security flaws, or leakage of private company data. Furthermore, open API can be problematic for developers due to the fact that the publishing company holds the majority of control. If a startup ever decides to change its API terms of use, for example, or introduce licensing fees, third-party developers have no choice but to accept it.

Despite these concerns on both the publisher and consumer sides, an ecosystem of tools and feature enhancements can be developed around this kind of API. Nonetheless, understanding human mobility data is now more vital than ever. Simply integrating an open-data service API is not enough; effective visualization is a key component in transforming data into actionable insights.

2.4.3.2 Data Visualization Dashboard

For human mobility data visualization, achieving a higher level of granularity means a more comprehensive understanding of consumers and more. Consumers can craft more specific visualization tools by building charts and graphs that draw from wider databases [76, 112]. In other words, visualization is a vital component of analytic projects, but the information presented without proper context can loses some of its insightful nature. However, data is more than just a set of numbers; insights often emerge from several parts of a set or from combining diverse sources into an actionable understanding. Thus, in visualizations, the ability to swift and logically organize information is crucial. Without it, conveying information becomes a less direct process, with a major challenge being the ease of searching for specific parameters.

By incorporating data visualization with ML tools to construct visual models, we build a more robust knowledge base for our operations, enabling the exploration of deeper connections in big datasets [144, 223]. These techniques facilitate the achievement of the desired level in user data. One of the most prominent applications is in search engines that can predict users' questions and formulate more effective data queries to deliver results. Moreover, these algorithms are designed to understand historical data and apply their findings to new information, constructing better visualization models. This process transforms datasets into well-defined narratives, offering a superior contextual framework for the information we are viewing.

In summary, regardless of the underlying infrastructure, the development of a suitable machine learning-driven visualization tool to understand human mobility phenomena has become more vital than ever. Such a tool should easily incorporate and integrate thousands of datasets available via service API, while simultaneously offering a better picture of the overall operations and room for continued growth.

2.5 Explainability for AI-based Systems

The history of Explainable Artificial Intelligence (XAI) related work dates way back to the 90s, but it was only in recent years that, significant progress was made in this field. The event which catalyzed work in this area was the Defense Advanced Research Projects Agency (DARPA) program launched in 2016 [82]. According to this initiative, explainable AI is essential to understand, appropriately trust, and effectively manage an emerging generation of AI partners.

The emergence of XAI has enhanced the lives of humans and envisioned the concept of SC using informed actions, enhanced user interpretations and explanations, and firm decision-making processes [102]. Additionally, this urban ecosystems are important to advance in the field of AI where researchers, experts, practitioners in relevant fields must work together to develop robust methodologies that empower individuals, giving them control over their information and respecting their privacy. In this sense, XAI is successful if these different stakeholders adopt and improve AI models with their knowledge.

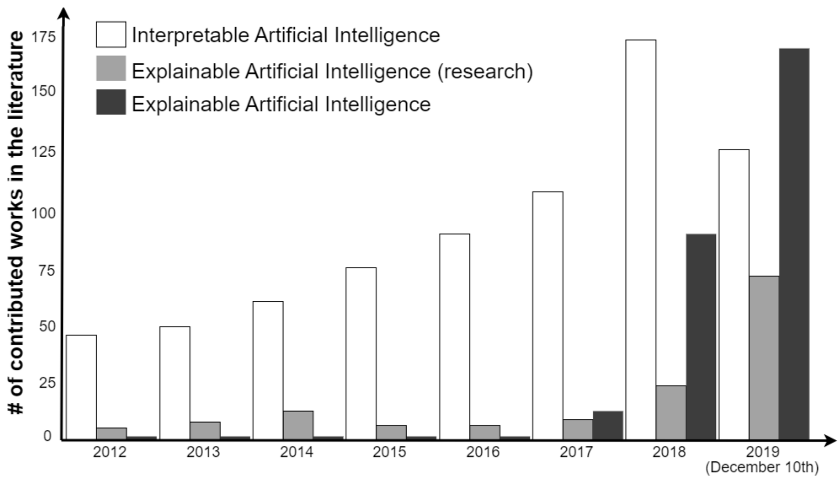

Figure 15: Evolution of the Explainable Artificial Intelligence literature (adapted from [19]).

The evolution of this explainable technique is illustrated in Figure 15. Data for this analysis were retrieved from Scopus (December 10^{th}, 2019) using the search terms indicated in the legend when querying this database [102]. It is interesting to note the latent need for interpretable AI models over time (which aligns with intuition, as interpretability is a requirement in many scenarios). However, it wasn't until 2017 that interest in techniques to explain AI models began to gain traction across the research community.

2.5.1 XAI Proposes

Undoubtedly, AI has a tremendous effect in our daily lives and its applications have already left their mark across the world. This technology's importance is underscored by substantial investments from companies in its development and, in an academic context, there are a number of topics for these and research [6, 103]. Despite its versatility, AI systems are notorious for their 'black-box' nature, leaving many users in the dark about how or why decisions are made. This is where explainable AI comes into play. XAI strives to render AI decisions both understandable and interpretable for humans. Additionally, it is more important than ever to expose how bias and the issue of trust are answered (e.g., in EU the General Data Protection Regulation provides the right to a clarification clause).

From a system point of view, an understandable explanation provides a human with information derived from and/or based on the internal model of the system, facilitating a human understand of (part of) the model's operation to comprehend its output. Importantly, explainability deals with extracting explanations from a system's model, which may not be inherently human-understandable [84, 189]. Thus, XAI considered the convergence of three areas:

- **Explainable Models:** this involves the development of new or modified Machine Learning (ML) techniques to produce more explainable models. For example, in [122] predictions about annual building energy performance were studied using explainable models. This work explains a ML model called the Quantum Lattice algorithm using a visualization approach. Localized DT models are used as surrogate models to explain decisions within a specific region in the explanation space. Additionally, XAI techniques have been applied to short-term forecasting models that predict energy performance of buildings.

- **Explanation Interface:** this area involves the Human Computer Interaction with new principles, strategies, and techniques to generate effective explanations. This area enables users to develop a level of trust in the system that accurately reflects the system's performance, and learns how users value the presentation of information (i.e., the face validity of design patterns). An illustrative instance includes explanations that furnish information for evaluating the system's accuracy.

- **Psychology of Explanation:** this aspect encompasses summarizing, extending and applying current psychological theories of explanation to formulate a computational theory. Currently, the validation of XAI is done empirically, but other alternatives and concepts have been proposed. For example, [222] posits a new psychology theory. According to this theory, when explanations are absent, humans expect AI to make decisions resembling their own, and they interpret an explanation by comparing it to the explanations they themselves would give. This comparison is formalized via Shepard's universal law of generalization in a similarity space, a classic theory from cognitive science.

In Figure 16, there is a continuous interplay among the three aforementioned areas in the evolution process of XAI performance. This performance is measured by the Explanation Scoring System (created by the US Naval Research Laboratory). It assesses multiple components within each user study (i.e., task, domain, explanations, explanation interface, users, hypothesis, data collection and analysis) to ensure that each study meets the high standards of human subject research.

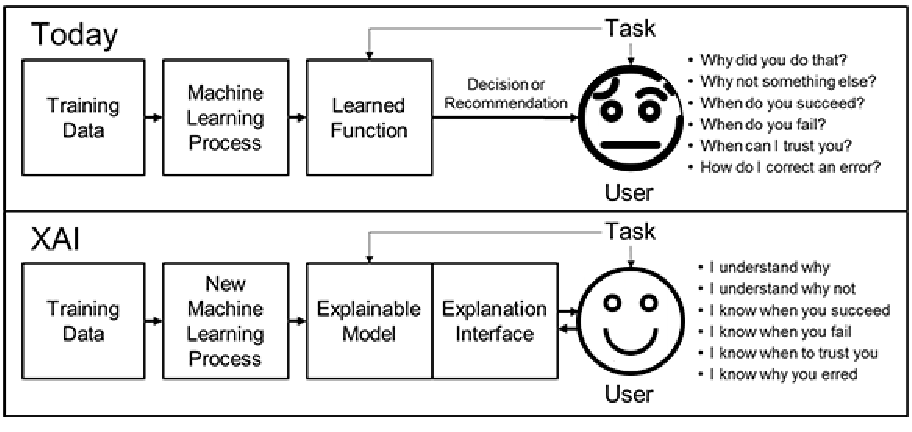

Figure 16: Explainable Artificial Intelligence project proposed by DARPA (adapted from [82]).

The preceding schema for measuring the evolution of XAI demonstrates the importance of carefully designing user studies in order to accurately evaluate the effectiveness of explanations in ways that directly enhance appropriate use and trust by human users, while also appropriately supporting human-machine collaboration [176]. Oftentimes, different types of measures (i.e., functionality, learning performance, and explanation effectiveness) are necessary to evaluate the performance of an XAI algorithm. XAI user study design can be tricky and the most effective research features often involve diverse teams with cross-disciplinary expertise(e.g., human-technology interaction and/or experimental psychology). The specific XAI evaluation metrics employed are shown in Figure 17.

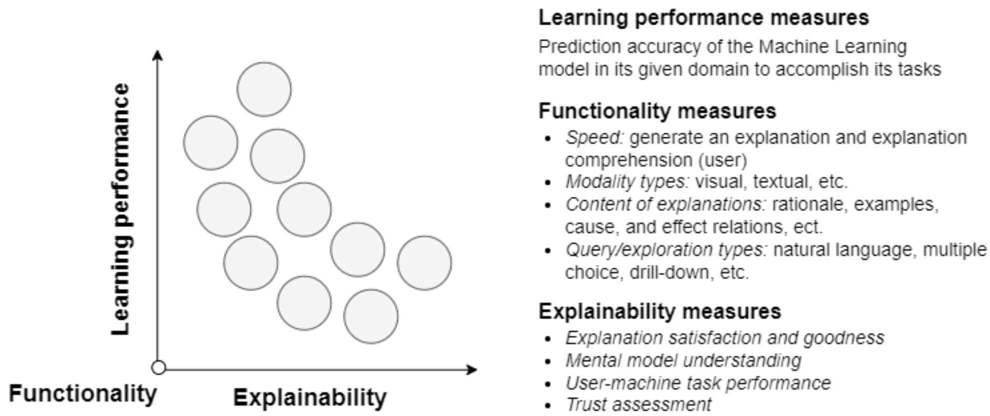

Figure 17: Evaluation measures for Explainable Artificial Intelligence algorithms (adapted from [82]).

The measures for XAI serve as guidelines to assist the author of this work in addressing the fundamental problem in an AI project, which is determining the accuracy of intelligence. These guidelines are particularly crucial when dealing with artificial systems that significantly differ from humans. When developing PoC, we use the definitions of these measures to extract their essential features for calculating the accuracy of intelligence. Moreover, the mathematical formalism for explanations and the tools for measuring how understandable an explanation is for humans enhance the practical applicability of most of the approaches presented in this thesis. Therefore, we believe that these measures formally capture the concept of AI explanations for humans.

2.5.2 XAI Research Methods

Examining the literature, we review the existing approaches regarding the XAI topic and present trends associated with it. These elements should be further analyzed within the human mobility context. In other words, an explainability project that aggregates a discussion of human mobility, XAI and their methods should be considered to provide insights into how ML algorithms operate, enabling governments or local authorities to understand and solve the diverse pedestrian challenges prevalent in smart cities. For that reason, in the subsequent sections, we synthesize a set of XAI methods useful for our regression problem.

2.5.2.1 Decision Tree

Based on Classification and Regression Trees (CART), this particular tree structure deals with all kinds of variables and predicts both numerical and categorical attributes. For our purposes, we employ a DT API developed by [38]. Inspired by the ML library, scikit-learn has an optimized version of the CART algorithm despite not yet supporting categorical variables. Demonstrating its explanatory potential, Miguel Guimarães et al. have implemented a scratch of a DT. An example of the execution of a DT via the Command-Line Interface is shown in Figure 18.

```
***************         Beginning DT training on play_reg dataset...      *********
(predict: 0.6428571428571429, std: 0.47915742374995496, var: 0.22959183673469388, num_instances: 14)
outlook = overcast
        (predict: 1.0, std: 0.0, var: 0.0, instances:[0, 1, 2, 3], num_instances: 4)
outlook != overcast
        (predict: 0.5, std: 0.5, var: 0.25, instances:[4, 5, 6, 7, 8, 9, 10, 11, 12, 13], num_instances: 10)
        humidity < 85.0
                (predict: 0.8, std: 0.4, var: 0.16, instances:[5, 6, 7, 12, 13], num_instances: 5)
                temperature < 68.0
                        (predict: 0.0, std: 0, var: 0, instances:[6], num_instances: 1)
                temperature >= 68.0
                        (predict: 1.0, std: 0.0, var: 0.0, instances:[5, 7, 12, 13], num_instances: 4)
        humidity >= 85.0
                (predict: 0.2, std: 0.4, var: 0.16000000000000003, instances:[4, 8, 9, 10, 11], num_instances: 5)
```

Figure 18: The generated Decision Tree from the command line.

The developed DT script is available as open-source, permitting programmers to freely download the source code and modify it, without necessarily depending on the API. Either through the API or executed using the DT script, the model training is not allowed if the settings given by the user are invalid. The user is also informed if there are possible improvements to be made in the settings.

2.5.2.2 LIME

This framework generates prediction explanations for any classifier or ML regressor. It manipulates the input data and creates a series of artificial data points containing only a subset of the original attributes [168]. It also provides local model interpretability. Essentially, this approach involves altering the values of features within a single data sample, then observing the resulting impact on the output. A notable feature of this technique is its ability to explain at the dataset level, discerning the importance of different features and elucidating the outcomes of models that operate on text, tabular, and image data. It accomplishes this by fitting the model using sample data points which are similar to the observation being explained. The explanations provided by Local Interpretable Model-Agnostic Explanation (LIME) for each observation x is obtained as shown in Equation 16:

$$\Phi(x) = argmin_{g \in G} L(f, g, \pi_x) + \Omega(g) \tag{16}$$

where G is the class of potentially interpretable models, including linear models and decision trees, $g \in G$: an explanation considered as a model. $\pi_x(z)$ denotes the proximity measure between an instance z and x. $\Omega(g)$ quantifies the complexity of the explanation $g \in G$. The goal is to minimize the locality-aware loss L without making any assumptions about f since a key property of LIME is that it is model agnostic. The measure L gauges how well g approximates f within the locality defined by $\pi(x)$.

2.5.2.3 SHAP

The Shapley Additive Explanations (SHAP) library, introduced by Lundberg and Lee [121], is a tool that draws inspiration from game theory. . It incorporates many attractive properties to explain the output of ML models. Additionally, it provides a means to estimate and demonstrate the influence of each feature's contribution to the model's output, allowing us to "debug"our model and observe how it arrived at a certain prediction. In its most general form, the Shapley value is the average marginal contribution of a feature value across all possible coalitions. If there are N features, the computation of Shapley values involves considering N different order combinations. From a computational perspective, it has been demonstrated that the only additive method that satisfies the properties of local accuracy, missingness, and consistency can be achieved by attributing to each variable $x'i$ an effect i as defined by Equation 17:

$$\phi_i(f, x) = \sum_{z' \subseteq x'} \frac{|z'|!(M - |z'| - 1)!}{M!} \left[f_x(z') - f_x(z' \backslash i) \right] \tag{17}$$

where f is the model, x are the available variables, and x are the selected variables. The quantity $f_x(z') - f_x(z' \backslash i)$ expresses, for every single prediction, the deviation of Shapley values from their mean: the contribution of the $i - th$ variable.

2.5.2.4 Seldon Alibi

This XAI technique is an open-source Python library designed for inspecting and interpreting ML models. It provides high-quality implementations of black-box and local explanation methods for both classification and regression models. Alibi boasts a selection of 8 different algorithms for model explanations, including popular algorithms like anchors, counterfactuals, integrated gradients, Kernel SHAP and Tree SHAP. For our work, we have opted for the Kernel SHAP algorithm.

Incorporating Alibi into the workflow necessitates the provision of training data and the invocation of the fit method. Then, the explain method is called to calculate an explanation for a singular instance or a set of instances. The outcome of this process is an Explanation object, which contains both metadata (e.g., hyper-parameter settings, names) in the meta dictionary, as well as the explanation data itself in the data dictionary. The structure of the Explanation object enables straightforward serialization in production systems, facilitating further processing tasks, such as logging or visualization.

2.6 Summary

The presentation of several concepts such as smart city hub, human mobility metrics, infrastructure for information acquisition, ML and XAI methods serves to demonstrate how cutting-edge technology incorporated into cities can significantly enhance human mobility in urban areas. From this human mobility topic, this work has discussed the potential of hub technology city indicators using a dense infrastructure network, and complemented by digital technologies that control and manage city services and systems.

Furthermore, the realization of interconnected and interoperable smart objects necessitates the establishment of information acquisition infrastructure. Once this type of infrastructure operates in smart spaces, using intelligent interfaces to connect and communicate across social, environmental, and user contexts. This infrastructure creates a worldwide network of objects based on a multi-source environment in cities, generating vast data streams. Part of this data is produced by interacting directly or indirectly with humans, depending on citizens' level of participation. This level of contribution can be used to appropriately model human mobility solutions.

Accurate characterization of data is important to adopt a better strategy to store all this information. From point of view of human mobility, cloud storage emerges as a suitable environment for managing this vast amount of data. Nonetheless, this database stores big data in a primary instance that must be analyzed. Metrics and exploratory analysis are valuable tools to unearth insights from human mobility data.

As the investigation delves deeper into the collected and stored data, some constraints should be considered, such as data anonymization of citizens and the security of open-data source API. These constraints are a relevant step to ensure they don't compromise DM process or the diffusion of forecasts and predictions from ML models' outcomes.

Finally, given that many AI implementations deal with personal data, customers need to know that this data is being handled with the utmost care and sensitivity. This entails providing customers with an explanations for decisions made. With explainable AI systems, customers can understand exactly where data is coming from and how it's being used, meeting these regulatory requirements and building trust and confidence over time.

Chapter 3

Related Work and Projects

This chapter highlights a selection of the many projects that have been developed in Smart City (SC). As previously mentioned, a city's classification as "smart"hinges on the evaluation of specific indicators. At the same time, any initiative that aims to make it smart depends on its unique attributes. In other words, city planning is shaped by the execution of intelligent projects. This requirement also extends to the human mobility project outlined in this doctoral thesis.

The architecture of human mobility proposed in this document is designed to meet citizens' needs. To this end, a detailed description of the potentialities and characteristics of Mobility as a Service (MaaS) is presented. We also highlight the interaction between the components comprising this service. The MaaS provider infrastructure has the ability to be integrated into solutions that align with a citizen-centered approach to cities. Therefore, this service is well-suited for application in the human mobility project put forth by this doctoral program.

3.1 City Planning and Development to Human Mobility

SC represents a significant breakthrough in contemporary urban development and planning processes. Three major driving forces sustain the paradigm shift towards intelligent cities. Firstly, the continuous urban population growth both sustains and challenges authorities to adopt contemporary policies for the effective city management worldwide. In its turn, the ubiquity and influence of the Internet and the web, as major technological innovations of our era, foster both voluntary and involuntary human participation. This participatory sensing can mean the acquisition of data about mobility contexts. These two views, supported by Artificial Intelligence-Driven Decision Making, have inspired the community to develop projects to revolutionize the future urban environment.

Various smart city projects offer solutions for city management wile incorporating architecture design, helping to identify main city components and their interactions. For example, Nominet, a company involved in nearly 150 projects [147], notably the Portuguese-engineered PlanIT Urban Operating System. This project converges and manages an infrastructure with a world of sensors, devices, and people across developments of scale and entire cities.

Another smart city project that is being implemented in many countries is the mobile application *TransitApp* developed by a Canadian company. It aims to be a partner for transit agencies and provide users with a seamless transit experience, through the fusion of high-quality data, great design, and thoughtful features [148].

Among governmental initiatives, the "Digital Catapult: Things Connected"in the United Kingdom stands out [48]. Supported by $3.8 million in research funding, the Digital Catapult has collaborated with over 2,000 small businesses to push forward more than 40 projects aimed at digitizing business processes in London. In Japan, the collaboration between the Fujisawa City local government and Panasonic addresses energy issues and encourages the country to become more sustainable.

Several projects have envisioned the concept of a smart city from different perspectives, taking advantage of existing knowledge and experience. Collectively, these analyses offer a "snapshot"of the evolution of smart cities.

3.2 Mobility as a Service

In recent years, MaaS has gained significant traction, particularly with the generalization of shared mobility solutions that prioritize individuals and their needs [86, 165]. MaaS operates by tailoring real-time information to users, offering a range of mobility solutions supported by different modes of transportation. While its role in changing the population's mobility patterns is not yet fully understood, the commitment of some of the main players (academic entities, private enterprises and governments organizations) suggests the rapid expansion and advancement of this concept [22, 27, 195]. These players, along with other enthusiasts, have contributed to addressing this knowledge gap and, furthering our comprehension of how end-users perceive and use this service. More specifically, they list core MaaS propositions, identified the key stakeholders involved in its development, elucidated potential ways for management (or, in other words, business models), and MaaS topologies.

3.2.1 MaaS Features

In the literature review, we have examined a comprehensive list of available MaaS definitions and similar concepts [34, 39, 216]. Researchers have explored various perspectives on relevant aspects of MaaS, analyzing existing MaaS services or analogous ones based on these aspects, outlining its distinguishing characteristics, defining and detailing MaaS levels, and aligning services with appropriate levels. While these perspectives may vary, they also share common features that characterize MaaS features:

- **Transport Mode Supporting:** Within an integrated MaaS system, the number of alternative modes of transportation's operators tends to increase. This expansion enhances the service's potential by optimizing resource allocation based on users' necessities, thereby reducing inefficiencies within the transportation system [15]. From the user's viewpoint, this will mean that, based on real-time information, he will be capable of optimizing their travel resources.

- **Technologically-based Platform:** The MaaS paradigm adopts a technologically-based definition, where MaaS is a technical system (a digital platform) accessible via web or (most likely) smartphone applications. This platform facilitates the seamless integration of the purchase process of several distinct journeys and/or travel elements and processes of transport integration and the commoditization of transport infrastructure [162].

- **Multimodal Trip Planning:** From the citizen's point of view, the journey planning tool, which interfaces with Information and Communication Technology (ICT) is central to MaaS [136]. This tool facilitates real-time information exchange between users and service providers, aiding trip planning, execution and problem-solving along the route.

- **Bundles of Products and Services:** Utilizing bundled offerings is a strategy commonly employed to enhance user acceptance and promote underutilized products and services [130]. In the context of MaaS bundle may combine less popular modes, such as bike sharing and car sharing with public transport, with the hope that this will result in increased adoption of these modes. As such, MaaS can be viewed as a soft mobility management tool, which aims to "repackage"travel services to cater to specific pedestrian needs.

- **Payment Plans:** MaaS embraces various payment plans and options. Periodic plans tailored to users' transportation needs, such as urban commuter, family, business, or casual plans, fall under the "Pay Option"category. Alternatively, "Pay on demand"allows users to pay per trip, where the entire journey is charged as a single payment, regardless of the transport means used, processed automatically via smartphone.

- **Decision-making Support:** MaaS resources are needed to produce, manage and optimize transport infrastructures and services, but also influence citizens' trip decisions, promoting environmentally friendly travel choices and minimizing the environmental impact of commutes.

- **Extra Services Add-on:** MaaS offers value-added services that enhance, the attractiveness and benefits of transportation choices for users.

- **Partnerships Promotion:** MaaS becomes even more robust when there is collaborative decision-making and cooperation among multiple stakeholders. While added values offer theoretical potential, practical implementation can be subject to technical and methodological challenges.

- **Smart Cities Promotion:** MaaS finds a natural fit in cities that use technology to improve services and the well-being of their inhabitants. Such SC is prepared to explore all the possibilities of MaaS.

Derived from these characteristics, MaaS emerges as a user-centered service which tailors itself to individual needs. Furthermore, according to Hensher, it has three structural properties: Broker (i.e., acting as the public/private sector's intermediary), Bundle (i.e., forming travel plans) and Budget (i.e., ensuring the most cost-effective trip). They are represented on Figure 19.

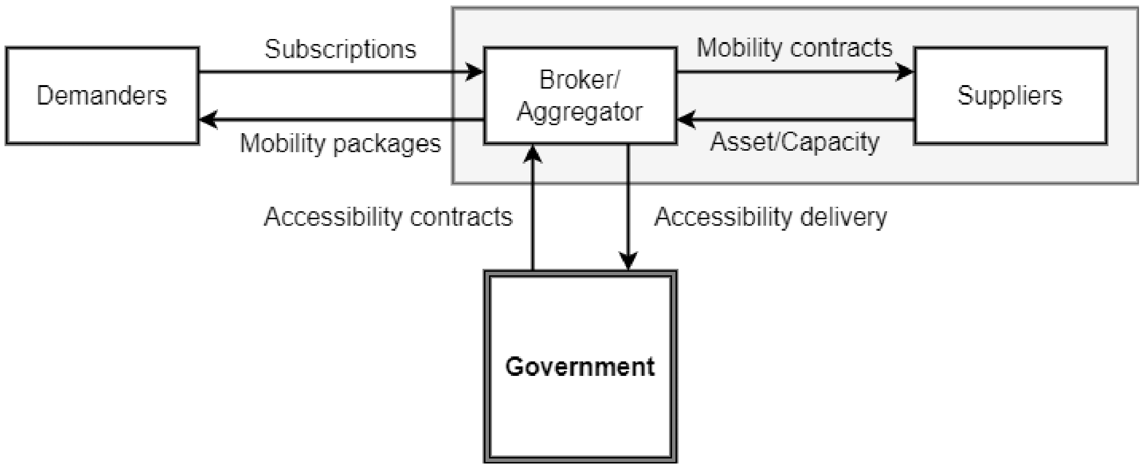

Figure 19: Main Mobility as a Service structural characteristics (adapted from [215]).

Brokers serve as intermediaries, facilitating access to different modes of transportation. These are bundled together into travel plans, referred to as Bundle. The collective strength of user demand for similar trips empowers negotiations for reduced fees. Therefore, mobility brokers aggregates different suppliers and deliveries, offering and integrated service to demanders. Moreover, this transformation necessitates authorities to reconsider their relationship with operators, influencing the design of transport modes and business models.

3.2.2 MaaS Components

As previously mentioned, MaaS is gaining significant attention in studies related to Human Mobility, Technology, and Smart Cities. Various strategies have been adopted. However, it's essential for each city considering a MaaS initiative to evaluate its local context and future strategic challenges. For these reasons, cities should prepare an infrastructure that involves its components, including both public and private devices. Simultaneously, this infrastructure should be relevant to various stakeholders, such as citizens, local authorities, and enterprises. Figure 20 illustrates a set of recommendations for developing a MaaS programme in cities.

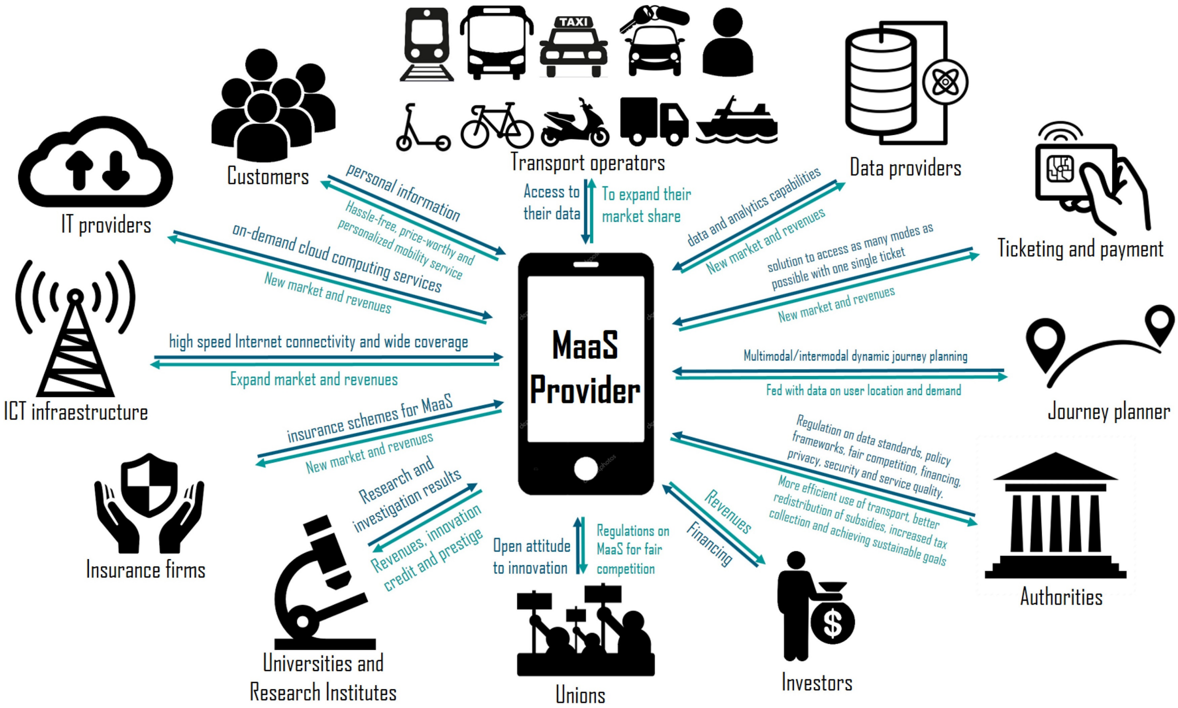

Figure 20: The interaction of Mobility as a Service components (adapted from [215]).

In the domain of smart cities, MaaS projects stand as strategic solutions, capable of identifying and adopting suitable business models tailored to specific applications or locales. These projects not only cater to diverse stakeholders but also foster user adoption while promoting environmental sustainability. Definitions of the "new"MaaS concept also include other significant elements, such as customer needs, personalized or tailored and comprehensive solutions, and user interfaces. MaaS outlines their main key components:

- **Information Technology Providers:** They offer on-demand cloud computing or fog computing services, enabling real-time transport planning, modelling and performance indicators, and environmental measures. This technology expansion translates into more efficient infrastructure utilization and innovative city development.

- **Customers:** They provide personal information, make reservations, payments, and give feedback. Despite payment and reservation being operations that occur on use-case specifics, the user interface typically facilitates information sharing and user feedback. In addition, customers expect hassle-free experiences, adjusted prices, and personalized services.

- **Operators:** These actors access their data and, at the same time, try to expand their market share. MaaS providers strive to on-board as many operators as possible to join their platform, ensuring a seamless integration of services.

- **Operations System:** MaaS facilitates various operations such as ticketing and payments, ensuring the availability of a high and varied number of services to the user. Consequently, it may attract new markets and revenues. Inversely, from these operations, the user can access a diverse range of services through a single ticket.

- **Data Providers:** MaaS platforms focus on matching individuals with suitable mobility options while expanding datasets over time to provide further details about both the individual and the transportation assets. Along with this expansion, operators and platforms envision a gradual and well-defined roadmap. They undertake a thorough evaluation of their current positions, guiding their subsequent actions, directing investing towards new markets, and capturing revenue opportunities.

- **Journey Planner:** This MaaS actor offers efficient multimodal/intermodal dynamic journey planning, utilizing user location and demand data. Mobility intelligence planning leverages user location data to analyze multiple factors such as population density, traffic patterns, trip parameters (origin, destination, distance, duration), driver intent, proximity to home/work or adjacent Points of Interest (POI)

- **Authorities:** They regulate data standards, and policy frameworks, increase accessibility and inclusiveness, improve service quality and reliability, and integrate tariffs. They also enable system-level optimization of mobility planning (and associated investments), real-time mobility flows, asset utilization, and achieve sustainable goals.

- **Investors:** MaaS platforms development relies on a "national"agenda, drawing attention to the concept and making it easier for start-ups to find investors and to convince mobility service providers to jump on the bandwagon. In other words, infrastructure and public policies should encourage people to opt out of traditional models of mobility. However, this migration to digital mobility platforms is only possible with players available to invest in public mobility solutions.

- **Regulators and Unions:** MaaS emerges as a user-centric, multi-modal alternative to conventional commuting practices and has harnessed the attention of the public and private sectors alike. However, regulators and unions require city authorities to recognize the balance of priorities in the new transportation mix and act as facilitators of partnerships, innovation, and uphold the interests of cities and citizens.

- **Academic Institutions:** Academic researchers have looked into the intricacies of MaaS, yielding valuable research findings that contribute to a better understanding of human mobility behaviours. Their insights aid policy makers in defining context-appropriate measures. In addition, their research can be published in journals, conferences, and media outlets, leading to revenue generations, innovation credit, and prestige.

- **Insurance Companies:** With many MaaS solutions available in cities around the world, insurance companies play a central role in the lives and mobility of transit users. They need to evolve to accommodate users' need for multimodality while also meeting the needs of service providers. Through their products, insurance companies can explore new markets and revenue streams.

- **ICT Infrastructure:** MaaS agglomerates heterogeneous displacement modes, physical infrastructure, and ICT tools. These components synergize to provide citizens to with high-speed Internet connectivity, extensive coverage, and efficient access to their destinations. Additionally, these infrastructures aim to create public value that in turn improves citizens' quality of life and contributes to market expansion and revenue growth.

These MaaS components offer a number of opportunities and challenges. They represent emerging human mobility models with considerations spanning technical, regulatory, and workforce domains. Mobility technology is also a space where insurance and actuarial sciences have an important role to play, contributing to safer mobility experiences. While it reduces driver-related mistakes, and the complexity of this evolving ecosystem requires careful management, the open market nature of MaaS providers offers cities intelligent possibilities for the future.

3.3 Citizen-centred approach to cities

As previously mentioned, studies on human mobility within urban areas are known to explore datasets containing thousands, and even millions, of data points. To address this challenge, conventional Machine Learning (ML) models (e.g., Deep Learning (DL), Decision Tree (DT), etc.) are commonly employed. Additionally, the advancements in parallel computing, and the training of Neural Network (NN) for classification and regression tasks have generally achieved superior results compared to other models.

On the other hand, the essence of the smart city concept lies in the established functional cycle of human engagement, particularly involving mobility, while fostering active participation within the cities. Numerous projects have used technology to enhance the quality of life inherent to this functional cycle, thereby moving closer to the optimal realization of a smart city. Below, we provide an overview of the issues and solutions pertaining to human mobility , enabled by Artificial Intelligence (AI) technology, that positions citizens as engaged and cooperative agents within smart cities.

3.3.1 Road traffic

SC prioritizes the secure and efficient transportation of their citizens from one point A to another. To achieve this, municipalities are increasingly turning to Internet of Things (IoT) development and implementing smart traffic solutions. These solutions encompass a variety of sensors and fetch Global Positioning System (GPS) data from drivers' smartphones to determine factors like the number, location and speed of vehicles [9]. Simultaneously, smart traffic lights connected to a cloud-based management platform enable the real-time monitoring of green light timings. system dynamically adjusts signal patterns based on the prevailing traffic conditions to prevent congestion. For example, consider Los Angeles, a city significantly impacted by traffic congestion. To address this challenge, Los Angeles has implemented a smart traffic solution to control traffic flow. Using road-surface sensors and closed-circuit television cameras, real-time traffic updates are relayed to a central traffic management platform. Additionally, the city is deploying a network of smart controllers to automatically make second-by-second traffic light adjustments, ensuring a swift response to evolving traffic conditions.

3.3.2 Public transport

The data from IoT sensors offer valuable insights into citizens' transportation habits. Public transportation operators can use this data to enhance the travel experience, ensuring higher levels of safety and punctuality. To carry out a more sophisticated analysis, smart public transport solutions can combine multiple data sources, such as ticket sales and real-time traffic updates. For instance, in London, some train operators employ predictive techniques to anticipate passenger loads on trains during journeys to and from the city. This predictive analysis combines the data from ticket sales, movement sensors, and Closed-circuit Television (CCTV) cameras positioned along the platform. By analyzing this data, train operators can predict how passenger loads will be distributed across each train car.

3.3.3 Environment

Mobility-driven smart city solutions allow tracking parameters critical for a sustainable environment, ensuring their maintenance at an optimal level. For example, in the context of water quality monitoring, a city can deploy a network of sensors across the water grid and connect them to a cloud-based management platform. Similarly, the surveillance of air quality is another pertinent application. This involves deploying a network of sensors along busy roads and surrounding industrial plants. These sensors gather data on the amount of CO, nitrogen and sulphur oxides, while the central cloud-based platform analyzes and visualizes the data derived from these sensors. This allows platform users to access an air quality map, which in turn aids in identifying areas with critical air pollution levels. Using this information, recommendations for citizens can be formulated to address the issue.

3.3.4 Public safety

To enhance public safety, smart city technologies offer real-time monitoring, analytics, and decision-making tools. By combining data from acoustic sensors and CCTV cameras dispersed throughout the city, alongside insights garnered from social media feeds, innovative public safety solutions are capable of predicting potential crime scenes. This will allow the police to stop potential perpetrators or successfully track them. For example, over 90 cities across the United States have adopted gunshot detection systems. These systems rely on data collected by microphones that transmit to a cloud-based platform. This platform then scrutinizes the audio data to identify the distinct sound pattern of a gunshot. Furthermore, the platform calculates the time it takes for the sound to reach each microphone, thus estimating the firearm's location. Upon the detection of a gunshot and its corresponding location, cloud-based software promptly dispatches an alert to the police via a mobile app.

3.3.5 Telecommunication Industry

The telecommunications industry stands as one of the most emerging industries providing various services such as fax, pagers, cellular phones, internet messengers, image transmission, e-mail, web data transfer, and more. Due to the evolution of computer and communication technologies, this industry is rapidly expanding. Consequently, the role of AI has become very important in comprehending and optimizing the business.

In this context, AI helps in identifying telecommunication patterns, detecting fraudulent activities, optimizing resource use, and improving the quality of service. The followig examples illustrate how AI improves telecommunication services:

- Multidimensional Analysis of Telecommunication data;
- Fraudulent pattern analysis;
- Identification of unusual patterns;
- Multidimensional association and sequential patterns analysis;
- Mobile Telecommunication services;
- Use of visualization tools in telecommunication data analysis.

3.3.6 Smart home

The concept of a smart home highlights the interactions among the various entities outlined earlier at a high level. Consider the following scenario: Person A buys a new smartphone and takes it home, plugging it into a power source. The smartphone automatically detects the presence of available Wi-Fi in the house. Further, assuming Person A is a registered user, the smartphone communicates with a Service Provider (SP), relaying information about its presence by providing information such as the available sensors (e.g. temperature sensors). Next, the SP communicates with Person A to check whether he would like to share the smartphone's sensor data with the cloud.

Person A has the authority to decide which sensors to make public, the eligible consumers allowed to bid for the data, and the desired compensation (be it fees or any other offer). Later on, Person A, from his mobile application, expresses interest in having access to the light sensor and solar energy sensor metrics. The temperature sensor, on the other hand, gauges heat or coldness generated by his presence, adapting to the temperature conditions of the home. In parallel with these processes, an energy company continually monitors Person A's real-time energy consumption. This data is made accessible via a web platform, readily available on Person A's smartphone, allowing him to track energy usage and associated costs in real-time.

This scenario also has a sensor owner, wherein Person A embodies the layer of sensors and sensor owners. Furthermore, in the realm of ownership categorization, Person A represents the Personal and Households scheme. The energy company, taking on the role of a consumer of sensor data, represents another layer. A SP who enables communication and transactions between Person A and the energy company is responsible for matching the sensor owner expectations with the requirements of sensor data consumers.

3.3.7 Marketing/Retail

DL models help marketing companies build models based on historical data to predict the responsiveness of individuals to new marketing campaigns, be it through direct mail or online marketing campaign, for example. Through this prediction, marketers can tailor their strategies, ensuring that they effectively promote profitable products to targeted customers with high satisfaction. Retail companies, too, reap substantial rewards from these models. Employing market basket analysis, stores can have an appropriate production arrangement to facilitate the purchase of frequently paired items, enhancing costumer convenience. In addition, such models enable retail companies to offer a certain discount for particular products, enticing customers. Here a comprehensive list of AI applications within the retail industry:

- Design and build of data warehouses;
- Multidimensional analysis of sales, customers, products, time and region;
- Analysis of the effectiveness of sales campaigns;
- Customer retention;

- Product recommendation and cross-referencing of items.

To conclude, as technology grows, the number of users and devices, each employing different types of communication protocols and technologies, also grows exponentially, and there is a need to provide interaction for an unbounded number of entities with significant differences in interaction patterns. Accommodating this multifaceted interaction is paramount, as it profoundly influences numerous aspects of SC technologies and infrastructure. The fundamental concepts and ideals of SC are shared with a large number of business prospects, fostering extensive growth potential.

3.4 Summary

SC projects offer solutions tailored to urban planning and management, with a specific focus on enhancing human mobility. However, they need to complement its definition with architecture that helps us to identify the main components and interconnections within a city. Furthermore, the environmental repercussions of construction projects underscore the importance of fostering, a participatory approach among citizens. That, together with other individual contributions with comprehensive analyses, makes it possible to have a robust set of decision-making processes, ultimately leading to huge benefits for the overall population's mobility.

The related work and projects involve different levels of application, spanning from city design/planning to the analysis of citizens' behaviours. Although these levels differ in conception, they also share some tools or concepts. For example, the concept of a smart city is based on the aggregation of data from multi-sources, encompassing nearly every facet of an individual's surroundings. In addition, the utilization potential of this data can manifest in multiple modes of execution and application strategies.

The generated data from human interactions with several available technology tools spread across urban settings harbors immense potential for empowering citizens. In fact, this abundance of data production, whether intentional or inadvertent, should be harnessed by cities. On the other hand, this data exploitation should be accompanied by robust measures to address privacy, safety and security concerns. Governing authorities must ensure the implementation of human mobility projects with a focus on multiple layers of security. Striking a balance between accessibility and safeguarding against misuse is paramount. Therefore, the stakeholders should adopt the best and most current practices to ensure a continuous improvement of the human mobility solutions.

Chapter 4

Innovation and Research

This chapter aims to study the scalability of cities in order to adopt solutions for human mobility. While previous chapters primarily examined existing theoretical and practical studies on technological tools supporting Smart City (SC), this phase of the document shifts its focus towards presenting the different stages and topics that constitute the architecture of the proposed human mobility service.

Beginning with an exploration of different urban phases, the incorporation of technology has significantly altered the dynamics of urban spaces. Additionally, it's imperative to understand the city-type to identify the most suitable context for the implementing our technological solution for human mobility. Although the proposed solution is idealized for less favourable environments, such as cities with lower population density, or other factors, limited telecommunications infrastructure or lack of stimuli (i.e., gamification) that promote the active participation of the citizen, the inherent conditions that cities offer play a pivotal role in its successful implementation.

Turning a city into an "intelligent"entity requires an examination of the architectural layers constituting a SC. The functionality of this architecture also depends on a dense and scalable heterogeneous network of diverse sensors and devices, facilitating data capture from pedestrians across various urban locations. Notably, existing projects already in the field offer insights into the advantages and challenges of implementing smart cities projects.

Since the proposed architecture extends beyond a mere infrastructure; we also want to make it inclusive and transformative, aiming to foster inclusivity in smart city prototypes. Inspired by this narrative, we must extend this new human mobility solution to the allocation of individuals. With citizens at the core of this solution, ease of participation plays a crucial role in its inclusiveness. Therefore, a comprehensive understanding of participation and inclusion levels and their key elements within a SC is imperative.

Alongside this analysis, the calculation of metrics and forecasting algorithms to comprehend trends and anticipate human mobility behaviour scenarios based on massive citizen participation are also relevant to our investigation. Furthermore, the integration of Explainable Artificial Intelligence (XAI) algorithms is seen as beneficial for ensuring the interpretability of predictions and aiding target audiences—governmental and non-governmental authorities, companies, and citizens alike . In short, citizen participation in this type of project must be incentivized and rewarded through clear information, easy acquisition and interpretation and, above all, security.

4.1 Human Mobility Infrastructures

Human mobility is one of the next challenges for cities. As we have already mentioned in several sections of this work, cities draw individuals like magnets, resulting in an increasing human presence (whether to live or simply to visit) in urban areas. This surge in urban inhabitants needs the adoption of unprecedent policies and investments. However, it's crucial to recognize that this growth, while carrying direct or indirect implications on cities, should not be automatically deemed negative. Instead, it presents a unique opportunity to implement infrastructures that optimize resources within urban environments, addressing challenges like traffic and security. On the other hand, leveraging this population density calls for an increased role of citizens within communities, fostering their active participation. Therefore, it seems clear to us that cities must be designed taking into account the virtuous acceleration of the digital transition while embracing the dynamic and continual involvement of residents and visitors alike. The evolution of SC has transpired through distinct phases since the 1990s [126]:

- **Smart City 1.0:** technology companies led and encouraged the adoption and implementation of new solutions;
- **Smart City 2.0:** local administrations used technological solutions as a way to improve sustainability and the quality of citizens' lives;
- **Smart City 3.0:** models of co-creation and collaborative approaches have emerged, uniting policy makers and citizens to find the best strategy and solutions for a common vision;
- **Smart City 4.0:** a city that capitalizes on the opportunities offered by sustainable development through technological tools.

During the initial phase of the digitization of cities, solutions were rudimentary and inadequately integrated with their surroundings. Over time, Even as cities evolved, integration issues persisted, lacking interoperability and third-party integration. Furthermore, they did not promote citizen participation. These shortcomings prompted researchers, focusing on the study of cities [3, 90]], to advocate for government policies supporting interoperable infrastructure and open standards in the most diverse areas (e.g., road mobility, human mobility, safety). Consequently, in this new phase, attention shifted to the relevant role of the citizen in building smarter cities [110, 142]. In the third stage, the citizen is instilled with the will to co-create, marking a pivotal shift in smart city strategies where citizen participation has become essential for achieving success [184]. However, merely involving citizens collaboratively is no longer enough, it is necessary to turn cities into sustainable spaces [46].

In the context of human mobility, several international entities have provided guidelines. The European Union's Europe 2020 strategy focused on three priorities [44]: sustainable growth (low-carbon economy), smart growth (education, research, and innovation), and inclusive growth (jobs and wealth). In the context of human mobility in cities, there are several ongoing initiatives. One of them involves implementing a strategy for smart and sustainable mobility in order to enhance the sustainability, intelligence, and healthiness of urban and interurban transportation networks. The European Commission (EC) is proposing revised transport guidelines for this purpose. The proposal envisions major European cities establishing public transport connection nodes and formulating urban sustainability plans by 2025, based on a set of criteria [64]. Similarly, internationally renowned entities such as the International Telecommunication Union, International Electrotechnical Commission or International Organization for Standardization set up a "task force"to align the Information and Communication Technology (ICT) with urban sustainability. This initiative includes implementing key performance indicators, encompassing human mobility [198]. Several frameworks pertaining to human mobility have been developed to make cities smarter and can be found in literature [11, 12, 111, 172, 211, 225]. However, it is important to study the best practices for establishing an infrastructure for continuous human mobility monitoring. Such an infrastructure would contribute a standardized basis for policy decisions, fostering the progress of cities without jeopardizing privacy and citizen security.

4.1.1 Functional Smart City architecture

In the digital age, technologies have improved people's lives, serving as tools for both leisure and work. This merger has made services more available to citizens. On the other hand, the influx of populations into cities requires a constant restructuring of communication infrastructures due to the growing interconnectivity of diverse sensors. Therefore, the evolution of ICT has led to the proliferation of devices, transforming the physical world itself into a knowledge and information-sharing system.

In human mobility, the incorporation of Internet of Things (IoT) and communication networks within cities has led to various infrastructure proposals. These infrastructures cater to pedestrians and runners by providing high- quality applications/services. In addition to this, government authorities are invested in making the infrastructure not only appealing but also accessible any time and anywhere thereby fostering active citizen participation and aligning with the attributes of a smart city 4.0. In this context, we propose the following functional organization for the construction of smart cities:

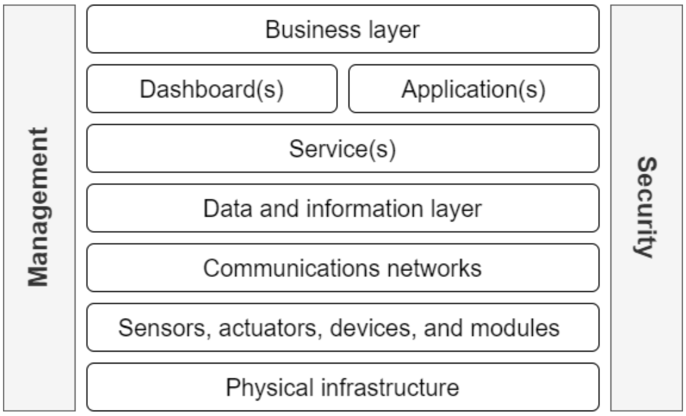

Figure 21: Functional layers of a Smart City.

Figure 21 illustrates the functional organization of a smart city, demonstrating the importance of ICT in each layer. Amidst ongoing discussions around this architectural framework, the data and information layer takes a prominent stance. Research work, demonstrations, and projects within this layer strive to address concerns of privacy, reliability, information sharing, standardization, Application Programming Interface (API), and commercialization. However, preceding these endeavours, it is necessary to facilitate comprehensive understanding among various stakeholders regarding the data they intend to collect, interpret, enrich, analyze, and effectively use in real-world scenarios.

While acknowledging the significance of the data, the layer dedicated to communication networks among digital objects also stands out. This layer is important to ensuring that the different layers are interconnected, through Local Area Network or Wide Area Network. This constitutes the main conduit for information transmission and serves as the central structure for modelling applications and services that can communicate with one or several networks simultaneously. It defines how data should be used within smart city tools. In other words, the implementation of ICT, supported by a scalable network, facilitates the modulation of the construction of smart cities.

4.1.2 Scalable Network and Internet of Things Infrastructures

To comprehensively address the study of human mobility within the functional architecture of a smart city, it becomes necessary to add a detailed design of a structure. This structure is engineered to accommodate a set of sensors or digital objects that interconnect and share data in real-time via a responsive communication network. Furthermore, the architecture presented in Section 4.1.1, which relies on the capabilities of ICT, is not intended as a single solution; rather, it provides an initial perspective of a viable solution needed to make a change for an application that uses the latest tools and technologies. In essence, with the incorporation of this infrastructure, the intention is for stakeholders to maximize the potential afforded by the sensor network at their disposal. Through the data harnessed by this network, solutions that acter to the community's needs are envisaged. Therefore, the authors of a given article [120] propose an IoT architecture (Figure 22) that helps to understand how pedestrian-oriented applications are supported in smart cities:

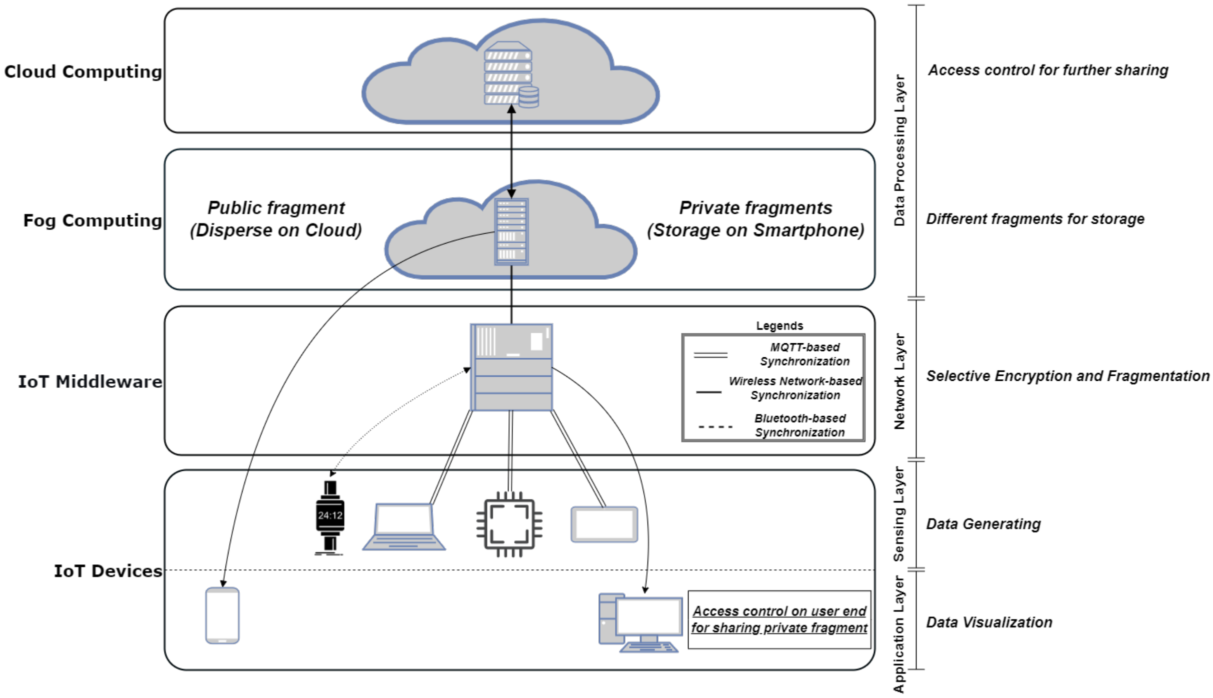

Figure 22: Architecture proposal for pedestrian-oriented applications.

Within the IoT Device layer, two distinct sub-layers operate: the Sensing Layer and the Application Layer. In the "sensing layer", components like IoT Device Data and Commands for IoT Devices are encompassed. They can be independent or connected to a system that aggregates them (i.e., IoT Middleware). Human mobility sensing entails a range of sensor types, such as smartphones, smartwatches, body sensors (e.g., temperature, movement), or other wireless and/or Bluetooth-enabled devices that facilitate synchronization among mobile devices and/or transfer data. On the other hand, the Application Layer is responsible for capturing the physical environment and building applications that can be accessed through various digital platforms by pedestrians. This layer is the Human-Technology Interaction illustrated in the scheme drawn in Section 1.2.2.

The IoT Middleware layer incorporates the Network Layer and allows users to control the devices individually and quickly. Although not explicitly described in the proposed architecture, the devices can communicate through different protocols (i.e., CoAP, MQTT and HTTP). REST API can facilitate the integration of other applications with the platform. Additionally, this layer serves as a central entity to support applications that use Fog and Cloud Computing architectures.

Figure 22 also shows the "Dataprocessing Layer"layer. The role of this layer is to use data collected from IoT sensors, and other data sources that have been sent to Fog or Cloud services. Storage provisions within this segment of the architecture facilitate the development of interfaces for pedestrian to receive information, recommendations, and suggestions. Through these interfaces, users can make informed decisions or request additional information from the application.

The organization of this architecture aims that applications or services are supported by an open data model. That is, citizens must have access to applications that use real-time data from different sources, which in turn improve the overall effectiveness of city services. By opening up data and sharing information, both pedestrians and authorities can gain insights and address issues proactively, mitigating their potential escalation. Therefore, the foundation of a smart city necessitates an agile and scalable connection to integrate different data sources, enabling prompt responses to requests and/or phenomena.

4.1.3 Smart City Architecture Models

Naturally, the evolution of communication technologies has prompted cities to adapt and keep citizens connected with faster access to services. The architecture that we will propose in Chapter 5 will be inspired by existing architectures and, based on the results, will allow us to better serve the different stakeholders (from the pedestrian to the ruler) within the city more effectively. One such is the LinkKiosk initiative [31, 118, 192] in New York City. These pluggable terminals provide citywide connectivity, ensuring citizens' mobile devices remain connected. In addition to this reason, it is important for this governmental entity to connect, directly or indirectly, the environment around the user in order to become aware of ongoing phenomena and anticipate the best political decisions.

With the connectivity of LinkKiosks, a pedestrian moving within the city stays up-to-date through the technological services provided by local authorities. In the context of human mobility, this showcases how a terminal can serve as a data source about its surroundings, its contribution depending on interactions with nearby tourists or citizens throughout the day. At the service level, each LinkKiosks offers a Wi-Fi hotspot with approximately a 50-meter radius, delivering internet speeds of up to 1 gigabit per second. Additionally, these kiosks provide national calling, a 911 button, mobile device charging ports, and access to maps, 311 services, and more, thereby increasing fast broadband access in many neighbourhoods of the city.

Another project that offers insights relevant to your thesis involves the utilization of call logs. Harvard researchers used location or movement information derived from telecommunications data, such as Call Data Records or x-Data Records (the latter generated with a mobile device connected to the Internet; smartphone application data), vehicle Global Positioning System (GPS) devices, Bluetooth exchanges, and social network data with location information, to study population movement and the evolution of the pandemic [52]. Subsequently, the survey results were made public, ranking neighbourhoods or districts based on their mobility patterns.

In the sensor domain, a project focused on the study of human mobility, such as PASMO project in Aveiro, Portugal, stands out. PASMO, an Open Platform for the development and experimentation of Mobility Solutions, covered road and urban areas with OVER 15 km of roads. This equipment included fixed radars and cameras, for traffic counting and classification, road and on-board communication stations, various sensors (i.e., parking lot, meteorological stations), LoRa gateways, public Wi-Fi, and Big Data Platform for IoT data aggregation [67]. To facilitate access to historical and real-time data from parking sensors and radars, an API was made available for public access. The project continues to improve, integrating new sensors, applications, and access to an experimental 5G network. Currently, the PASMO platform already supports other projects in the field of human mobility.

4.2 Participatory Sensing, Metrics and Community Patterns

As already mentioned, the study of human mobility within a SC depends on the production and analysis of data sourced from diverse sensors that detect various types of information, from weather conditions to human activities in public spaces. However, the concept of a "Smart City"can be extended to the allocation of participation civic resources through the use of digital devices [138]. While sensors tell part of the story, citizens tell the other part of the narrative that shapes a city: they know the real needs of the city and the emergence of certain phenomena. Sharing this valuable information is essential for designing a more human-centric, pedestrian-friendly urban environment, consequently enhancing the quality of life in the community [23, 124, 207, 213]. The perspective of citizens and their insights, combined with technological advancements, must be part of the process that allows them to express their concerns or needs concerning smart city initiatives. Therefore, in the study of human mobility within urban areas and with the purpose of fostering civic participation, Goldman categorized sensory participation into three different types:

1. **Design and research collective:** In this category, citizens are not just data collectors, but active participants throughout the investigation. They participate in the different stages of the project, contributing to the definition of the objective(s), sensing and analysis of the samples, and make decisions about data processing and management;

2. **Community contribution:** This category involves citizens actively participating in the data collection phase, but not necessarily being part of the group that defines the objective(s) of the investigation and manages the sensors and data infrastructure;

3. **Personal use:** Here, the focus is on individual self-discovery or constant improvement: the citizen engages with the technology primarily for personal benefits.

In this investigation the individual is considered an integral part of a larger whole – an active agent within a city. That is, based on the categorization proposed by Goldman, our study aims at community contribution, which means citizens can play a pivotal role in collecting data about human mobility.. Consequently, collectively gathered data can then be made accessible to, the municipality, driving city innovation. This approach transforms citizens into key contributors to a better understanding of human mobility patterns, enabling the creation of data-driven solutions for human mobility that benefit the entire community.

4.2.1 Key Principles to Community Contribution

As previously discussed, in the context of human mobility research, the SC foresees the creation of IoT and big data-connected infrastructures that offer benefits to local authorities, citizens, and companies. In addition, this approach plays an important role in addressing the ecological impact of urban centers, as optimized resources use can cater to pedestrians' preferences and lead to cost reductions. However, the dynamics of smart cities, spanning the areas of design, organizational conception and community engagement [30, 155], consistently incorporate sensory participation grounded in five core principles:

- **Active Citizens:** users work collectively for the common good;
- **Building Capacities:** promoting the digital transformation of cities involves cultivating active participation and digital transition among citizens;
- **Build infrastructure:** the construction and availability of infrastructure allows for citizen participation and impact;
- **Community Engagement:** a community-centric approach facilitates the implementation of city solutions and acts as a catalyst for large-scale transactions;
- **Encourage creation and innovation:** a smart city must have a dynamic community that can actively contribute to innovation and creativity.

Cities that implement these principles progressively transition into as environment characterized by cutting-edge technology and dynamic citizens. In it, authorities will be able to provide free access to public data sources, fostering city innovation. Collaborative partnerships with private entities that share a mutual vision for the future of cities enable the development of data-driven services and applications tailored to the city's requirements and its citizens [63, 217]. Cities that harness collective involvement from diverse stakeholders adopt a strategic approach to smart city development, fostering improved human mobility quality. In the next sections, we will explore solutions for cities arising from the collective contributions of citizens.

4.2.1.1 Participatory Sensing: a crowd-sensing paradigm

The concept of SC envisions the creation of urban spaces that are not only more environmentally friendly and healthier but also capable of accommodating the rapid population growth [4]. Currently, different stakeholders are tasked with understanding how people use the city, creating better access to services through digital solutions, providing data-driven operations and supervision, and building detailed maps, such as those detailing pedestrian-perceived temperatures across different city zones. Therefore, with the objectives of smart cities in mind, a group of researchers from the University of Minho proposes a pedestrian-oriented architecture approach. This innovative framework allows citizens to actively engage in measuring both body and environmental temperatures through a dedicated mobile application [171]. Promoting people's participation in providing relevant data for urban planning services or analysis of human behaviour patterns forms an integral part of this project. The project's driving force was to showcase the potential of the participatory sensing paradigm—a practical example of urban sensing—in illustrating the ambient temperature in different areas of the city. Hence, the features of this paradigm can hold valuable insights for the crowd sensing needed for this scientific work.

4.2.1.2 Understanding dynamic urban areas from human behaviours perspective

Supporting "quality of life"implies a particular focus on citizens and their surroundings. Nonetheless, most smart city sensors, along with the quantitative and abstract data they produce, only tell part of the story of what is being measured [70]. The data collected by these sensors fails to elucidate how individuals perceive certain events in cities or the reason for the concentration of people in a certain area of the city; it merely informs us of their presence and reasons for being there. Therefore, this citizen-centered perception, facilitated through participatory sensing services or intelligent applications, facilitates the gathering of data that enhances the comprehension of ongoing events in their surroundings.

Through the collection of data concerning citizens' perceptions, several research projects have been carried out so that the results are the subject of discussion on the phenomena of human mobility. In this project undertaken by Synthetic Intelligence group (ISLab), at Centro ALGORITMI, University of Minho, researchers make predictions of time series related to individuals' sentiments in a certain Points of Interest (POI) of New York City by employing Machine Learning (ML) models [169]. The extensive results obtained from the dataset collected from LinkNYC Kiosk devices illustrate that the Long Short-Term Memory (LSTM) model significantly enriches the understanding of population sentiments within the analyzed location.

The same group of researchers improved their investigation with additional indicators: well-being and comfort. In their publication [173], they leverage smart urban development strategies and digital tools to magnify geolocalized social issues. This research incorporates data not only from an external environment (i.e., LinkNYC Kiosk), but also from an internal environment, the Metropolitan Transportation Authority (MTA) of New York City. The integration of these datasets aims to examine the influence of comfort and well-being indicators on human mobility. By comparing predictions from Statistical Method (SM) and Deep Learning (DL) models using aggregated mobile data, the study reveals that Deep Neural Network (DNN) models exhibit a reduced maximum forecast error. These projects collectively illustrate that the use of technological resources, coupled with the direct or indirect contributions of citizens, help to overcome the challenges that local or governmental authorities face, whether they pertain to social, environmental, or mobility-related concerns.

4.2.2 Human Mobility Metrics Overview

The data collected from participatory citizens serves as a valuable source for measuring the performance of a smart city. Along-side all the indicators enumerated in Table 1, metrics can be analyzed to address specific urban challenges or contribute to city development [117, 175, 219]. The various metrics can be examined within several contexts, detailed in the following segments:

- **Individual Mobility:** Understanding the dynamics of individual human movements is important to understand the relationship between the effort individuals invest in travelling and the underlying purpose of those journeys. This involves several mechanisms that interact across different spatial and temporal scales [177]. Many of these mechanisms have been explained by the "Lévy Flight"model, as formulated by the French mathematician Paul Pierre Lévy. This model has three fundamental properties: (i) similar to the Gaussian law, this model is formed by the sum of random variables; (ii) presents fluctuation processes characterized by bursts or large outliers; (iii) meets the description of random fractal processes. Therefore, this underscores that individual mobility motivations may differ due to real issues, objective circumstances, or social and environmental implications.

- **Collective Mobility:** The study of large agglomerations of people combines aspects of long-range human mobility with location-specific pedestrian dynamics [221]. Recently, some researchers have used the Origin-Destination (O-D) matrix to investigate group mobility in a given area. Each O-D matrix corresponds to a cell associated with a trip between an origin and a destination. The density of these trips directly correlates with the density of routes. With this method, it is possible to make full use of spatio-temporal information to predict pedestrian flow based on various mobility patterns.

Within the realm of smart city metrics, georeferencing data collected from millions of mobile devices assumes a prominent role. Benefitting from their widespread adoption, these metrics can be categorized based on their geographic attributes (global vs. local) and network characteristics (returner vs. explorer), yielding four distinct categories.

In Figure 23, the categories of *Global returner* and *Global explorer* depict data obtained from geographically remote and dispersed sources. While in the first category, the mobility of individuals is based on international travel (the destination is outside the country), in the second category, international travel is still a factor, but the places visited are distributed in a more decentralized manner. Conversely, *Local Returner* and *Local Explorer* contrast with the previous categories, as the georeferenced information is not geographically remote or dispersed. The *Local Returner* category typically involves individuals who visit nearby locations or a frequently visited center. On the other hand, the *Local Explorer* category comprises places that are visited more frequently than others, despite being relatively far from each other, resulting in significant travel distances. Furthermore, the places visited in this category are less centralized.

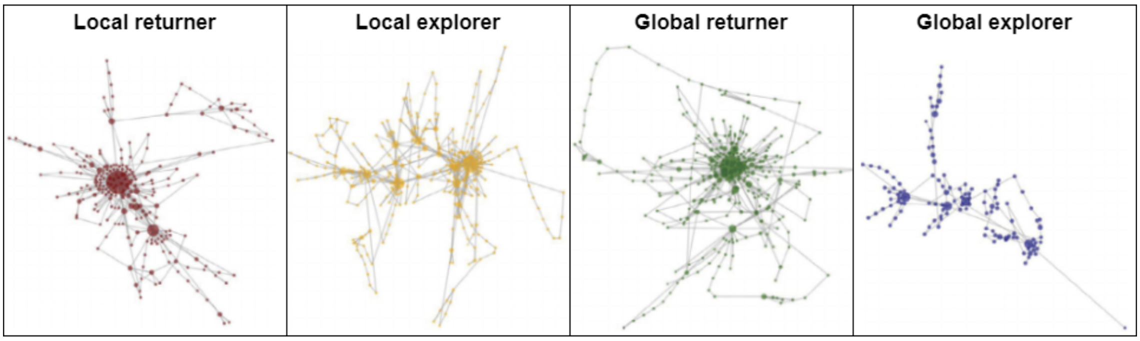

Figure 23: Network visualisation of four typical individual trajectories (adapted from [114]).

Following each metric outlined in Section 2.3.3.2, a set of solutions or policies were preventively adopted. That is, many institutions, organizations, and political authorities have actively engaged with these potential solutions. Notably, a project prepared by a group of researchers from the University of Minho [174], associated with the ISLab laboratory has presented a comprehensive analysis. In its initial phase, this project introduces metrics commonly associated with human mobility. Subsequently, it compares additional metrics such as resilience, displacement, interval, and duration from different types of data (i.e., CDRs, GPS and Social Media) collected by two individuals. Another project that has been working towards the challenges of human mobility is the C2SMART Center at New York University [108]. By harnessing anonymous spatial and temporal information from mobile devices, this centre strives to estimate real-time urban populations. Using metrics, they differentiate between various types of populations, assess individual or collective mobility patterns, and quantify historically challenging-to-measure human activities. Finally, [202] quantifies the effect of a new metric on predictability and evaluate the routine of a predictably typical individual. In doing so, they employ previously proposed metrics, alongside this new metric, to understand factors influencing the predictability of an individual's routine. Although with different purposes, to be successful, all these projects take into account the needs of all actors, especially those of citizens. As any approach aiming to improve human mobility within smart cities must be centered on the individual and transition towards collective good.

4.2.3 Representing Human Mobility Patterns

Alongside metrics, the study of human mobility patterns constitutes another field of study that requires citizen initiative to create and implement an ecosystem of smart city solutions, ultimately generating added value for the community. This area is focused on the application of forecasting models to address urban challenges.

4.2.3.1 Deep Learning and Statistical Models

As previously discussed in Section 2.4, the exploration of such patterns necessitates the creation of models capable of forecasting individual whereabouts at specific instances, along with other individual or collective phenomena. This facilitates the extraction of valuable insights, with practical applications in different contexts such as traffic management, disease contagion control, and recommendation systems in advertising and marketing [40, 133, 175, 218]. Crucially, these models operate on the premise that the abundance of data on pedestrian movement directly corresponds to the precision of the predictions. Therefore, we use a set of algorithms based on Neural Network (NN) to carry out an experimental study of human mobility patterns:

- **Long Short-Term Memory (LSTM):** This model has a specific architecture from Recurrent Neural Network developed for sequential data forecasting. It excels in analyzing time series data over extended periods and defines optimal time intervals. The training of the model can be carried out with be univariate or multivariate data.
- **Convolutional Neural Network (CNN):** This model is used to capture spatial patterns in data, especially in crowd flow prediction where the distribution of people in a geographic region is represented as an image.
- **Convolutional Neural Network-Long Short Term memory (CNN-LSTM):** This model combines the ability of CNN to process population movement maps and the ability of LSTM to handle sequences. This amalgamation facilitates the swift calculation of the downstream flow through maps and the upstream flow sequence.
- **Bidirectional LSTM (BLSTM):** This model applies the concept of incremental learning and deals with long-range contextual processing. In addition, compared to Unidirectional LSTM, it tends to show better accuracy results in recognizing human activity [169].
- **Stacked LSTM (SLSTM):** This model acts as a reliable technique in solving prediction problems in human mobility. Its architecture consists of multiple LSTM layers, where each layer produces string outputs, rather than single-valued outputs, which are then employed by subsequent LSTM

layers. This model is expected to employ a more sophisticated decoder than just using a simple LSTM model.

In addition to DL algorithms, the prediction of human mobility has also attracted great attention and efforts to investigate the existence of statistical methods. Statistical methods prioritize the inference and exploration of individual mobility patterns. In addition to these, we outline different types of models:

- **Autoregressive Integrated Moving Average (ARIMA)** - This model aims to describe the autocorrelations within the data, representing the degree of similarity between a given time series and a lagged version of itself in successive time intervals. In short, it fits a regression model to a time series to predict new values in human mobility.

- **Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX)** - A variant of ARIMA, this model incorporates additional variables into the forecasting process. The effects of calendar changes are one of the variables that are often used in modelling the prediction of population movements.

- **Seasonal Autoregressive Integrated Moving Average (SARIMA)** - Differing from the ARIMA model, this one takes into account the influence of seasonal trends in many time series data. SARIMA is an extension of ARIMA that supports the modelling of time series with the seasonal component, making it advanced an evolution of the grouping of the various time series models.

- **Seasonal Autoregressive Integrated Moving Average with Exogenous factors (SARIMAX)** - The SARIMAX model is a SARIMA model with external variables. External variables can be modelled by a multilinear regression equation.

In summary, ARIMA stands for Autoregressive Integrated Moving Average, while the *"S"* in SARIMA stands for seasonal. Therefore, the ARIMA model can be either seasonal, in which case it is a SARIMA model, or non-seasonal, in which case it is an ARIMA model, depending on whether it accounts for seasonal trends.

The biggest difference between ML and Statistical models is their purpose. While ML models are designed to provide accurate predictions without requiring explicit programming, whereas statistical models are used to find and explain relationships between variables. Furthermore, while ML models can provide better predictions, they are often more difficult to understand and interpret. In contrast, statistical modelling is generally more accessible in terms of interpretation and, identifying relationships between variables.

4.2.3.2 Model Evaluation Metrics

To measure the impact of SM and ML models on the training of human mobility data, a set of experiments can be carried out. Based on their results we can analyze which method is most effective for predicting the demand of people's movement. However, the accuracy comparison can be evaluated through the following criteria:

- **Mean Squared Error (MSE)** - This model (Equation 18) measures the average of the squares of the errors, that is, the average squared difference between the estimated values and the actual value ion human mobility data.

$$\frac{1}{N} \times \sum_{i=1}^{N} (ot_i^{t+1} - \hat{ot}_i^{t+1})^2 \tag{18}$$

  where ot_i^{t+1} is the i^{th} observed value, $ot_i^{t+1})^2$ is the corresponding i^{th} predicted value and n the number of observations. The MSE is a measure of the quality of an estimator—it is always non-negative, and values closer to zero are better. However, it is sensitive to outliers.

- **Mean Absolute Error (MAE)** - This is a robust metric. As defined in Equation 19, it involves summing the magnitudes (absolute values) of the errors to obtain the "total error"when evaluating the predictive performance of a single model:

$$MAE = \frac{1}{T} \sum_{t=1}^{T} |\ln(y_t + 1) - \ln(\widehat{y_t} + 1)| \tag{19}$$

  where y_t and $\widehat{y_t}$ are the observed and average counts for forecast month t. The strength of these approaches is that MAE will be the same, regardless of magnitude, as long as the context variables are the same. This is an important characteristic, given the differences, for example, in the population size of a city and the real-time count of individuals between different urban zones.

- **Root Mean Square Error (RMSE)** - This metric (Equation 20) determines the similarity between two sets. It is similar to the MAE metric, except that each absolute error is squared before being summed and the final result is square-rooted before being returned.

$$rmse = \sqrt{(\frac{1}{n}) \sum_{i=1}^{n} (y_i - x_i)^2} \tag{20}$$

  where $\sum$ is a fancy symbol that means "sum", y_i is the predicted value for the i^{th} observation in the dataset, x_i is the observed value for the i^{th} observation in the dataset and n is the sample size

These three metrics for analyzing the performance of a ML model have different characteristics, making some more suitable for a loss regression context than others. That is MSE and RMSE penalize large forecast errors more compared to MAE. However, the RMSE metric is more used than the MSE to evaluate the performance of the regression model with other random models, as it has the same units as the dependent variable (Y-axis). In turn, MSE metric facilitates the execution of mathematical operations compared to MAE. Therefore, in many forecasting models, RMSE is used as the default metric to calculate Loss Function, despite being more difficult to interpret than MAE. In short, RMSE quantifies how well a linear regression model fits a dataset. In addition, it tells you how well a model can predict the value of a response variable in absolute terms.

4.3 Prediction Explainability on Pedestrian Environments

XAI technique is increasingly adopted by researchers due to the challenge of explaining the reasoning behind the predictions generated by ML models. In the same sense, many studies have been carried out to interpret these prediction processes and validate the reliability of the results on human mobility [113, 170, 187]. That is, it is important to emphasize that confidence in the results is an important factor for pedestrian acceptance of the recommendations provided by AI applications managed by authorities that govern Smart Cities.

Furthermore, when AI models are employed to predict human mobility and seamlessly integrated into citizens' daily lives, transparency becomes paramount. XAI techniques play a pivotal role in shedding light on the intricacies of human mobility patterns. An AI model for predicting population movements that is comprehensible and transparent not only instills confidence in individuals but also encourages their active participation in refining the intelligent recommendation system. Briefly, ethically what is expected between AI and XAI in the context of human mobility is:

- **Intelligibility:** The user must understand the result of the AI presented to him without any need for intricate explanations of its internal architecture or processing methods [201].
- **Comprehensibility:** Knowledge learned by the ML algorithm must be interpretable by humans. In other words, results generated from "black-boxes"should align semantically and structurally to what a human observer would deduce from the same data [79]. In addition, both quantitative and qualitative outcomes must be presented in simple and natural language.
- **Interpretability:** If an individual wants transparency and understands exactly how the model is generating predictions, he needs to look at the internal structure of the AI/ML methods. This includes model weights and features translated into terms that humans can understand [104].

- **Explainability:** As models become more complex, the significance of their explainability grows, even though it might come with a trade-off in terms of performance. This is particularly relevant when individuals seek a comprehensive comprehension of the model's behaviour. Explainability is facilitated through an interface between humans and the decision-maker. That is, the human needs an additional method/technique to be able to examine the ”black-box”and understand how the model works [73].

- **Transparency:** In AI a model is only transparent to the user if it is understandable [61]. However, as there are different degrees of comprehensibility, as such transparent models can be divided into three categories:

  1. **System explainability to augment trust and decision-making:** Systems have the challenge of presenting model outputs in an intuitive way for individuals not familiar with ML models. Although many model interpretation approaches (some mentioned in Section 2.5.2) may identify significant features, but these might lack organization or intuitiveness. To deal with these challenges, a model explainer should cater to pedestrians, offering insights that reflect the logic behind the model's predictions. This approach can lead to improvements in other areas (such as user trust) where we extend explainable AI to user interactions.
  2. **Explainable AI for modellers to understand their systems:** The entity responsible for building the AI explainer itself needs to understand how the ML models make decisions to identify blind spots and thus perfecting the system. Tools for explainability should offer model authors finer insights and features. However, just identifying weaknesses might fall short. It is recommended that the system be able to automatically refine the models with the underperforming segments and automatically improve.
  3. **Transparency beyond AI User Interface:** An AI system should work as a whole that offers the best possible experience to users. From a holistic point of view, there are initiatives (unrelated to AI) that help to increase the transparency of the services provided by these platforms. For example, in the back end, documentation of the dataset is one of the essential points for transparency, as well as its verification and validation of data quality. In turn, on the front end, functionalities are improved in order to preserve the trust of users. Enhancing the application elevates transparency at every level, ultimately enhancing content clarity. Other points could be mentioned since there is a long way to go to demystify the behaviour of AI Systems, however, we believe that every step we take towards transparency is a step in the right direction.

Thus, nowadays, ML have transcended the ”black-box”paradigm, with various avenues for interpretation now available. Unlike other forms of AI, which cannot offer evidence, reliability or error correction methods for the results,XAI sheds light on the processes, reasoning and origins of errors within these models. Furthermore, it doesn't solely provide outcomes; it identifies the reasons behind the results achieved and explains the roots of AI errors. This is realized through an explanatory interface, enhancing existing research by rendering the understanding and interpretation of outcomes immediate and accessible, rather than just presenting calculated result values.

4.4 Summary

To conclude, this chapter presents a visionary framework for the future development of urban areas, aiming to expedite and optimize the implementation of mobility services. It is proposing cross-concept collaboration in mobility research to establish a Proof of Concept (PoC) platform. This helps us understand that the IoT has an important role in the evolution of smart cities and illustrates from a practical standpoint the wide applicability of ML.

Taking into account the potential of XAI models, the PoC initiative focuses on promising projects in the domain of smart human mobility. Its objective is to achieve an early mobility project at a low cost. Basically, the pilot project serves to demonstrate the potential between the DL models, providing additional mobility capabilities, creating service samples, and identifying the stakeholders. Notably, the PoC collaboration format was also extended to research areas beyond smart mobility research. Additionally, the adoption of a tactical approach addressed the immediate, highly-visible challenges associated with device proliferation and user expectations in terms of accessibility. Overall, this marks the beginning of a new phase in our research journey, poised to elevate our endeavours to the next level of development, as expounded in the forthcoming.

Chapter 5

Design and Implementation

This chapter describes the planning and defines the elements comprising the architecture of the proposed human mobility solution. Building upon previous chapters where we identified the key topics and explored existing evidence indicating that Smart City (SC) are fertile grounds for innovative and cutting-edge projects in several areas, the time has come to idealize and implement the proposed smart service.

To preliminary assess the solution's components, we implemented an prototype for the study of human mobility based on real data provided by already established services in the field, along with Machine Learning (ML) prediction models that align with our intended objectives. Furthermore, to some extent this archetype not only aids in supporting the architecture we intend to unveil but also lays the foundation for it. Subsequently, we define the interconnected components that provide the service, while the implementation phase identifies and defines these components, providing stakeholders with a comprehensive overview of the proposed architecture.

5.1 Analysis of Human Mobility Prototype

From this important part of the research plan and investigation methodology, we idealized a prototype to address both the research hypothesis and research question posed in Chapter 1. This prototype takes into account SC concepts and spans the several areas of intervention mentioned throughout this section, with a particular focus on human mobility. For preliminary results, we used data collected from LinkNYC Kiosks, situated across five boroughs of New York City. These kiosks stored a historical list of their devices, locations, and the operational status of the Link's Wi-Fi, tablets, and phones. Notably, LinkNYC represents a pioneering communication network, aiming to offer the fastest available free public Wi-Fi to millions of New Yorkers and visitors [31].

5.1.1 Materials and Methods

We used a sample spanning the last 3 years, between 2016 and 2018, with a monthly interval of 1. The original dataset consisted of the count of registered individuals at specific kiosks for each day of the month, within a 12-hour interval. As a result, we had a set of 1739 links measured between 01-2016 and 12-2018, located in different locations across New York City (Table 4).

Borough	Date	Location	Census	Temp ($^{\circ}$C)	Humidity (%)	Precip	Press (Pa)	Dew-point ($^{\circ}$F)	Wind (MPH)
1	2018-02-12	(40.71, -73.95)	514260	12	50.3	0	30.3	13	2
2	2018-05-17	(40.85, -74.10)	2208318	14	59.8	0	29.9	12	2
1	2018-03-08	(40.71, -73.95)	3387972	13	61.2	0	30.6	11	3

Note: We ignored Wi-Fi, Tablet and Phone features because they present constant values.

Table 4: LinkNYC Kiosks locations dataset.

The collected data provides a wealth of insights. It not only allows us to detect devices within specific areas but also facilitates the monitoring of human activity. Furthermore, the data grants us control over authorized individuals connecting to these devices, and by utilizing multiple Wi-Fi sensors, we gain the capability to track real-time movement. Additionally, these findings are further refined through pre-processing techniques.

During the data pre-processing phase, data gathering methods are often loosely controlled, resulting in missing attribute values (e.g., install date, active date) or certain attributes of interest (e.g., Wi-Fi status, Tablet status, Phone status), impossible data combinations (e.g., install date: yes; activate date: yes), or data only presented in aggregated form. Thus, a thorough analysis of data quality and representation is essential prior to conducting any analysis.

For pre-processing we are encouraged to use a list of tools but assuming that data pre-processing has three main components like extraction, transformation and loading, we opted for the *KNIME* [194]. As we mentioned in Section 2.4.2.3, this open-source data analytics, reporting, and integration platform gives us a graphical user interface to allow for the assembly of nodes for data processing. It encompasses modular data pipelining, leveraging ML and Data Mining (DM) concepts, and is versatile for building business intelligence reports. Figure 24 shows the workflow of data precessing.

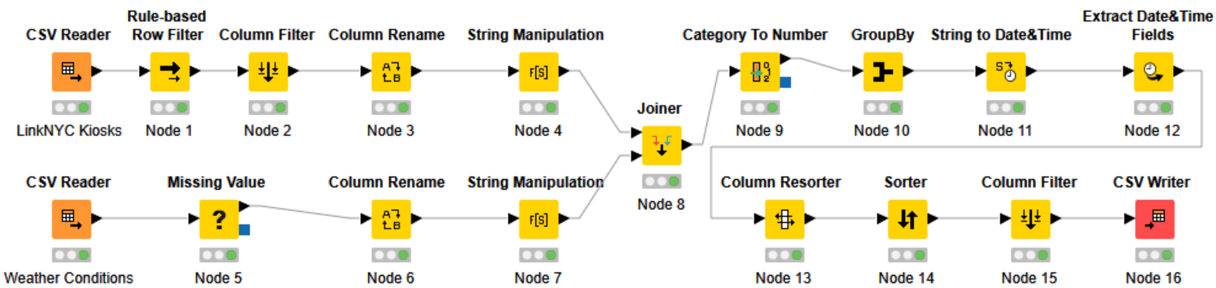

Figure 24: The process of data pre-processing.

In the first nodes are responsible for importing the original data from the LinkNYC Kiosks and Weather Conditions Comma-Separated Values (CSV) files. Since there are missing values in the imported data, these are excluded in node 1. In Node 5, the data gaps in weather information are addressed by filling them with the most frequent value. Node 2 trims the dataset to include only the most important columns, specifically $Borough$, $Date$, $Location$, and $Census$. Node 3 and node 6 serve to standardize the column names across both datasets. In node 4 and node 7, the process changes the date format to facilitate data aggregation. Node 8 conducts the calculation of feature importance, an essential phase where datasets are merged based on the date. At the same time, it filters the unnecessary columns to reduce the noise. In node 10, the outcome includes the calculation of the number of people living in a particular place by data aggregation such as region, geo-coordinates, and date, calculating the mean for all numeric features (e.g., temperature, humidity). Notably, this calculation takes place post- categorical to numerical conversion (e.g., weather description), in node 9. Node 11 converts the date string to datetime format. In node 12, the process extracts the datetime fields to $Year$ and $Month(ofYear)$. Prior to data export, a final adjustment to the columns and the removal of redundant features from the merged data is performed. Node 16 writes the samples into a CSV file in order to use the Deep Learning (DL) algorithms.

During feature extraction, census figures within the same time interval are summed, and a distance threshold is applied, taking into consideration the maximum length of each borough for better visualization. Points mapped to the same street are unified under one density colour, and the location is illustrated through the density colour segment. To enhance clarity, after the feature extraction, the location sequence is represented between blue and green points. The visualization output is depicted on the map showcased in Figure 25.

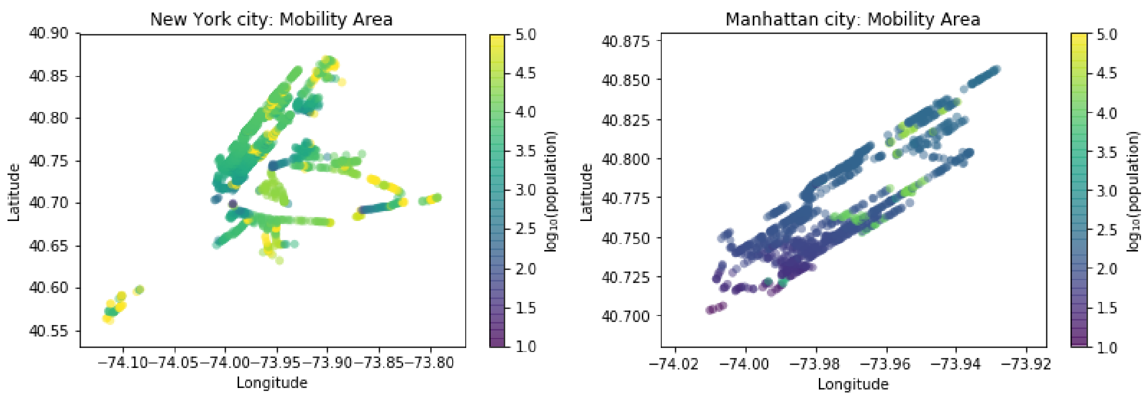

Figure 25: Density of tracked people in New York City and the city of Manhattan via LinkNYC kiosks over the past 3 years.

We solely consider filtered data from the city of Manhattan to evaluate human mobility predictions. As detailed in Section 5.1.2, we transform the LinkNYC kiosks dataset to render it compatible with the Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN) and Convolutional Neural Network-Long Short Term memory (CNN-LSTM) models. Firstly, we split the supervised data into two matrices: the training matrix X, encompassing the first 24 months, and the testing matrix Y, containing the remaining data. Matrix X facilitates model training, while Y is used to evaluate and validate its accuracy. Just as with the training set, we need to scale our test data. Subsequently, we use Keras library to build and train our model.

The exploration of user patterns involves an analysis of data trends over a specific time span. As exemplified in Table 4, the number of people who access the kiosk located at lat 40.71 and long -73.95 varies in different time series. The utilization of time series data is recommended for the prediction of the future value of an item based on its past values, with the help of DL algorithms like LSTM and CNN.

We will be predicting the future number of people in New York City based on the number of tracked people via LinkNYC kiosks over the past 3 years. The calculations were conducted within the Python environment using the main packages *pandas*, *sklearn*, *numpy*, *keras*, *matplotlib*, *seaborn*. Figure 26 shows typical time series graphs depicting kiosk foot traffic, with values normalized using $\log_{10}$ units.

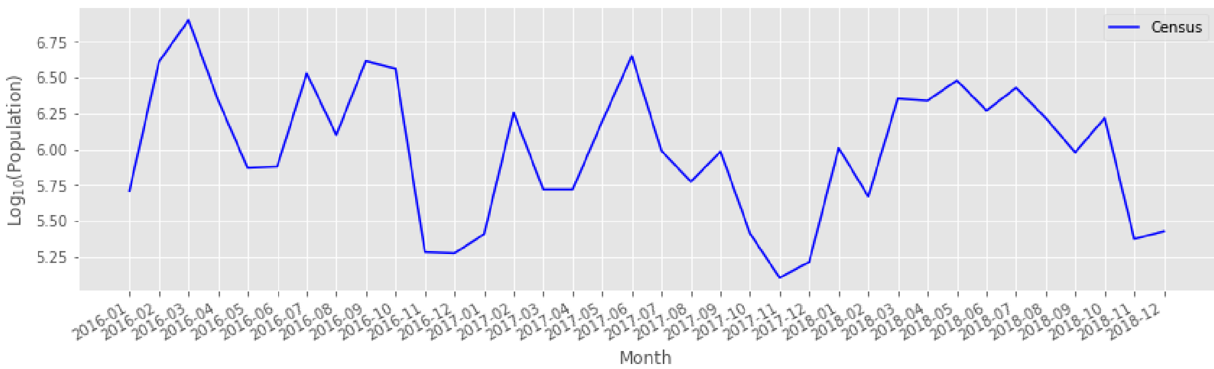

Figure 26: Number of tracked people via LinkNYC kiosks over the past 3 years.

The depicted trend exhibits a pronounced non-linear nature, characterized by its multiscale and long-range dependence. Capturing this intricate trend proves to be a challenging endeavor with the available information. Hence, the requirement for an apt model for time series forecasting becomes evident. This is where the combined power of CNN and LSTM (or CNN-LSTM model) comes into play. While the former is adept at forward-looking predictions through its self-learning method, ensuring more realistic forecasts, the latter excels in memorizing past data, facilitating the prediction of tracked foot traffic over an extended sequence of time steps.

5.1.2 Experimental Protocol

To robustly assess the accuracy of the prediction model, the Root Mean Square Error (RMSE) metric proves invaluable. The main benefit of using RMSE is that it penalizes large errors. Moreover, its score scaling aligns with the forecast values' units (i.e., per month, for this study). After, we repeat the model construction and prediction several times. The average RMSE is an indication of how well our configuration would be expected to perform on unseen real-world LinkNYC kiosks data. Finally, we compare our predictions with actual trends in human mobility which can be inferred from historical data.

The LSTM model was trained with an 80 batch size, 100 nodes, 200 epochs, and Adam optimization. Moreover, to develop a robust result, we repeat the experiment 20 times, a number deemed sufficient to provide a good distribution of RMSE scores. The outcome comprises a line plot that juxtaposes the test data (blue) and the predicted values (red), contextualizing the model's skill. Additionally, a box-and-whisker plot encapsulating the distribution is presented for further visualization.

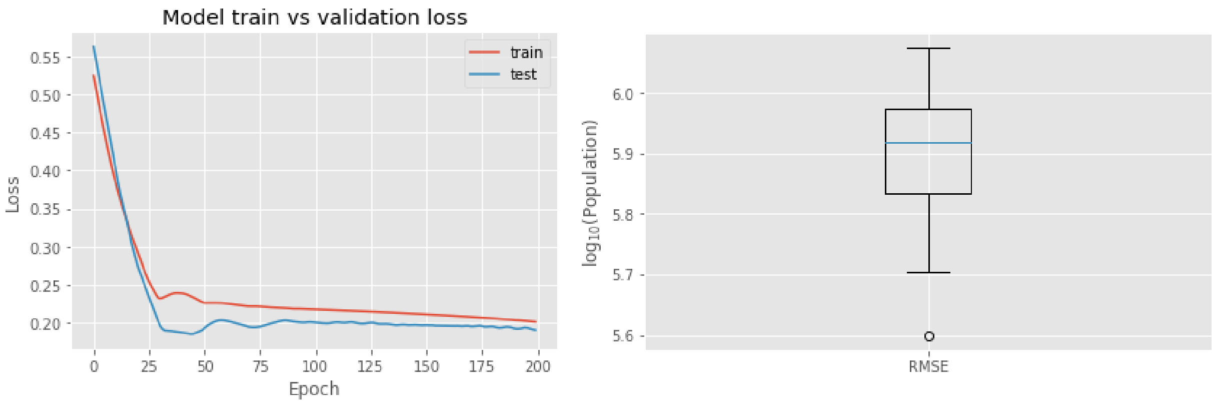

Figure 27: Plot of Loss on the Train and Test datasets and Root Mean Square Error score in a box and whisker plot from LSTM model.

As observed in Figure 27, the left side shows the printing of train and test loss after each training epoch; the right side summarizes the distribution of RMSE scores. The first chart is important for predicting human mobility based on the train-line chart and test-line chart results. Interestingly, we can see that the test loss drops below the training loss, potentially indicating the model's overfitting to the training data. To be more precise, the chart exhibits a substantial exponential reduction up until the 30th epoch, where—after 50 epochs—the lines tend to run parallel.

Concerning the measurement and plotting of RMSE during training, as outlined in Table 5, we can see that the mean and standard deviation stand at $10^{6.16}$ and $10^{0.04}$ respectively for the monthly number of connected individuals. Correspondingly, the RMSE value tallies at $10^{0.18}$. Furthermore, it shows that this model is the second lowest among the human mobility RMSE chart representations.

In the CNN algorithm, we develop a suite of CNN models with layers, represented in Figure 28. In other words, we choose an one-dimensional CNN design, featuring a convolutional hidden layer that operates over a 1D sequence. This is followed by a potential pooling layer, responsible for condensing the convolutional layer's output to essential elements. Subsequently, the convolutional and pooling layers are followed by a dense fully connected layer which interprets the features extracted by the convolutional part of the model. A flat layer serves as an intermediary between the convolutional layers and the dense layer, reducing the feature maps into a single one-dimensional vector.

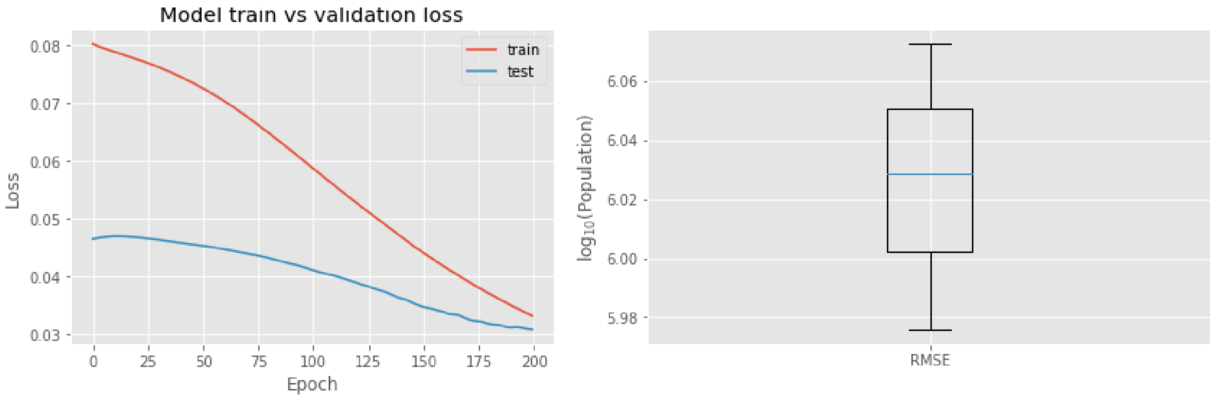

Figure 28: Plot of Loss on the Train and Test datasets and Root Mean Square Error score in a box and whisker plot from CNN model.

Training data involves 64 filters, 5 kernel sizes, 80 batch sizes, and 200 epochs. To find a good result, the experiment is repeated the same number of times as the previous model. Evaluating the out-of-sample results, they are gauged by the $10^{0.23}$ RMSE, alongside a mean of $10^{6.02}$ and a standard deviation of $10^{0.05}$. As a result, we see in Figure 28 that the model chart has a worse performance than that of the LSTM chart. This illustrates the superiority of LSTM model over the CNN model in the context of predicting human mobility.

For a mixed CNN-LSTM model, we define the CNN model first, then add it to the LSTM model by wrapping the entire sequence of CNN layers within a TimeDistributed layer. The idea is to merge the two models, creating a third model. Through experiments with this fusion model, we can draw performance comparisons. The construction of this model assumes that the parameters used are the same as the previous models.

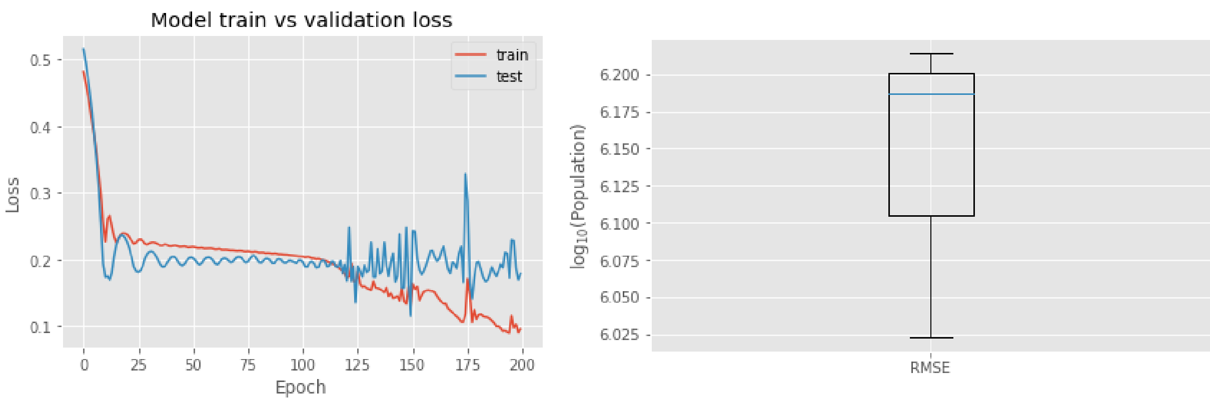

Figure 29: Plot of Loss on the Train and Test datasets and Root Mean Square Error score in a box and whisker plot from CNN-LSTM model.

The loss values represented in Figure 29 show that there is a good overfit model up to 125 epochs. During this period, the model presents good performance on the training set, steadily improving and resulting in a decrease in RMSE. Simultaneously, the model's performance on the validation set demonstrates initial improvement followed by degradation. Next, we check which model is more appropriate for predicting human mobility.

Transitioning to Section 5.1.3, our approach involves dissecting the models, first the LSTM model, followed by the CNN model, and finally the CNN-LSTM model approach. The objective is to provide the best model for time series forecasting of our survey.

5.1.3 Preliminary Results and Discussion

Our survey findings are concisely summarized in Table 5, offering an overview of out-of-sample results of experiments with human mobility density in the city of Manhattan. We evaluated three different models. First, we determined the RMSE using the LSTM model to predict human mobility. Second, we showed that the CNN model performs worse than LSTM and CNN-LSTM models. Lastly, we needed to make sure that our proposed merged CNN-LSTM model is meaningful. Thus, such as in previous results, we resort again to the RMSE equation but add other performance metrics.

	Model		
	LSTM	CNN	CNN-LSTM
Mean	6.16	6.02	6.14
Standard deviation	0.04	0.05	0.10
Minimum	6.07	5.93	6.02
Maximum	6.29	6.07	6.21
Accuracy$^{*}_{T=25}$ (%)	6.00	5.97	6.10
Accuracy$^{*}_{T=50}$ (%)	6.02	6.02	6.18
Accuracy$^{*}_{T=75}$ (%)	6.05	6.01	6.20
RMSE (log_{10})	**0.18**	**0.23**	**0.14**

* Train set size percentage.

Table 5: The human mobility modeling results (best values in bold).

The results reveal a general trend of improved prediction accuracy across all three models as the dataset size increases. In terms of model comparison based solely on metrics, the out-of-sample outcomes indicate that the CNN-LSTM surpasses both the LSTM and CNN models in accuracy and the variation between maximum and minimum values. This validates our proposed CNN-LSTM model as a superior choice, demonstrating heightened accuracy compared to the other models.

5.2 The WalkingStreet Platform: Project Plan

The *WalkingStreet* platform is designed to make use of human mobility in a computing environment. Built upon the architecture of Mobility as a Service (MaaS), this platform is composed of different components, each contributing to a specific step in the workflow. This workflow is characterized by the involvement and process of sensor infrastructure, data acquisition, treatment and storage of data via fusion method, data acquisition based on the performance of mining and modelling data, the integration process of compatible services, including predictable data, and its interpretation, explanation, and visualization in explainable data.

Predictive and explainable processes are executed using sophisticated DL methods hosted on the MaaS core. Within the SC assessment system, these and other tasks focusing on human mobility, are analyzed by measuring a set of key human mobility metrics. These metrics are are also showcased in practical examples provided for a comprehensive investigation of their properties and potentialities, although, they were designed to be also used in generic applications. However, the degree of citizen participation can significantly impact the outcomes, as many implementations rely on metrics influenced directly by the level and quality of user engagement. Therefore, this MaaS platform has a generic structure, customizable for specific applications, and is open to the active participation of different players.

This platform operates as an intelligent environment, with certain core modules functioning independently of firect user control. This autonomy extends from a dense sensor network embedded in physical objects and infrastructures to background processes such as ML models or XAI methodologies. Additionally, these components are developed for continuous adjustment to user requirements and environmental conditions. This means that the adjustment process often begins with the collection of user environment data, behavioural monitoring, as well as other associated workflows.

This MaaS project, beyond its adaptability for different functions, it also authorizes the installation of external devices, even those not initially envisioned by the developer but introduced by users. Additionally, its open architecture enables data exchange between multi-device applications, without proprietary restrictions. This seamless communication across heterogeneous systems is only possible through unlocked network capabilities with other domains, such as cloud and web services. This means that this open approach adopts an interoperable infrastructure capable of aggregating bulk data from diverse domains and tailoring them to specific urban scenarios.

The architectural overview, main components, and additional services of this MaaS application are described in the following sections. This Proof of Concept (PoC) needs to be built using available technologies or deliver value to various human mobility actors. From the specification of individual sensors for detecting and responding to some type of input from the physical environment to the modules for metric analysis and predictive modeling that generate human- readable outputs, this platform encompasses a comprehensive data storage system and interpretable policies. These collectively facilitate the understanding and explanation human mobility phenomena based on Artificial Intelligence (AI) outputs. The chapter concludes by illustrating how the research was conducted and the PoC deployed.

5.2.1 Architecture Overview

The elucidation of smart human mobility elements in Chapter 4 resulted in the conceptualization of the *WalkingStreet* platform. This platform combines structures from a hub technology with an infrastructure for information acquisition, sensing, and tracking supported by a sensor network architecture, exploratory analysis of human mobility data employing a set of metrics, the engagement of smart device users in participatory sensing tasks to derive value-sensitive applications, and the integration of XAI to make an AI platform more transparent and understandable to end-users.

The selected software development approach breaks the process of developing this platform application into modular units. Each module represents the whole iteration development process. With the core platform module, the AI system resides, alongside an API service defined by a set of protocols and access permissions based on user roles, as well as a data storage infrastructure. Other modules can function independently or dynamically combine as needed. Moreover, this development approach specifically recognizes that modules will require adjustments over time, with these considered integral to the platform's lifecycle. It particularly emphasizes the organization and integration of different developments into system integration modules, where hardware is an essential part of the solution. For example, it involves creating an intelligent module focused on data-driven decision-making [62]. This leads to gathering all relevant available data and then implementing modern tools for monitoring and assessing different human mobility indicators [200]. Figure 30 illustrates the physical representation of the *WalkingStreet* platform.

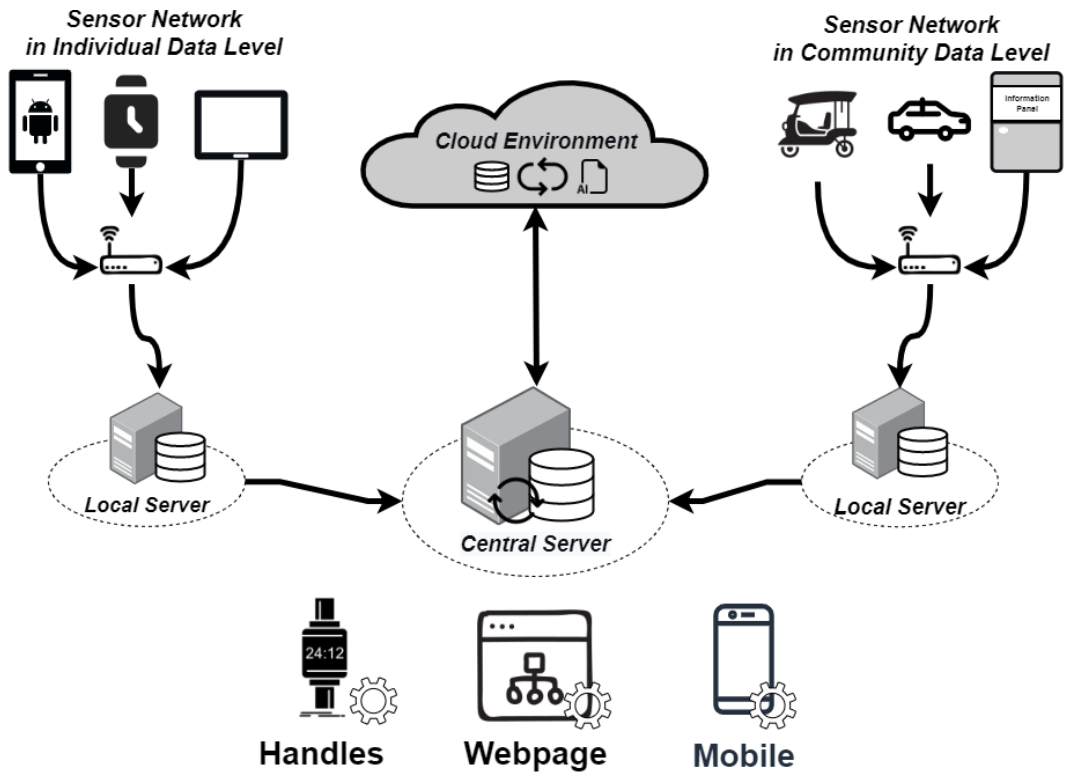

Figure 30: WalkingStreet components platform.

At the infrastructure level, this platform intends to collect human mobility information from different sensor network sources, such as individual data sources like smartwatches, tablets, smartphones, and more (referred to as *People as Sensors*), as well as community-level data sources (e.g., Public Information Displays, Interactive Urban Connectivity, etc.). Each sensor requires fast and reliable connectivity to efficiently contribute raw data, which is then stored in local servers (or decentralized modules). These decentralized modules manage data communication between sensors (or devices) and the local servers within a sophisticated network infrastructure. Each local server is accessible remotely through a central server (or core module) via a REST API. This server centralizes all data, including personal user-to-environment information generated at each local server. Additionally, leveraging the capabilities of a fog computing architecture, the central server sends the consolidated data to a cloud-based server. Then, in a cloud environment, AI-Driven Big Data is processed and made accessible through restful web services and APIs, providing services related to human mobility for the city.

At the application level, the core module takes advantage of infrastructure-level technologies to support efficient sensor data aggregation mechanisms and semantic web technologies for unified data representation. Simultaneously, an ML kit on a cloud server employs predictive mechanisms and renders outcomes that are explainable and comprehensible to end-users. Moreover, this explicability can be manages through an authentication system or made openly accessible to encourage community collaboration and knowledge sharing regarding intelligent solutions.

5.2.2 Decentralized Module

As depicted in Figure 30, the decentralized module includes a multi-sensor network system that operates at both individual data and community data levels. In this system, each group of sensors located in various positions transmit data to a local server, where this data is stored. Furthermore, the server can be accessed by multiple external applications via web servers. To provide a user-friendly interface for viewing and analyzing stored data, a web application is also included in this module.

As discussed in Section 2.2.2, the sensor network is an important module to support any server application, ranging from monitoring a single device, such as a smartphone, to tracking different modes of transportation or interactive public displays. In other words, these sensors facilitate the monitoring of the environment. Moreover, they are responsible for ensuring the privacy and security of the targeted public data. To achieve this, communication protocols are developed by establishing a connection between the sensor and the local server. The following sub-sections detail the architecture and features, communication, and storage within the decentralized module.

5.2.2.1 Components and integration

As previously mentioned, *SiteWhere* employs a fully distributed architecture using Kubernetes and a variety of micro-services to construct the system. Figure 31 illustrates how these micro-services enable well-defined interactions between multiple areas of the system, facilitating independent scaling of individual functional areas or their exclusion as needed. Moreover, Istio provides a service mesh for the system's micro-services, providing a more robust, fault-tolerant system, and providing detailed monitoring and tracing of data flow between micro-services [100].

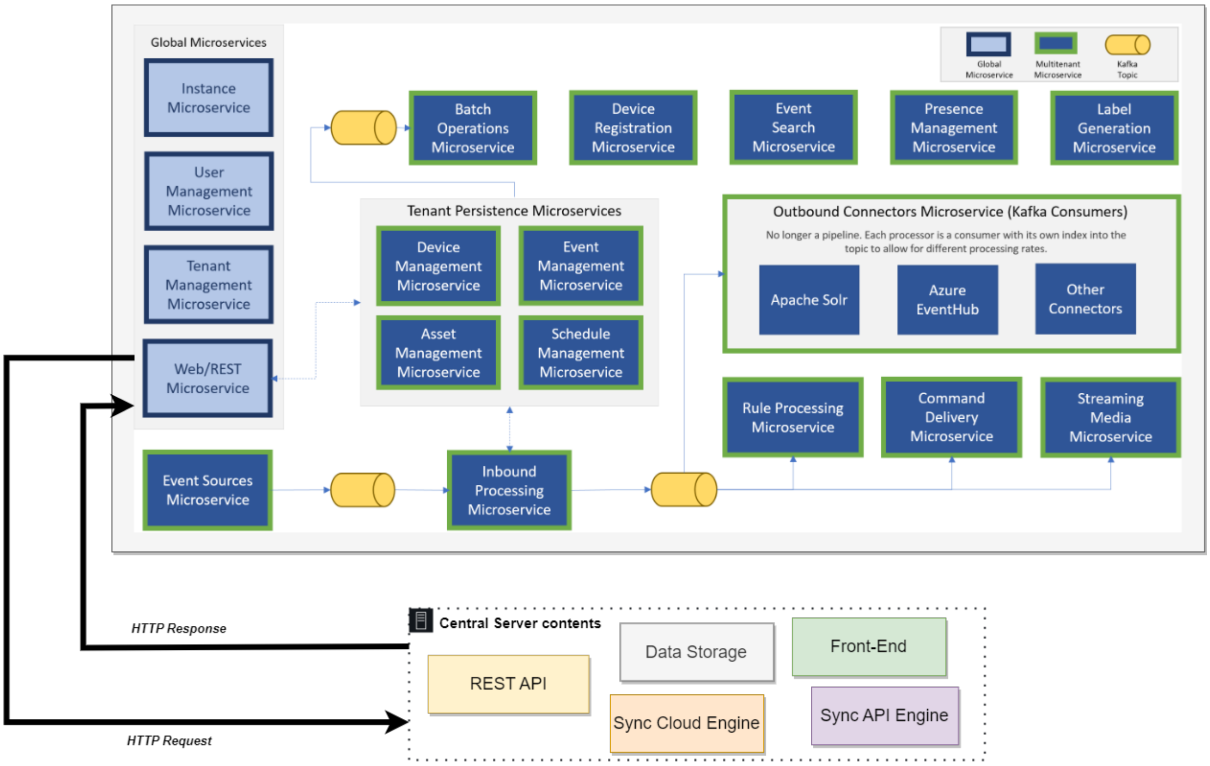

Figure 31: Local server architecture (central server transition).

SiteWhere also includes technologies such as Apache Zookeeper. Its configuration management is centralized, but it can be externalized for scalable configuration management. ZooKeeper also contains a hierarchical structure that represents configuration data for one or more *SiteWhere* instances, with this configuration data being replicated across all micro-services.

For scalable storage, *SiteWhere* uses Rook.io. Rook distributes and replicates block storage, making it resilient to hardware failures. As storage and throughput requirements increase over time, new storage devices can be dynamically added. This platform also supports various databases including MongoDB, InfluxDB, Cassandra, and other supporting technologies such as Message Queuing Telemetry Transport (MQTT) brokers.

To facilitate data movement between micro-services, *SiteWhere* uses Apache Kafka. Micro-services process data asynchronously through pipelines, reading data from well-known inbound topics, processing it, and then sending the processes data to well-known outbound topics. In its turn, external entities use *SiteWhere* topics to consume the data as it moves through the system. Data loss is prevented, as data is always backed by Kafka's high-performance storage since all data in the distributed log is stored on disk and can be replayed to the event stream based on previously gathered data, which is useful for debugging processing logic or load testing the system.

SiteWhere also offers a set of APIs, ensuring that new technologies can be easily integrated as the IoT ecosystem evolves. These APIs facilitate real-time operations between the micro-services, such as storing persistent data, initializing micro-service-specific services, or managing devices and events. Although REST API is exposed directly as Remote Procedure Call (gRPC) services and available via the Web/REST micro-service (i.e., API gateway), they use the gRPC API underneath to provide a consistent approach to accessing data.

The complete *SiteWhere IoT ecosystem* incorporates multitenancy data storage. Each tenant operates within its own processing pipelines, ensuring that in-flight data is never co-mingled. By default, tenants do not share database resources or pipeline processing and have a completely separate configuration lifecycle. With this approach, each tenant may use its own database technologies, external integrations, and other configuration options. Moreover, parts of a tenant's processing pipeline may be reconfigured or restarted without causing disruptions to other tenants, ensuring security and scalability.

5.2.2.2 Additional Experimental Environment

SiteWhere serves as an example of the rapidly evolving Internet of Things (IoT) market is evolving at a rapid pace. As mentioned, the reason behind this growth is the huge demand for platforms, devices and other components on IoT environment. As for its architecture, the IoT landscape encompasses a wide array of elements, protocols, sensors, cloud services, actuators, and layers, typically organized in four stages (see Section 2.2.2.1). These stages include sensors and actuators, Edge IT, Internet getaways and data acquisition systems, as well as data centers and the cloud, which together enhance scalability and provide extensive storage for sharing human mobility information [193].

It's important to note that *SiteWhere* is the only kind of local server that can be seamlessly integrated with the *WalkingStreet* platform. Moreover, the system architecture also supports open data infrastructures. As mentioned in Section 4.1.3, some organizations and companies offer such infrastructure and advocate for policies that encourage data sharing. Many of these entities specialize in building platforms which offer public access to a collection of datasets. In fact, one of these platforms, LinkKiosk, is employed for the study of human mobility, as detailed in Section 5.1. At the architectural level, these infrastructures are designed to facilitate applications before transferring data to a centralized database for further analysis or distribution to other applications through our API and Web Services integration points.

5.2.3 Core Module

The *WalkingStreet* ecosphere starts with the deployment of the most basic sensors located in the most remote corners of the city. These sensors are then connected to a local network that sends data and receives commands from a local server. Subsequently, packetized data will traverse various channels until it reaches a large data center. The strength of this platform doesn't hinge on a single local server but rather on the collective power of hundreds, thousands, and potentially millions of servers and other external services. Therefore, this module must incorporate a flexible architecture that groups a set of rules and guidelines to determine how micro-services and other components must behave with each other on a docker host.

As a central server, security is one of the most critical pillars to help protect data and ensure the integrity and availability of the *WalkingStreet* platform. Scalability and applicability are also fundamental considerations. To address these concerns, container technologies play an increasingly important role, executed directly on more constrained edge devices. Workloads are structured as applications based on micro-service. As shown in Figure 32, these applications are effectively managed and orchestrated through an on-premises manageability solution, working in conjunction with a local server (hosting either a *SiteWhere* or RESTful API infrastructure) and a cloud server.

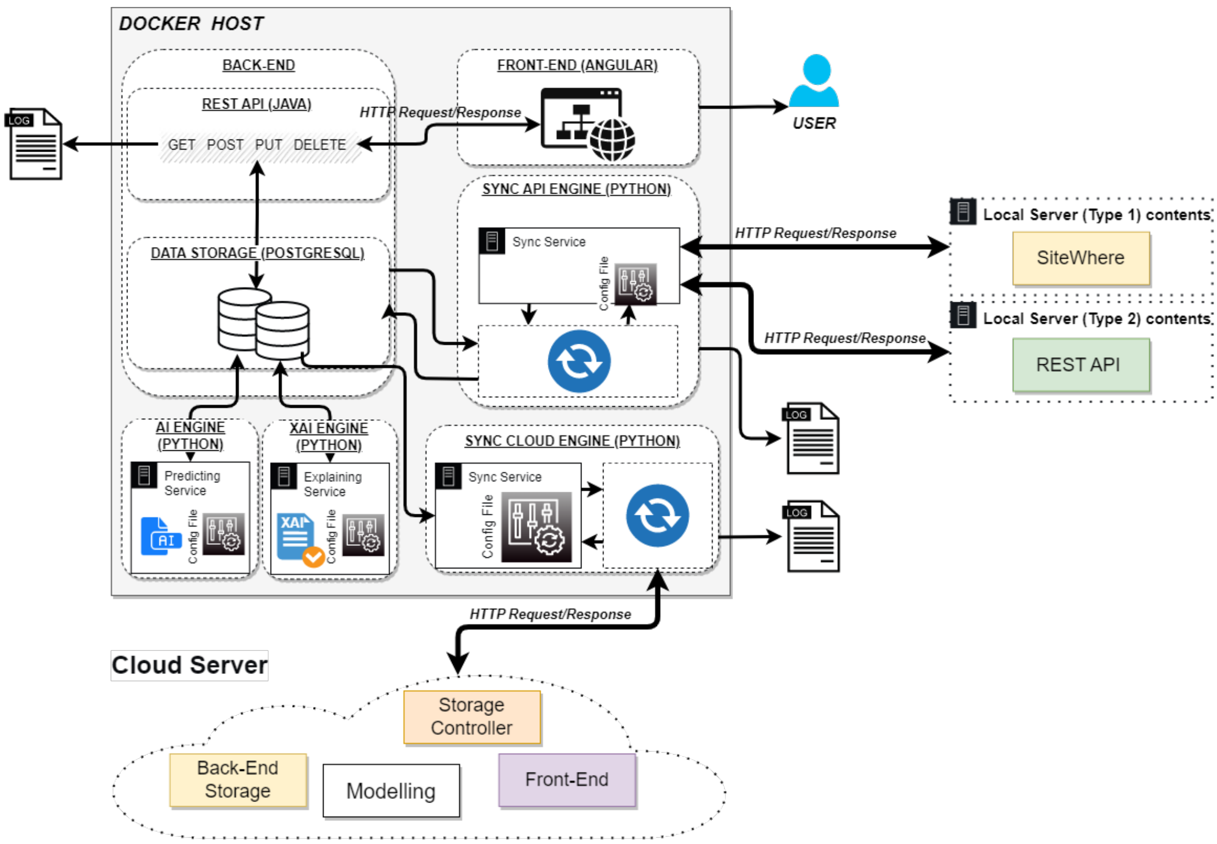

Figure 32: Central server architecture (cloud transition).

As previously mentioned, the main module of the *WalkingStreet* platform, referred to as *Core Module*, is supported by external servers, a *docker host* dedicated to data storage and modelling, and it transmits the outcome to a *cloud server*. But, specifically, the docker host is designed to calculate and explain predictive metrics about human mobility. To accomplish this, the docker-based containers are composed of six different micro-services:

- *Sync API Engine:* This service synchronizes data from external REST services with the central server;
- *Data Storage:* As one of the components in the *Back-End* architecture hosted on Docker, it stores data primarily originating from the *Sync API engine* micro-service. Additionally, it includes AI data such as metrics and predictive data created from the *AI Engine* micro-service;
- *AI Engine:* This module is responsible for computing ML models and metrics using Human Mobility data stored in the *Data Storage* micro-service;
- *XAI Engine:* It performs explainable AI models that facilitate transparency, comprehensibility, and an ability to question or query AI outputs generated from the *AI Engine* container;
- *Sync Cloud Engine:* This service syncs processed data (i.e., predicted data and computed human mobility metrics) between the *central server* and the *cloud server*;
- *REST API:* Integrated into the *Back-End* architecture, this component is dedicated to serving the *Front-End* service. The API is also offered as a service to 3rd party applications;
- *Front-End:* This element enables users to view and edit the stored data in the *Data Storage* module by accessing an API incorporated into the *Back-End* architecture.

All these micro-services are hosted within a centralized environment, by encapsulated within Docker containers. Although these Docker containers (each container typically represents a process) promote resource allocation and distribution, each service instance must interact using an inter-process communication protocol. These could include Hypertext Transfer Protocol (HTTP), Advanced Message Queuing Protocol (AMQP), or a binary protocol like Transmission Control Protocol (TCP), depending on the nature of each service. In the following sections, we will delve into the specific roles of each micro-service and their interactions with one another.

5.2.3.1 Sync API Engine

The Sync API Engine micro-service is responsible for synchronizing data between external systems and central server applications. It is configured to manage one or a set of external REST API Engine is configured to control one or a set of external REST API and is synchronized to retrieve information from multiple web servers where this service has the appropriate permissions to read or write their information.

In general, this service operates by periodically running, taking into account parameters defined in the configuration file. However, its operations should also consider other parameters, which include:

- *Cron Jobs*: an API for scheduling jobs, running periodically (i.e., at a given time and date) Python scripts;
- *API Endpoint(s)*: this endpoint is essentially one end of a communication channel that interacts with the platform. It can include the URL of a server or service;
- *Frequency*: it provides the ability to customize the sync frequency for a specific endpoint. It allows you to specify the time interval between syncs in terms of minutes, hours, or days;
- *Number of Rows*: it defines the maximum number of rows that are allowed per synchronization run. This limitation is in place to optimize performance and not overload the micro-service.

Before starting the synchronization process, the service must create an instance of the *SyncAPIEngine*. Then, it updates records in the metadata tables with data. To facilitate this process, the mechanism within the Central Server responsible for synchronization with the data storage component is known as the *Data Engine*.

5.2.3.2 Data Storage

This service enables the storage of the data within the system through synchronization with external systems. Then, it manages the storage of data and facilitates its export for training in the *AI Engine Service* or accessed from a Back-End platform. In the case of synchronization, this task is performed using *SyncDataStorage* class, ensuring data integrity and consistency. The database can be shared with other micro-services and can be queried directly or via a Rest API service.

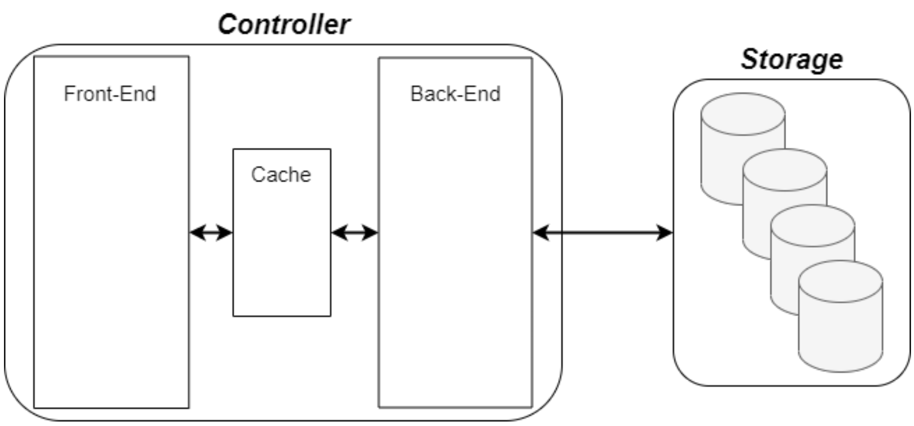

Figure 33: Block-based storage system architecture.

The centralization of the database serves to foster the necessary interdependencies between micro-services. If there is an update to the schema of one micro-service's database, this update should directly impact other micro-services. Furthermore, other micro-services should not need to be aware of these database changes. By maintaining a centralized database location, the platform ensures that changes to the database schema do not limit the scope of changes in micro-services when they occur.

In more detail, storage systems used for training ML models typically deal with large volumes of unstructured data. In other words, these environments leverage vast amounts of unstructured data rather than organized blocks or databases. But, in general, having access to more pertinent data leads to better insights. However, the challenge lies in providing high-performance storage at a scale that remains cost-effective and suitable for long-term data archiving.

To meet all the storage requirements for ML and analytic models selected for use on this platform, an option that combines block-based storage is considered ideal. This option balances the workload against factors such as data volume, data types, and the speed of decision-making, especially in an AI training environment where low latency is important. The *Block-based Storage System* (Figure 33) represents a conservative yet effective approach to storing data analytics within the *Central Server* architecture. This system not only supports data centralization but also provides one or more advanced features such as controller(s) to enhance repeated operations and offloading storage-specific processing for the array and the number of storage disks.

5.2.3.3 AI and XAI Engines

This container is responsible for modelling the stored data and operates alongside a separate data-wrapping micro-service (or *Data Storage* container). It can handle multiple tasks running independently, including data ingestion, pre-processing, training, prediction, and explainability. The Docker of the *Central Server* runs two containers to load-balance the training (*AIEngine* class) and explaining (*XAIEngine* class) components. This is usually a good practice to follow when using both AI and XAI models [50]. Additionally, each *Config file* for these containers contain the following parameters:

- *Cron Jobs*: specifies the time and date for running periodically jobs;
- *Frequency*: customizes the sync frequency of *Cron Jobs*;
- *Number of Rows*: defines the maximum number of rows to be used on AI or XAI models.

Each model can be handled by different micro-services (i.e., *AI ENGINE* and *XAI ENGINE*), which are deployed in separate containers. Additionally, this diagram highlights the differences in internal operations for both use cases, like train service, including the services responsible for handling outcomes from model input, namely the *Predict Service* and *Explain Service*. The *Predictive Service* is a process to address potential issues and interruptions before leading to breakdowns in operations with *AI ENGINE*. While the *Explain Service* is responsible to express why an AI system reached a particular decision, recommendation, or prediction. In its turn, this service is member of *XAI ENGINE* frameworks to help understand and interpret predictions made by ML models. These engines can be visualized in Figure 34.

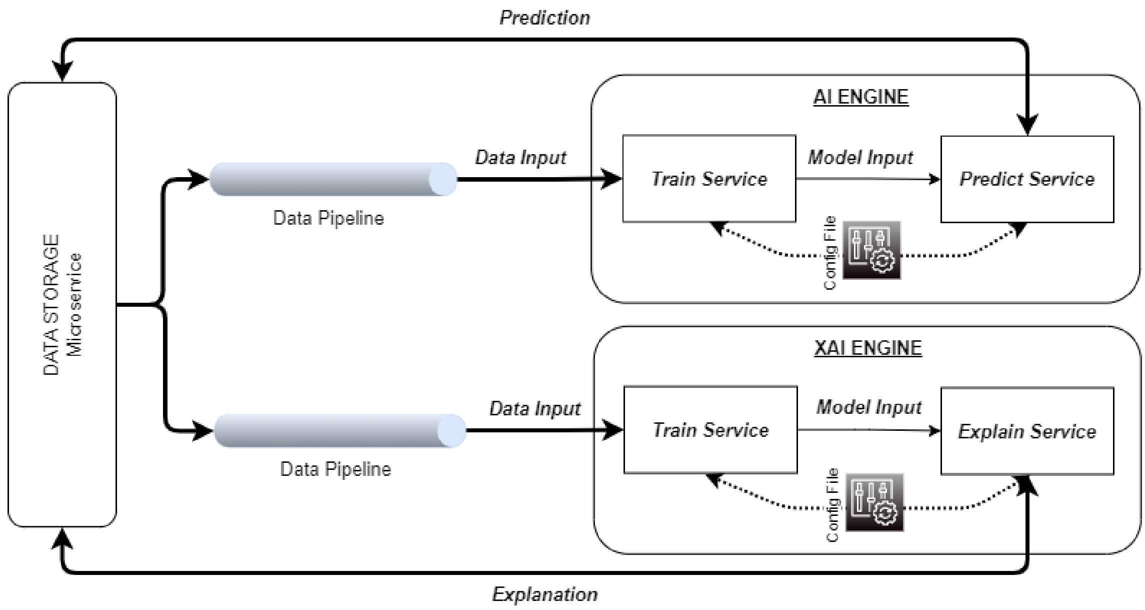

Figure 34: Artificial Intelligence Engine and Explainable Artificial Intelligence Engine Architectures.

The proposed Docker architecture is designed to accommodate several AI or XAI models seamlessly. This flexibility is crucial for adding and removing services without disrupting other services. However, scaling these models in a production environment can be challenging. Thus, the micro-services architecture should be operational and flexible, allowing the setup of multiple pipelines and effective workloads management.

Metrics Analysis Based on the *WalkingStreet* architecture, which encompasses data from GPS traces to location-based social network data captured by personal and public devices, the platform offers the opportunity to analyze and visualize large sets of geospatial human mobility data. This data, initially collected by a local server, is used by the Central Server to calculate the key human mobility metrics and measures as described in Section 2.3.3. Additionally, it provides a brief overview of the fundamental physics behind human mobility. Nonetheless, prior to this modelling process, the *WalkingStreet* platform proceeds with a set of techniques (i.e., generative and phenomenological models at the individual, population, and combination of both levels). These techniques are used to address human mobility data landscapes and privacy concerns associated with such data.

Predictive Analytics These models are also applicable to specific human mobility-related issues. With a particular focus on the models presented in Section 4.2.3, they are capable of accurately predicting future displacements and locations visited by individuals, among other domains. Therefore, in this section, we calculate some of the fundamental representations used to characterize human mobility and extract relevant information from the data storage.

Interpretable and Explainable Intelligent Environment As previously mentioned, these XAI models are important for end-users to have a full understanding of the AI decision-making processes. The models proposed in Section 2.5.2 are crucial for promoting AI explainability while assessing the human mobility impact of using such algorithms, helping stakeholders understand the behaviours of AI models. models. By displaying positive and negative values, we intend for these models to generate real-time explanation model evaluations. This means that the *WalkingStreet* platform will generate feature attributions for model predictions, enabling users to visually explore model behaviour through interactive charts.

5.2.3.4 Sync Cloud Engine

This sync process works by aligning data, which is primarily (managed by the *Data Storage* micro-service, with a predefined set of queries, and subsequently transmitting this data to a cloud server. In addition, it can compare, analyze, and synchronize data schemas, streamlining group collaboration and refining query results. The synchronization process is also automated via a *Config file*.

As previously mentioned, this project has at least two distinct server environments: one local and one in the cloud. The same duality is mirrored in the database infrastructures. Initially, data is stored in a *MySQL* database and then stored in a cloud-based database. Thus, to facilitate the maintenance of the cloud database, it is advisable to conduct periodic synchronizations with the local database. This entails the centralization of data from numerous local databases into the cloud database.

From a synchronization methodology perspective, this micro-service assures data availability and accuracy. Over time, *MySQL* databases will have scenarios in which database schema and data should be synchronized with the cloud server. When this happens, the *SyncCloudEngine* tool can be used, configured with a specific time range for job execution.

In essence, this proposed micro-service is responsible for managing the new content to ensure synchronization with a centralized repository, like a cloud server. Moreover, it supports a synchronization mechanism that is based on timing and includes a transaction queue to address situation in which many users attempt to access the shared data concurrently. These approaches are designed to mitigate issues related to data consistency when multiple users and devices interact with shared data across diverse environments.

5.2.3.5 REST API

This container plays a crucial role in interacting with the *Data Storage* module. In other words, its primary responsibility is to facilitate the transmission of data from the *Data Storage micro-service* to the *Client*, without giving the user direct access to stored data content. Essentially, it serves as the server-side component of a *Front-End* End micro-service, encompassing all the aspects that the user doesn't see [210]. Moreover, it empowers the web application that is based on the *Front-End*. As shown in Figure 35, the communication between the *Front-End* and *Back-End* is implemented using a *REST API*.

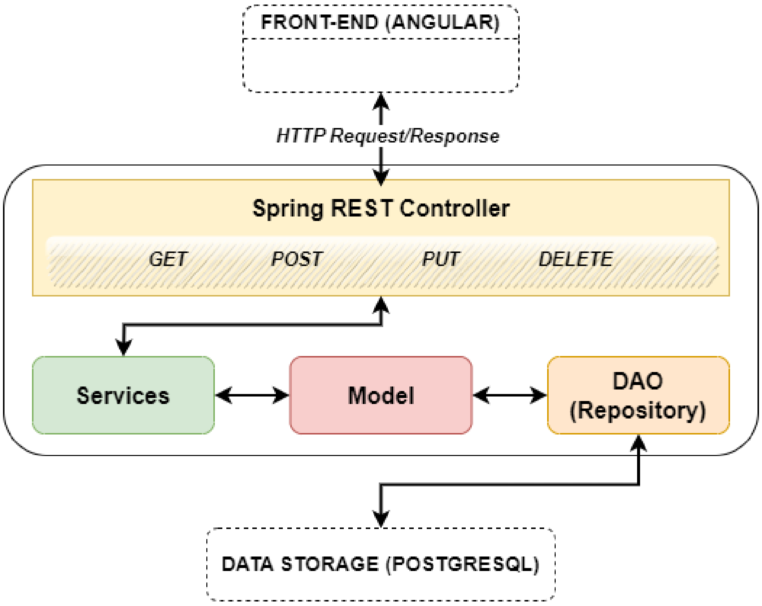

Figure 35: REST API application architecture with Spring Boot.

On the front-end side, HTTP Client communicates with the *Back-End*, which handles these HTTP requests through a *Spring REST Controller*. The *Angular* application on the front-end side creates a fully -functional User Interface (UI) that enables users to manage posts, including tasks like (adding, editing, and searching). Thus, this *REST API service* delivers relevant information about the health database. The requests are prepared to be translate into a format that clients can easily interpret, offering insights into metrics and predictions that have been previously calculated by the *AI Engine* and *XAI Engine*, in addition to other parameters that may be relevant to understand the remaining stored data in the *Data Storage*. All data is securely stored in the *PostgreSQL* database and is integrated into the application using a repository that is implemented from an abstraction of data persistence (i.e., Data Access Object).

5.2.3.6 Front-End

This micro-service hosts a website that functions as a back office for administration. It provides users with the tools to efficiently manage the stored data. The website is developed in *Angular*, and is tightly integrated with the *Back-End* are connected with each other through a *REST API micro-service* structured in *Java Spring Boot* code.

5.2.4 Cloud Module

In addition to the core module, the *WalkingStreet* platform includes a complementary module known as the *Cloud Server*. This module is hosted on the Google Cloud Platform and leverages the Internet to store, manage, process, and integrate data collection from different remote local server sources for future use. Additionally, the Cloud Server can facilitate the integration of several applications, optimizing cloud computing-related business processes. Real-time dataset integration is employed to facilitate data exchanges among the components illustrated in Figure 36 that are involved.

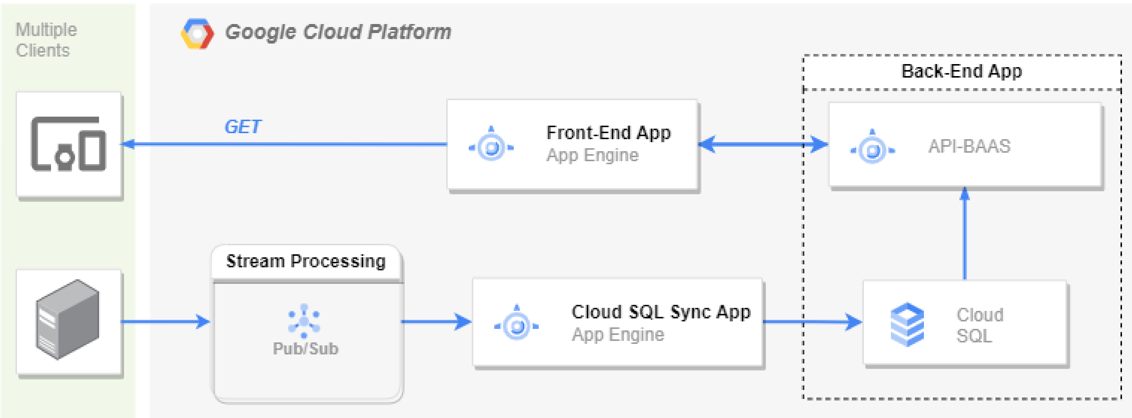

Figure 36: Cloud server architecture (central server transition).

The architecture of the *Cloud Server* is designed to integrates applications, improve workflows, and store insight-based data models. These models encompass predictions of future behaviour using AI and ML, as well as their explainability through XAI. This setup modernizes, the infrastructure of the platform, allowing to remain adaptable and scalable. Cloud-based servers offer virtually unlimited scalability in terms of increasing storage space and provide a stable environment for any component, in which any issue detected remoting doesn't affect another person co-renting the server. Furthermore, as a service offered by a third-party, users can access a series of resources accessible on the cloud from anywhere, enabling fully online operations.

5.2.4.1 Stream Processing

The *WalkingStreet* platform comprises independent modules, and to facilitate their development, deployment, and maintenance, the architecture of the cloud server incorporates several blocks. One of these key blocks is the Google Cloud Publish (Pub)/Subscribe (Sub) block, which serves to disseminate messages asynchronously across different parts of systems, specifically within a cloud server. In other words, the Pub/Sub allows messages to flow between different system components without these components needing to know each other's identities. Our simulator will read data from the events table in Cloud SQL and then publish messages to Cloud Pub/Sub. The architecture of this messaging system is shown in Figure 37.

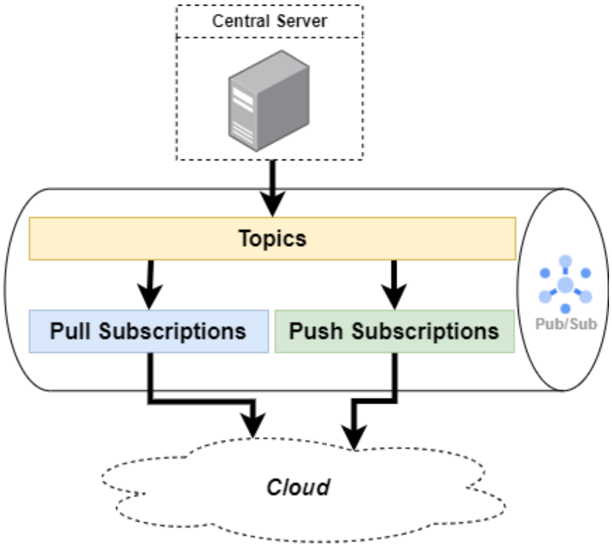

Figure 37: Google Cloud Pub/Sub architecture.

The Pub/Sub functionality is used for streaming processing and data integration pipelines in *Cloud Server* Architecture. It operates as a message-oriented middleware for services integration and establishes a framework involving event producers and consumers known as publishers and subscribers. Compared to older design patterns like message queuing and event brokers, Pub/Sub is more versatile and scalable. This stream processing tool handles data from multiple sources and streams it simultaneously to Cloud SQL databases.

5.2.4.2 Cloud SQL Sync App

This micro-service primarily focuses on synchronizing Pub/Sub information with Cloud SQL. From a structural viewpoint, this module contains a set of Continuous Integration/Continuous Deployment pipelines responsible for publishing events. These events are then subscribed to and acted upon by other services. However, its primary role lies in data processing.

When a *Local Server* places information on the *Data Storage* (hosted on the Central Server), the *Sync Cloud Engine* service comes into play. It synchronizes the information, carries out some preliminary processing, and then sends the event to Google Cloud Pub/Sub. Within the *Cloud Server* environment, the *Cloud SQL Sync* service, subscribed to the events generated by the *Sync Cloud Engine* service, starts the synchronizing process for the information with *Cloud SQL*. Thus, this micro-service extends cloud support for the existing modular data processing infrastructure. To summarise, the *Cloud SQL Sync* app is deployed on Google Kubernetes Engine (GKE), where it connects to a *Cloud SQL* database.

5.2.4.3 Cloud SQL

Cloud SQL of Google Cloud Platform (GCP) is a fully managed relational database service for PostgreSQL, boasting a wide range of configuration flags and developer ecosystems. One of the standout features that makes it an integral part of the Cloud Server architecture is its graphical User Interface. This interface simplifies the process of creating a database instance, including setting up a network firewall, as well as storing and managing data with a high level of security [54]. However, although Cloud SQL supports private connectivity through Virtual Private Cloud, it also allows public network access to database instances.

WalkingStreet architecture allows real-time data pipelines for Google Cloud SQL. These pipelines facilitate low-impact, real-time data ingestion from several on-premises to cloud systems. This process is executed and logged without causing downtime or data loss. It serves the purpose of offloading operational workloads to GCP, ensuring that data is available at the right time and in the right format for operational decision- making.

5.2.4.4 Back-End and Front-End App

As depicted in Figure 38, these services collectively create an infrastructure that allows web and mobile applications to connect to backend cloud storage (i.e., Cloud SQL service). This infrastructure offers a range of features such as user management, instant messaging, remote updating, and hosting, to name a few. In addition, it provides a highly flexible environment with minimal interface restrictions and a large stack of resources to build a variety of applications and appropriate widget integration. Therefore, the *Back-End* and *Front-End* services within the *Cloud Server* are characterized by their flexibility.

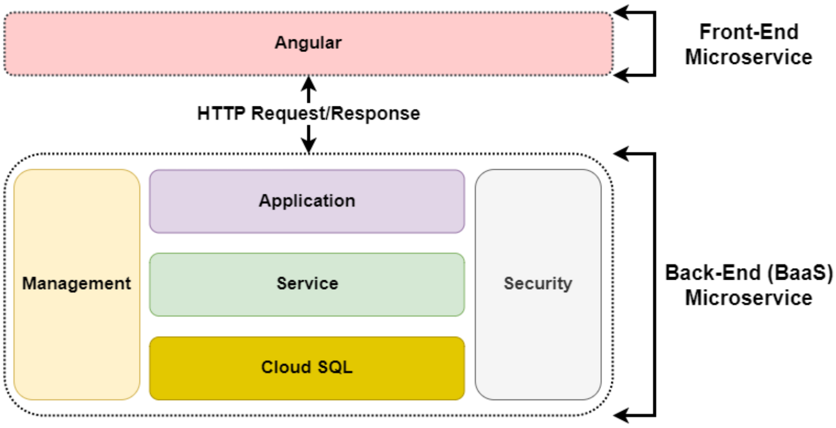

Figure 38: Back-End and Front-End architecture.

Inspired by Backend-as-a-Service (BaaS) service, the *API* micro-service serves as a third-party service that offers several advantages over Front-End development. It provides an additional layer of security during information exchange and manages infrastructures around data and resources. This cloud-based microservice operates at scale, bringing together the proper resources into meaningful sets. This allows for efficient application development without the need to source individual components from around the web and other information repositories. On the other hand, the *Front-End* micro-service is the user-facing part of the cloud computing architecture. This is where end-users interacts with cloud computing. It includes a web browser that allows users to access the data provided by the Back-End architecture as a whole. At the applicability level, the architecture can leverage serverless applications or Functions as a Service (FaaS), due to their auto-scaling capabilities, cost-effectiveness, and flexibility. However, as with other similar projects, the main components of this micro-service are made up of three parts:

- **User Interface**: In this component, end-users directly interact with the cloud system.s It encompasses all the elements that end-users access to send requests or perform any task in the Cloud Server.
- **Software**: This component runs from the user's side and often takes the form of a web browser application. In fact, it is the software that runs on the user's device to enable interaction with the cloud-based resources.
- **Client Device or Network**: These client-side components are integral to the front-end architecture and include the client-side hardware, such as the user's PC or other devices. These devices do, not require exceptional processing capabilities, as most of the heavy lifting happens on the cloud server.

The separation between *Back-End* and *Front-End* applications allow for scalability. When the system needs to migrate to a different database or change *API-based BAAS* endpoints, it can do so without disrupting data consistency. In addition, it enables the seamless switching of *Front-End* frameworks. Finally, this setup also ensures that cloud computing services can be accessed across multiple platforms, offering these services to end-users via public networks.

5.3 The WalkingStreet Platform: Implementation Plan

In the implementation phase of the *WalkingStreet* platform, we present a detailed description of the data flow, from the data capture process to the end user. This step is crucial as it transforms our project plan (Section 5.2) into actionable steps to achieve the goals of this thesis. In fact, it serves as the where we execute the planned strategy, outlining the tasks needed to complete the project, identifying necessary resources, and ensuring alignment with our strategic goals [109]. Therefore, while data capture is a standard procedure for many data sources, such as private or public platforms, tendering innovative human mobility products often requires careful consideration of the local, core, and cloud environments. This is essential to ensure the successful launch of this novel service, like MaaS, and its adoption by stakeholders. This is the main goal of our implementation plan.

5.3.1 Decentralized Module

As we mentioned in Section 5.2.2, this module represents the deepest node within the *WalkingStreet* architecture. In detail, it is characterized by its granularity, containing a network of devices and REST API services linked to the edge of the local server, web server software, databases, and various tools designed to handle remote requests. These remote entities are also supported by a solid IoT infrastructure, which is physically located elsewhere. Thus, the local server establishes separate connections to the services and machines, with their identities determined by configuring files. In essence, all configurations for connected clients are centrally managed, ensuring secure remote access to these services and devices.

In Section 5.2.1, the local server interconnects different types of smart applications (or devices), such as smartwatches, smartphones, or a Web Server hosted on a remote server (Section 2.2.1). These sensors and technologies create the infrastructure upon which the IoT thrives. They gather sensory information and monitor changes in local environments, measuring specific parameters related to their physical surroundings, and producing outputs, often in the form of electrical signals, for further processing. This section provides a fundamental understanding of which sensors and servers work and what roles they will fulfill within the local server. This local server hosts a *SiteWhere* (version 2.0) platform. Therefore, the following sections outline the implementation flow and communication between these diverse devices and this platform.

5.3.1.1 SiteWhere on Docker

SiteWhere 2.0 adopts Docker as its core deployment model, streamlining the installation of the application into a container for efficient resource deployment. In addition, Docker simplifies the process by allowing all micro-services to be deployed collectively, eliminating the need for manual individual deployments. For local installations, Docker Compose is a convenient tool, providing a straightforward method for deploying the entire service in one go, while also aggregating logs to facilitate easier debugging. Thus, the *SiteWhere 2.0* recipe acts as a starting point for building instances. Once cloned, one can access the *sitewhere-recipes/docker-compose* sub-directory, which contains configurations for the *SiteWhere* [1] deployment scenario.

5.3.1.2 Remote Server

In Figure 31, we discussed the integration of IoT devices into the *WalkingStreet* platform using the *SiteWhere* infrastructure. This infrastructure is hosted on a remote server, situated outside the scope of the elements that comprise the proposed thesis platform. The *SiteWhere* application is designed to manage

[1] Command line to clone of the SiteWhere repository: git clone `https://github.com/sitewhere/sitewhere-recipes.git`

extensive sensor deployments and facilitate the seamless addition of new devices. It introduces micro-services, where each one is a Spring Boot application encapsulated within a Docker container, specializing in a specific task. Many of the process pipelines are supported by Apache Kafka. In addition, certain micro-services can access external service functionalities solely only via well-defined API. Below is an example, presented in Listing 1, demonstrating the creation of a new device endpoint:

Listing 1: Request and response for the creation of a new device.

```
PATH: /devices
METHOD: POST
REQUEST BODY SCHEMA: application/json
{
  ”comments”: ”string”,
  ”deviceElementMappings”: [
    {
      ”deviceElementSchemaPath”: ”string”,
      ”deviceToken”: ”string”
    }
  ],
  ”deviceTypeToken”: ”string”,
  ”metadata”: {
    ”property1”: ”string”,
    ”property2”: ”string”
  },
  ”parentDeviceToken”: ”string”,
  ”removeParentHardwareId”: true,
  ”status”: ”string”,
  ”token”: ”string”
}

RESPONSE SCHEMA:
{
  ”comments”: ”string”,
  ”deviceElementMappings”: [
    {
      ”deviceElementSchemaPath”: ”string”,
      ”deviceToken”: ”string”
    }
  ],
  ”deviceTypeToken”: ”string”,
  ”metadata”: {
    ”property1”: ”string”,
    ”property2”: ”string”
  },
```

```
  "parentDeviceToken": "string",
  "removeParentHardwareId": true,
  "status": "string",
  "token": "string"
}
```

An alternative communication method supported by *SiteWhere* for handling device events is the MQTT protocol, using JavaScript Object Notation (JSON) format. In the example provided, payloads are published to the MQTT topic and transmitted as JSON as per the tenant configuration. The Listing 2 illustrates an example of sending a device location:

Listing 2: Registering device location.

```
{
  "type": "string",
  "originator": "string",
  "deviceToken": "string",
  "request": {
    "latitude": "double",
        "longitude": "double",
        "elevation": "integer",
        "updateState": boolean,
        "eventDate": "timestamp"
    }
}
```

The JSON object contains fields such as $type$, which provides a description of the device location, latitude, longitude, elevation of the location, and *updateState*, which, when set to true, updates the *DeviceState* of the assignment, as well as *eventDate* and associated metadata for the event. The Android app serves as an IoT gateway and/or client device for *SiteWhere*. If configured as an IoT gateway, it can send location and measurement events and register the Android device. When operating as an IoT client, it can receive real-time event pushes from *SiteWhere*. The configuration for pushing data to a specific device is done using server-side filters and Groovy scripts.

From an infrastructure perspective, this remote platform deploys and orchestrates micro-services from a set of Kubernete infrastructures, ensuring highly-availability in a distributed system. These kinds of infrastructures can be deployed on-premise or on various cloud services like Microsoft Azure, Amazon WebService, Google Cloud, or OpenShift. To simplify installation and configuration, Helm is used to supply both the micro-services and the necessary dependencies for a complete *SiteWhere* deployment. Additionally, there is an administrative application based on Electron that remotely manages *SiteWhere* instances.

5.3.1.3 Open-Source API Services

While the previous paragraph focuses on data communication elements between a remote device and a local server, this one groups a set of open-Source REST API designed to support the monitoring of human mobility conditions. Some of these services, as presented in Section 2.4.3.1, involve sharing data either through nodes that consolidate information at a remote server. This server acts as a centralized point, connecting local sensors and hosting a REST API Infrastructure.

As we mentioned in Section 5.2.2.2, other experimental environments can be an opportunity to engage citizens in the human mobility information that is used by multiple stakeholders. One such environments has been the subject of study in various scientific articles. We believe that every citizen can benefit from Open Data and Open Data can benefit every. An example of this is the New York City (NYC) Open Data service, managed by the Open Data Team at the NYC Office of Technology and Innovation. This service offers a wealth of datasets covering many aspects of life in New York City, ranging from the New York Police Department (NYPD) to the Metropolitan Transportation Authority (MTA). These datasets are accessed through the Socrata Open Data Application Programming Interface (SODA). This API identifies and retrieves data from the NYC OpenData platform by, applying filters stacking parameters in the Uniform Resource Locator (URL) and Socrata Query Language (SoQL) queries to search and change the results. For example, one can filter the Bicycle Counters dataset using a query similar to Listing 3.

Listing 3: Parameters and response for bicycle counts.

```
DATA_URL: ”data.cityofnewyork.us”
DATA_SET: ”uczf-rk3c”
APP_TOKEN: ”app token created”

RESPONSE SCHEMA:
[
  {
    ”countid”: ”10014848”,
    ”id”: ”100009425”,
    ”date”: ”2022-06-24T00:00:00.000”,
    ”counts”: ”15”,
    ”status”: ”0”
  },
  {
    ”countid”: ”10014849”,
    ”id”: ”100009425”,
    ”date”: ”2022-06-24T00:15:00.000”,
    ”counts”: ”12”,
    ”status”: ”0”
  }
]
```

5.3.2 Core Module

This module is built around a single Central Server hosting a Docker that runs multiple services. If this server goes down, the Docker goes down with it. This centralized infrastructure manages critical functions related to human mobility data processing. In addition, several types of REST services can connect to this master server. These applications are created using different architectures, with one of them being the micro-service architecture.

In this phase, AI and XAI models are other components developed for containers as micro-services. These applications, isolated at the micro-services level, function as independent, interconnected, and scalable providers. This design ensures that the failure of one service does not impact the functionality of any other service. This way, DL models (Section 4.2.3.1) analyze data patterns, while XAI (Section 2.5.2) aids in helping humans understand the decisions or predictions made by the AI. This approach enhances resource management efficiency and ensures the stability and robustness of the micro-services.

As described in Section 5.2.3, client systems and users cannot directly access resources without first passing through the back office. They interact with a platform that has a separate database, but this server relies on the micro-service for tasks like clock synchronization, identity management, authentication, and other vital micro-services. Subsequently, the stored information will be remotely stored in the cloud. The following sections provide detailed descriptions of all these components.

5.3.2.1 Sync API Engine

Once there are various open-source solutions available, we need to integrate them into the same job *job scheduler* component within the *WalkingStreet* architecture. In other words, this *"job scheduler"* is responsible for triggering events at predefined time intervals, and it must be fault-tolerant and scalable. Developed in $Python$ language as part of this work, these scheduled *Jobs* execute at specific times or on a recurring basis using *Unix style* expressions called *cron*. The Listing 4 shows the necessary imports and key services involved in the job synchronization process.

Listing 4: Syncronization API service.

```python
# import the necessary packages
from sodapy import Socrata
from database import DataBase
from apiService import ApiService

class SyncApiService:

    def __init__(self):
```

```
        self.database = DataBase()
        self.apiService = ApiService()

    # other methods needed
```

This job application handles the scheduling of routine tasks, including system maintenance and the synchronizing of remote API and device data, all configured through a central Configuration file. Moreover, from the same file, we can define the synchronization intervals, integrating various asynchronous services, and avoiding any arbitrary timeframes. Thus, to orchestrate all these parameters and primary settings at this level of the *WalkingStreet* platform architecture, we integrate the *Configuration file* presented in Listing 5.

Listing 5: Configuration file of local server.

```
[
   ”nyc-open-data” : {
      ”config”: {
         ”url”: ”https://data.cityofnewyork.us/”,
         ”token”: ”askljfhlashflasfipiooiyoyqwuoryuioqywroyoiqwroiyoy”
      },
      ”paths”: [
         {
            ”path”: ”browse?Dataset-Information_Agency=311”,
            ”type”: ”cron”,
            ”hour”: 24
         }
      ]
   },
   ”open-data-hub-mobility”: {
      ”config”: {
         ”url”: ”https://mobility.api.opendatahub.com/v2/tree,node/”
      },
      ”paths”: [
         {
            ”path”: ”Bicycle/*”,
            ”type”: ”cron”,
            ”hour”:24
         }
      ]
   }
]
```

5.3.2.2 Data Storage

This *WalkingStreet* platform node is responsible for synchronizing data with a database, such as a *Data Storage* container, while ensuring that mapped fields and relationships remain synchronized. To maintain data consistency, we aim to synchronize daily, although the exact frequency may vary depending on the scalability of the external system's data source (i.e., REST API or remote device). During synchronization, an automated check is performed to determine which program contains the most recent updates for linked records. Any new records are then added accordingly. These measures are in place to prevent the loss of new records in either or both of the programs.

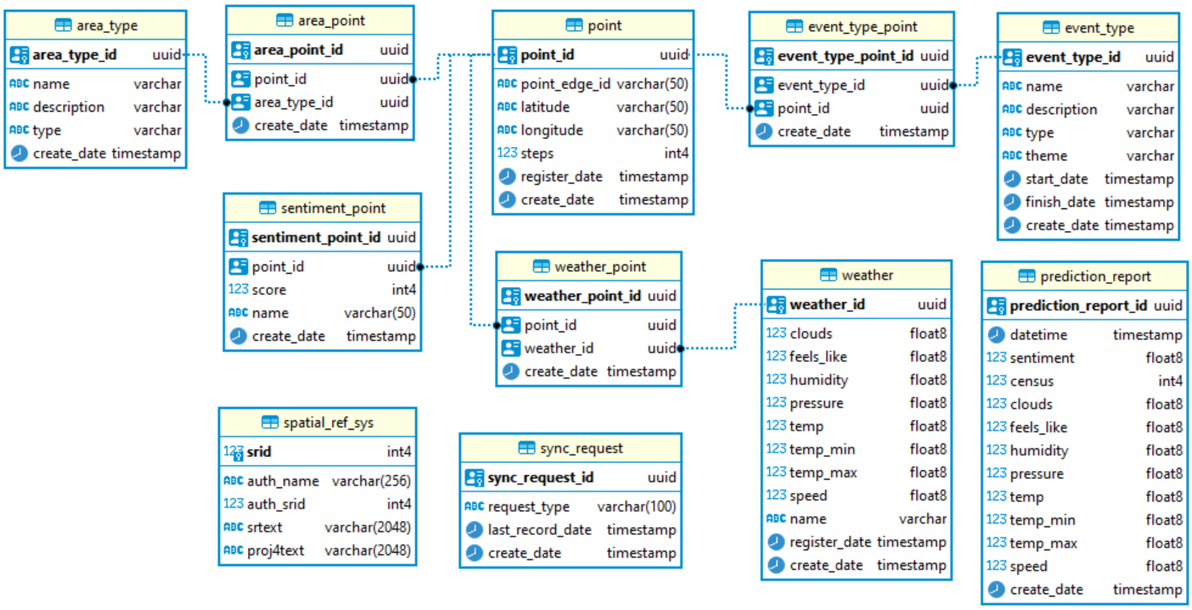

Figure 39: Central Server basic Entity-Relationship diagram.

As shown in Figure 39, the data model has been built using *PostgreSQL*. This open-source object-relational database system is renowned for its stability and is deployed using Docker. It is well-suited for handling routing, optimizing queries, and storing complex spatial human mobility data topologies. Additionally, *PostgreSQL* benefits from continuous community contributions, ensuring ongoing performance enhancements through the release of new versions.

5.3.2.3 AI and XAI Engines

Developed in Python, AI and XAI are a set of models designed to execute specific tasks, adhering to the principles and architecture of micro-services. While AI algorithms, particularly DL, are evolving thanks to improved data collection methods, data training sets, advanced data processing mechanisms, and enhanced analytic techniques, the Interpretable AI models play a crucial role in addressing "explainability" within the central server platform. In both cases, the horizontally layered platforms provide a structural foundation, but it's the vertical components that effectively tackle domain-specific problems. Key parts of the AI and XAI elements are presented in Listings 6 and 7, respectively.

Listing 6: Syncronization AI service.

```python
# import the necessary packages
from database import DataBase
from aiService import AIService

class SyncAIService:

  def __init__(self):
    self.database = DataBase()
    self.aiService = AIService()

  # other methods needed
```

Listing 7: Syncronization XAI service.

```python
# import the necessary packages
from database import DataBase
from xaiService import XAIService

class SyncXAIService:

  def __init__(self):
    self.database = DataBase()
    self.xaiService = XAIService()

  # other methods needed
```

In both the DL and XAI architectures, the vertical components follow a modular and layered structure. Concerning the layers of the DL engine, the first layer is responsible for reading the data, which includes tasks such as data processing, aggregation, and transformation. In the next layer, ML models, as presented in Section 4.2.3.1, are applied to the processed data to perform various tasks. The final layer involves storing the outcome generated by the previous layer in the *Data Storage* service. As for the XAI engine, it also comprises three layers. The base layer reads the stored outcome from the previous architecture. Then, this retrieved data is used to calculate the models outlined in Section 5.3.2.3. Finally, in the third layer, the results are stored in the *Data Storage*, ready to be later consumed by a job.

5.3.2.4 Sync Cloud Engine

This engine syncs the data from the *Central Server* to the *Cloud Server*. From a Job service. It transfers new data to the cloud using a class we have named *SyncCloudService* class. Developed in Java Spring (version 4.0), which is developed in Java Spring (version 4.0). This container enables asynchronous processing jobs, and a recommended tool for this kind of service with the required features is Google Cloud Publish (Pub)/Subscribe (Sub) [80]. In addition, given the potentially large flow of messages exchanged between the Central and Cloud environments, the design of Pub/Sub is focused on high reliability and scalability.

In this work, we use the *Spring Cloud Stream* infrastructure to implement the concept of publish-subscribe messaging. This infrastructure simplifies the use of messaging technology across different platforms and provides predefined annotations for declaring bound input and output channels. Moreover, it also applies the *@EnableBinding* annotation to both of the application's configuration classes, such as OutputChannel and InputChannel. Thereby, the micro-service that adopts this infrastructure is prepared to trigger the send or receive message methods. The Listing 8 contains the class which represents the *cron job* for sending messages between the micro-service and the cloud via an output channel.

Listing 8: Message sender service of synchronization Cloud micro-service.

```
import walking.street.app.delivery.OutputChannels;
import walking.street.app.service.MessageSenderService;
import walking.street.app.service.DatabaseService
import org.springframework.scheduling.annotation.Scheduled;
import org.springframework.stereotype.Component;

@Component
public class AIMessageSink {

   @Autowired
   private DatabaseService databaseService;

   @Autowired
   private MessageSenderService messageSenderService;

   @Autowired
   private OutputChannels outputChannels;

   @Scheduled(fixedDelay = 5000)
   public void send() {

      //get data from DatabaseService
      AIMessage message = databaseService.getAIMessage()

      //send AI message
      messageSenderService.send(outputChannels.aiMessageOutput(), message);
   }
}
```

To publish messages from a message channel to a Pub/Sub topic, we use the outbound channel adapter. The outbound channel adapter converts Plain Old Java Object (POJO) into Pub/Sub messages and then sends the messages to a Pub/Sub topic. The Listing 9 presents the implementation of the outbound channel adapter used in this work, which sends messages from the output message channel to either the *”aiMessageOutput”* or *”xaiMessageOutput”* topics.

Listing 9: Output channel (publisher) of synchronization Cloud micro-service.

```
import org.springframework.cloud.stream.annotation.Output;
import org.springframework.messaging.MessageChannel;

public interface OutputChannels {

   @Output(”aiMessageOutput”)
```

```
  MessageChannel aiMessageOutput();

  @Output("xaiMessageOutput")
  MessageChannel xaiMessageOutput();
}
```

In addition to the output message channel, Spring Cloud Stream supports configuration for bindings and binders via a Yet Another Markup Language (YAML) file.These binding properties follow the format *spring.cloud.stream.bindings.<channelName>.<property>=<value>*. Here, *<channelName>* represents the name of the configured channel (e.g., output or input), and *<property>* must be filled with a specific set of properties like *destination*, *group*, *contentType* and *binder*. However, input bindings should be prefixed with *spring.cloud.stream.bindings.<channelName>.consumer*, including binding properties such as *concurrency*, *partitioned*, *headerMode*, *maxAttempts*, *backOffInitialInterval*, *backOffMaxInterval*, *backOffMultiplier*, *instanceIndex* or *instanceCount*. Conversely, output bindings should be prefixed with *spring.cloud.stream.bindings.<channelName>.producer* and can include binding properties such as *partitionKeyExpression*, *partitionKeyExtractorClass*, *partitionSelectorClass*, *partitionSelectorExpression*, *partitionCount*, *requiredGroups* and *headerMode*. The Listing 10 shows the various binding properties specified within the YAML file of this micro-service.

Listing 10: YAML file of synchronization Cloud micro-service.

```
spring:
 profiles:
  #profiles parameters
 datasource:
   #database configuration parameters
  cloud:
   stream:
     default-binder: pubsub1
     bindings:
      aiMessageOutput:
       destination: ai_message_output
       content-type: application/json
       binder: pubsub1
      xaiMessageOutput:
       destination: xai_message_output
       content-type: application/json
       binder: pubsub1
 binders:
  pubsub1:
   type: pubsub
```

5.3.3 Cloud Module

This module centralizes all information from multiple Central Servers. As discussed in Section 5.2.4, it allows for the creation of multiple Pub/Sub clients during the Stream Processing step. In this step, Java uses separate client classes for each Pub/Sub server connection, which are then used to create clients for both publishing and subscription. Thus, stream processing enables applications to collect and process data immediately as it's generated.

As mentioned in Section 5.2.4, this module also leverages the Java Spring language for micro-services, Kubernetes, and the cloud using Pipeline. For an application server, the GCP lets Kubernetes resize clusters based on different metrics and resource requests. At the micro-service level, the Cloud Server also provides a listener annotation to convert incoming messages without the need to specify the content type of an input channel. During the dispatching process, methods with this annotation automatically apply the message where it is required.

After passing through the preceding architecture components, the data is stored in the database service for later access by all end users. This access is only possible from the other two platforms within the cloud architecture: Back-End and Front-End. These cloud infrastructures support everything that the end-user interacts with, forming an essential part of how end-users connects to the cloud computing infrastructure. Subsequent sections provide detailed insights into the connectivity and functionality of these cloud components for web applications.

5.3.3.1 Stream Processing and Cloud SQL Sync

This section includes a Spring Cloud application with stream processing, operating within thr Spring Cloud Kubernetes environment, which is advantageous for building cloud-native applications. Additionally, it also effectively manages the available messaging service, which automatically scales with demand and is known as Google Cloud Stream. To understand the Cloud SQL Sync service, we must familiarize ourselves with four key components: topics, subscriptions, subscriber clients, and Cloud SQL.

This service receives messages from subscriptions using Spring Cloud Stream, which spans across different processes. The typical scenario for Stream involves the creation of multi-application pipelines, where micro-service applications exchange data with each other. As demonstrated in Listings 9 and 12, this scenario establishes a correlation between the output and input destinations of adjacent applications. Its turn, Listing 11 presents the design of Stream design responsible with the Time Source application (output) sends data to the Log Sink application (input), referencing a common destination named $ai\_message\_output$ (AI message) or $xai\_message\_output$ (XAI message) for bindings within both applications.

Listing 11: YAML file of synchronization Cloud SQL micro-service.

```
spring:
```

```yaml
profiles:
  #profiles parameters
datasource:
  #database configuration parameters
cloud:
  gcp:
    logging:
      enabled: true
    pubsub:
      enabled: true
      project-id: walking-street-app
      credentials:
        location: file: #credential file location
      stream:
        default-binder: pubsub1
        bindings:
          aiMessageInput:
            destination: ai_message_output
            group: ai_message_output_consumer
            content-type: application/json
            binder: pubsub1
            consumer:
              concurrency: 20
          xaiMessageInput:
            destination: xai_message_output
            group: xai_message_output_consumer
            content-type: application/json
            binder: pubsub1
            consumer:
              concurrency: 20
  binders:
    pubsub1:
      type: pubsub
```

The Listing 12 introduces the *AIMessage* class, one of two receiver classes in the synchronization Cloud SQL micro-service (the other being *XAIMessage*). It uses the *@StreamListener* and *@Payload* annotations to handle incoming messages by specifying the content type of the input channel, such as *aiMessageInput*. Moreover, during the dispatching process, the argument of the *receive* method's argument will be automatically populated with POJO containing the unmarshalled form of the JSON String. In other words, this application processes incoming messages with a String content and a content-type header of *application/json* are received on the input channel. The following application exemplifies a sink application that receives external messages:

Listing 12: AIMessage receiver of synchronization Cloud SQL micro-service.

```
import org.springframework.cloud.stream.annotation.EnableBinding;
import org.springframework.cloud.stream.annotation.StreamListener;
import org.springframework.messaging.handler.annotation.Payload;

@EnableBinding(InputChannels.class)
public class AIMessageSink {

  private static final String AI_MESSAGE_INPUT = ”aiMessageInput”;

  @StreamListener(AIMessageSink.AI_MESSAGE_INPUT)
  public void receive(@Payload String message) {
    #save the 'message' on Cloud SQL (PostgreSQL relational database)
  }
}
```

After the applications receive external messages, they are stored at the expense of a relational database, such as PostgreSQL Database. However, as previously discussed in Section 5.2.4.3, we use third- party software to run a SQL database provided by the GCP. One notable feature of this platform is the Cloud SQL environment. Which offers a straight-up PostgreSQL relational database. In essence, the PostgreSQL functionality provided by a Cloud SQL instance is on par with that of a locally-hosted PostgreSQL instance. This setup optimizes database operations, allowing more time for application-related tasks.

5.3.3.2 Back-End and Front-End App

In this section, we will outline the development of the Web Back-End in three steps: choosing the framework, programming, and building the Back-End and Front-End. Throughout this process, these components are hosted on two distinct micro-services. As already mentioned in Section 5.2.4.4, during the framework selection phase, we opted for Java Spring for the Backend-as-a-Service (BaaS). Spring offers a comprehensive platform with various components highly helpful for the Front-End application. For example, Spring security is instrumental in robust access control, while Spring data aids in connecting with databases. For the Front-End, we chose the open- source Angular framework.

Back-End App To start, we initiated a Maven Project for the Back-End application. Maven is a tool that automates the build process of the Back-End application. Additionally, dependencies are later manually added to the project. Then, within the Java Spring context, we selected Spring Boot 4.0.0 version. Basically, we use spring to program the application and Spring Boot to handle the application's runtime intricacies. The Front-End project is developed using TypeScript, a JavaScript-based language. TypeScript eliminates unnecessary features and code, resulting in a faster application.

With the frameworks for the Back-End and Front-End micro-services in place, we proceed to program them in separate subparts. In the Back-End, we begin by adding the following libraries to develop web functionalities in the *POM.xml* file of the Java Spring project:

- spring-boot-starter-web
- spring-boot-starter-test

The inclusion of spring-boot-starter-web provides us with additional libraries, including dependencies like spring-boot-starter, jackson, spring-core, spring-mvc, and spring-boot-starter-tomcat.

In the next step, we define endpoints, often referred to as controllers in the Spring Web MVC Framework. These controllers fetch data by referencing the model and provide the viewer with the necessary information for the Front-End micro-service. In other words, these endpoints essentially act as controllers, responding to web-requests. Thus, based in Listing 13, we define a set of endpoints for the main controller, named by *AIMessageController*, in the Back-End micro-service.

Listing 13: AIMessageController class of Back-End micro-service.

```
import org.springframework.web.bind.annotation.GetMapping;
import org.springframework.web.bind.annotation.RestController;

@Api
@RestController
@RequestMapping(value = ”/api/walkingstreet/aimessage”)
@Validated
public class AIMessageController {

  @Autowired
  private MobilityAreaService mobilityAreaService;

  @GetMapping(value = ”/{areaKms}/{rangeDates}”)
  @ResponseStatus(HttpStatus.OK)
  public MobilityAreaResponse getMobilityArea(
    @PathVariable(”areaKms”) Integer areaKms,
    @PathVariable(”rangeDates”) String rangeDates) {
      return mobilityAreaService.getMobilityArea(areaKms, rangeDates);
  }
}
```

First, to make this micro-service function with data, it must be connected to a database. As detailed in Section 5.3.3.1, for the sake of simplicity and to maximize the potential of GCP, we will work with *Cloud SQL* for the relational database called PostgreSQL. Here are the steps involved:

Step 1: Create a Repository. Start by creating a repository to interact with the PostgresSQL database. Using Spring's features, we simply create an interface that extends the *BaseRepository*.

Step 2: Create a Service. After addressing the database component of the micro-service, create a new class as a service and annotate it with *@Service* annotation. As Listing 13 shows—using the annotation *@Autowired* to make sure the Spring runtime an instance—a service, named MobilityAreaService, was implemented in the Back-End application. This is where business logic happens and works with data stored on *Cloud SQL*,using the repository as a private property. In this specific case, the service retrieves information about the density of human mobility in a specific area and within a range of dates.

Step 3: Declare Controllers. Declare classes such as *AIMessageController* class, *XAIMessage-Controller* as controllers for the micro-service. Annotate these controllers with *@RestController*. This annotation, a specialization of the *@Controller* annotation, simplifies the setup of the application, avoiding explicitly the need to use the *@ResponseBody* annotation for endpoints. Furthermore, these controllers have an annotation called *@RequestMapping*. For example, in the case of *AIMessageController* case, all endpoints are mapped at *"/api/walkingstreet/aimessage"*. Then, we create a set of endpoints. As shown in the previous listing, we create a method, the path and a get request under *getMobilityArea* to be mapped to the *"/areaKms/rangeDates"* variables and the *@PathVariable* annotation in the method header. This endpoint is responsible for filtering the predictions within a radius of *areaKms* kilometres and a range of *rangeDates* dates.

Step 4: Implement Security. To ensure the security of endpoints, implement a security service in the Back-End. This can be achieved using a configuration class called *WebSecurityConfiguration* and adding the dependency called *spring-boot-starter-security* dependency to the *POM* file. In the configuration class, use the *@Configuration* annotation from Spring to define access management. Within this class, configure *HttpSecurity* (an instance of HTTP) in the *configure* method. This method helps secure the endpoints. There's an example of how the configure method from the *WebSecurityConfigure* class might look in Listing 14.

Listing 14: WebSecurityConfigure class of Back-End micro-service.

```
@Configuration
@EnableWebSecurity
public class WebSecurityConfig extends WebSecurityConfigurerAdapter {

  @Autowired
  private LoginSuccessHandler loginSuccessHandler;

  @Override
  protected void configure(HttpSecurity http) throws Exception {
      http
      .authorizeRequests().antMatchers("/api/**")
```

```
   .authenticated()
   .and()
    .httpBasic()
   .and()
    .exceptionHandling()
    .authenticationEntryPoint(restAuthenticationEntryPoint)
   .and()
    .formLogin()
    .loginProcessingUrl("/api/login")
     .successHandler(loginSuccessHandler)
     .failureHandler(new SimpleUrlAuthenticationFailureHandler());
 }
}
```

Regarding security, it's important to note that the request endpoints need to be authorized under *"/api/"*. In this configuration, any HTTP request is authenticated using a HTTP-basic authentication. Otherwise, if a user doesn't have the necessary privileges, the entry point should send back a *401 Unauthorized* response. Additionally, to facilitate user sessions and avoid the need for credentials with every request, it's common to use tokens. These tokens can be automatically set in the header of HTTP requests after the initial authentication. This allows for easy identification of each user session. The *Front-End* micro-service can then use the token stored in the header, often in the form of a cookie, to maintain user sessions without requiring users to repeatedly enter their credentials for each request.

Front-End App Before building the *Back-End* micro-service, it is essential to develop a *Front-End* application. This application is hosted on the Google Cloud Platform. There are many examples of the successful building of this paradigm of application [186]. Its resources are deployed, scaled and updated on a third-party cloud service, spending a smaller amount of time configuring and stress-testing applications. To do this, the following processes should be implemented to integrate an *Angular* application into the cloud.

This *Front-End* application is executed using the GKE Frameworks. These frameworks are capable of running *Angular* and versions of the *NodeJS* version at runtime. At the architectural level, these frameworks function as a container orchestration system. They allow you to containerize both the *Back-End* and *Front-End* applications using declarative configurations specified in YAML files. For a more detailed view of how the *Front-End* development version is controlled, you can refer to the YAML file provided in Listing 15.

Listing 15: YAML file of Front-End micro-service on Cloud Kubernetes.

```
apiVersion: apps/v1beta1
kind: Deployment
metadata:
  name: front-end-development
spec:
```

```
replicas: 1
template:
metadata:
labels:
  label-key : label-value
spec:
containers:
  - name: front-end-development-container
    image: inyee/walking-street-app-demo:v1
    imagePullPolicy: Always
    ports:
    - containerPort: 80
```

To deploy the Angular application in the Kubernetes environment, the deployment manifest for angular is listed above. Even before deploying the application, we can modify the manifest file, named *walking-street-development.yaml*. For instance, we can change the name of the Deployment, adjust labels, specify the Docker registry, and set the image tag accordingly. Additionally, two other YAML files, namely *walking-street-service.yaml* and *walking-street-load-balancer-service.yaml*, were created to facilitate internally and external access to the application within the Kubernetes cluster.

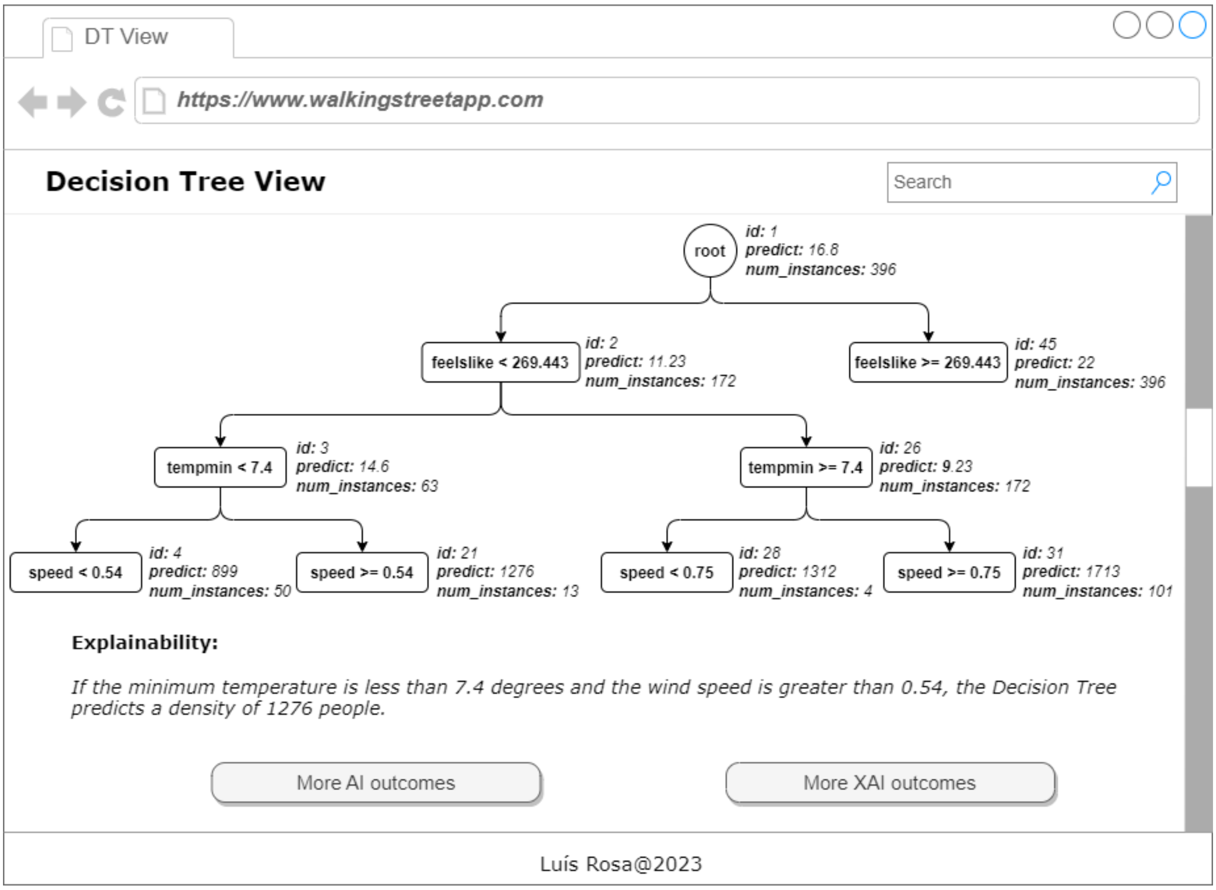

Figure 40: Element of the Front-End application.

With all this information, Figure 40 shows part of the Front-End application with the help of the *Angular* framework. This kind of application is a complex client-side application where a significant portion of the code runs on the client's web browser. As mentioned in Section 5.2.4.4, this approach offers several advantages. The basic building blocks of the Angular framework are Angular components that are organized into *NgModules*. *NgModules* group together related code into functional sets, and an Angular application is formed by a collection of these *NgModules*. Every application has at least one root module that facilitates bootstrapping and often includes multiple feature modules.

Based on the same figure, the *Human Mobility XAI* section of the application is responsible for explaining predictions related to human mobility phenomena in different scenarios. In this scenario, users can select a set of settings to see the multiple predictions generated by a Decision Tree (DT) model stored on *Google Cloud SQL*. The model consists of various decision nodes, each characterized by a field and threshold values. The selected information is presented in a straightforward text format on the web platform. The final node indicates whether there will be a concentration of people in the NYC center at a specific hour. The user's decision path through the model is highlighted in black. With the decision tree model, it is possible to identify the factors contributing to the density of people in NYC, as well as the associated threshold values. For example, if the user's decision path goes through a node marked as "average density level"with a threshold value of 1200, this suggests that a density level of more than 1200 in the city center is likely to attract citizens. This information can be valuable for advising authorities on managing and potentially reducing population density to levels below 899.

5.4 Summary

The architecture presented in this archetype is the result of a collaborative effort involving laboratory experiments and contributions from master's students. Through these experiments, we have designed an infrastructure capable of integrating new data sources, combining multiple devices, and enriching the dataset related to human mobility. Additionally, master's students have also played a crucial role in the course of this doctoral program. They were dedicated to researching different testing scenarios and disseminating scientific knowledge about the *WalkingStreet* Platform. These combined efforts have resulted in a robust architecture that aligns with both current and future human mobility demands of citizens, namely trends of human spatiotemporal movement.

The *WalkingStreet* infrastructure is composed of a set of components that are instrumental in transforming cities as they help to enhance the quality of human mobility, improving the performance and interactivity of urban services, and optimizing resource allocations. These components urge in sophisticated computing structures and operate within intelligent environments that can interact with smartphones or smartwatches like data sources. The combination of intercommunication technologies with an affordable distributed system that acts both locally, centrally and in cloud environments, enables the platform to collect and provide data for a wide range of applications. Devices, whether installed or carried by individuals, capture data and send it to the local server using IoT services such as *SiteWhere* platform. This local server employs basic functions such as data collection and synchronization with the central server. This central node, an approach developed within the *WalkingStreet* platform, facilitates the scalable transfer of human mobility data to a cost-effective cloud environment, striking a balance between local control and the advantages of storing data in larger databases. Users can interact with this central component using a Web application, and the cloud serves as the ideal environment for continuously storing the data transmitted via the central server. This data includes predictions, metrics related to human mobility, and other features, all of which can be accessed through a responsive browser on devices like smartphones.

The *WalkingStreet* platform is designed to actively engage users as contributors to improving the quality of mobility in their community. It goes beyond simply observing and analyzing citizen movements in a SC; it encourages user participation by responding to citizens' needs and feedback. In other words, this strategy involves users actively interacting with devices, whether individual devices or those integrated into city infrastructure, and providing instantaneous and real-time feedback. This user involvement is complemented by suggestive algorithms, which assist authorities in enhancing or developing existing smart city products while empowering citizens to propose strategies for improving mobility in their city.

One of the key contributions of the *WalkingStreet* platform is its adaptability to address human mobility challenges in urban environments. The presented case study shows that within the technology ecosystem, it is feasible to provide MaaS with a focus on assessing and understanding human mobility behaviour. The platform not only produces results from laboratory experiments but also houses AI that are valuable assets for addressing mobility-related issues.

Chapter 6

Discussion of Results

In this chapter, we present prototypes designed to demonstrate the versatility of the proposed architecture, addressing a variety of challenges related to human mobility. These projects are also equally important to instill confidence that the components that make up this proposed service can effectively tackle issues even on a smaller scale. Additionally, these prototypes reveal the diverse facets of human mobility, spanning from predictions based on cognitive interactions to infrastructures that analyze the number of people located in a certain location of a smart city.

Furthermore, we emphasize the interconnected nature of this architecture, showcasing its bidirectional nature. Our aim is to demonstrate that several scenarios are possible to implement, ranging from the evocation of Application Programming Interface (API) and external services to the capillary integration of a heterogeneous set of sensors and mobile devices. On the other hand, we describe the project's evolution, highlighting its substantial potential impact among stakeholders within Smart City (SC). At the same time, over the course of these 5 years, we have collaborated with diverse scientific entities to establish the merits of this thesis as a recommended solution for solving the multifaceted challenges related to human mobility.

6.1 Prototypes and Case Studies

Within the context of this thesis, relevant studies take the form of prototypes. This section provides a brief overview of the development of experiments and highlights the advantages and disadvantages of the employed approaches. Additionally, it delves into the preliminary studies carried out on each module constituting the *WalkingStreet* platform architecture.

6.1.1 Emotion-based Recommendation Platform

The objective of this experimental case study is to ascertain the validity of an event recommendation application based on a time series prediction model. This model aims to predict a specific emotion based on factor such as event location and type, and other user reviews. Our focus is investigating whether it is reliable to use an emotion feature database to develop an event recommendation engine.

The model's dataset includes weather information, sourced from the Braga/Porto region, collected through the *rp5.ru* website [167]. The training data used by the model corresponds to the period between January 2020 and November 2021. Out of all the information collected, only specific columns related to temperature, atmospheric pressure, humidity, and precipitation level were kept. In terms of data processing, unnecessary columns were removed, and missing values were cast to the double data type. Additionally, the date information for each entry was split into different columns: hour (each entry corresponds to one hour), month, day of the month, and day of the week.

Another component of the dataset pertains to events that occurred within the Braga/Porto region. The data was collected through the *dados.gov.pt* platform [160]. Initially available in Extensible Markup Language (XML) format, the data was altered to Comma-Separated Values (CSV) format for processing. A notable challenge with this data was that despite listing various place names for each event, the provided coordinates remained constant. To rectify this, manual efforts were undertaken to identify and replace coordinates for each specific location. Creating separate training datasets for chosen locations required isolating weather and event data for each place. This was also done manually by merging weather and event information from each distinct place into its own individual dataset.

Due to the lack of available data on this topic of emotions associated with specific locations, it became necessary to fabricate fictional data. Given the project's focus on creating a Proof of Concept (PoC), this was deemed acceptable for assessing its feasibility. Since there was no viable emotion dataset available for exploration, a new dataset had to be devised to answer the project's questions, while adhering to several key considerations to ensure the operational and reasonable recommendation system is the main goal:

- **Richness:** A scarce dataset would have an adverse effect, there wouldn't be enough data to train the Machine Learning (ML) model and, thus, assure its effectiveness;
- **Categorical Values:** A nominal feature is needed for the sake of the model's prediction of emotion concerning specific places/events;
- **Correlation:** An association with this new fictional data had to be created to promote the most realistic prediction possible since a weak or non-existent correlation would undermine this proof of concept's effort. Emotion can't be predicated if there's zero to no association with the other features: the model's ability to predict emotions relies on understanding the relationships between the nominal variable and other attributes.

- **Distinguishable Emotions:** To facilitate straightforward user assessment, five distinct emotions were selected: Sad, Boring, Happy, Excited, and Overwhelmed. This choice was made to better resonate with users and ensure a intuitive rating system.

The previous distinguishable emotions were assigned numerical labels, forming a scale that represented a continuum of excitement levels. This approach aimed to facilitate handling the data as well as enhancing its compatibility with the model. The chosen scale was the following: 1 - Boring; 2 - Sad; 3 - Happy; 4 - Excited; 5 - Overwhelmed.

The final architecture of the dataset contained various columns, each corresponding to distinct attributes such as temperature, pressure, relative humidity, precipitation level, hour, month, day of the month, day of the week, Boolean value for whether or not an event is occurring, and emotion. Each entry in the dataset corresponds to a single hour. In order to train the Long Short-Term Memory (LSTM) model across the chosen locations, a separate dataset was used per location. The model was tailored to each dataset individually. The Artificial Intelligence (AI) models used for time series prediction were LSTM variations. The model was implemented in Google Colab.

Of all the available LSTM variations, four were selected for experimentation: the basic LSTM, the Stacked LSTM, a Convolutional Neural Network-Long Short Term memory (CNN-LSTM) hybrid with a one-dimensional convolutional layer, and the Encoder-Decoder LSTM. Throughout the training process, in order to measure the model performance evolution the main loss metric used was Root Mean Square Error (RMSE). This metric, often abbreviated as RMSE, calculates average distance between the model's predicted values and the actual values in the dataset. After running LSTM and Stacked LSTM algorithms, we find the following results:

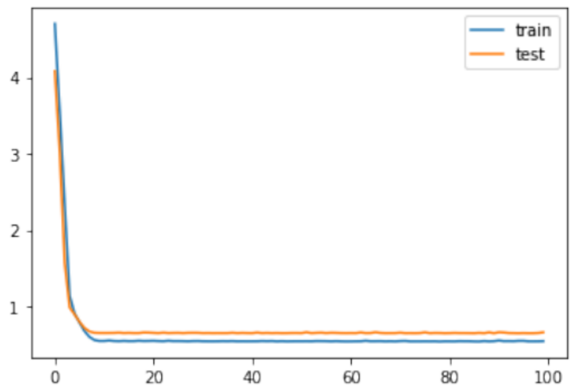

Figure 41: Results from LSTM.

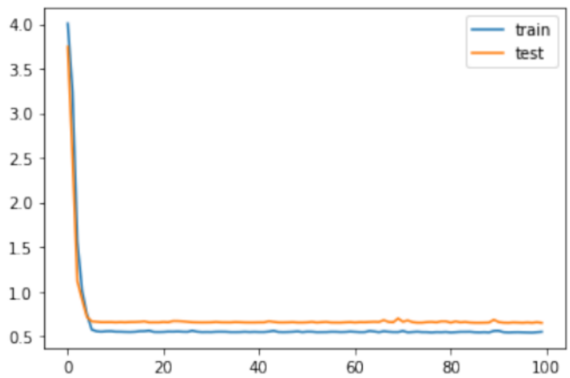

Figure 42: Results from Stacked LSTM.

Relative to the LSTM, three distinct models were implemented. The first model (Figure 41) was the classic LSTM with a single layer. The second model (Figure 42) was an extension of the classic LSTM model, which is the LSTM Bidirectional. This entails the introduction of two models instead of one. The first model learns the sequence from the input, while the second model learns the inverse of that sequence.

For this, it is necessary to have a mechanism that can combine both models, the merge step, which can be achieved through addition, multiplication, average, or concatenation functions. The third and final model (Figure 43) implemented was another extension of the classic LSTM model, and is referred to as the Stacked LSTM. While the classic LSTM consists of a single hidden LSTM layer followed by a feed forward output layer, the Stacked LSTM comprises multiple hidden LSTM layers, each of these containing multiple memory cells.

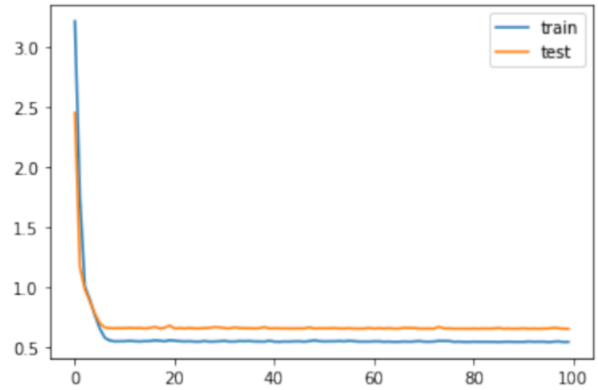

Figure 43: Results from LSTM Bidirection.

These three graphs illustrate the behaviours of the models in terms of Mean Squared Error (MSE) loss function, with optimization conducted using thee Adam optimizer. As we can see, all three models of LSTM exhibit a rapid reduction in loss initially, followed by a stabilization phase. Upon a fast observation, it becomes evident that the Stacked LSTM slightly outperforms the LSTM Bidirectional and the LSTM single layer. However, the computational cost associated with training the models is higher in comparison to the LSTM single layer.

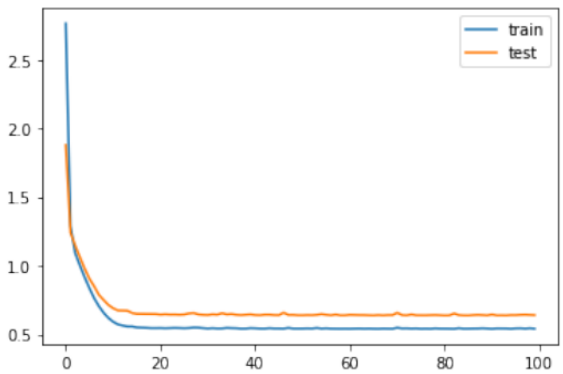

Figure 44: Results from CNN-LSTM.

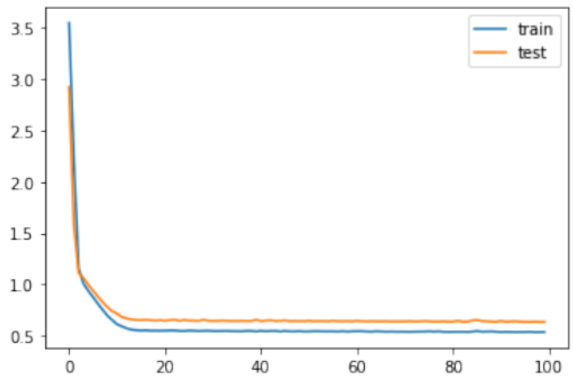

Figure 45: Results from ConvLSTM2D.

Regarding the ConvLstm (Figure 44), two models were implemented. The first model is the CNN-LSTM, which is an integration of a Convolutional Neural Network (CNN) with an LSTM. In the first phase of this model, the CNN part of the model processes the data, and in the subsequent phase, the one-dimensional result of the first phase is fed into an LSTM model. The second model, ConvLSTM2D (Figure 45), resembles the traditional LSTM, but both the input transformations and the recurring transformations are convolutional.

In summary, conclusions lean toward the Stacked LSTM and the Encoder-Decoder LSTM, as these two showed consistently better results than the other two models. Although neither really performed significantly better than the other, the Stacked LSTM was ultimately chosen for two main reasons. Firstly, it had the single best performing model, albeit with only a slight edge. Secondly, its simplicity and proximity to basic LSTM architecture were considered advantageous. However, it is important to acknowledge that, like many algorithms, it has limitations in terms of transparency and interpretation. Therefore, to bridge this scenario, the next section describes how an Explainable Artificial Intelligence (XAI) algorithm interprets and explains to the user the output of the previously employed algorithm.

6.1.2 Human Mobility from Machine Learning models Platform

This work is a comparative case study of ML models applied to a specific human mobility problem. The research uses a set of public datasets provided by NYC OpenData, more specifically the (1) LinkNYC Kiosk, (2) Yellow Taxi Trip Data, and (3) Green Taxi Trip Data [149, 150, 203]. The LinkNYC Kiosk dataset provides data about the LinkNYC kiosks, encompassing various data such as location and the operational status of the Link's WiFi, tablet, and phone of the person connected to the network. This dataset comprises 2161 rows and 29 columns. The Yellow Taxi Trip dataset includes fields capturing timestamps for pick-up and drop-off events, along with corresponding locations, trip distances, itemized fares, rate types, payment methods, and driver-reported passenger counts, all from 2018. Similarly, the Green Taxi Trip dataset records all trips conducted in green taxis in New York City (NYC) starting from 2016. Thus, in order to better understand the data, an exploratory analysis was carried out, followed by a data treatment that stemmed from the results we obtained during the analysis phase.

6.1.2.1 Predictive Process Analytics

Following the exploratory analysis, we used data preprocessing to input in ML models. The developed models are based on two architectures: LSTM and ConvLSTM. Regarding LSTM, we implemented three distinct models. The first model was the classic LSTM with a single layer. The second model was an extension of the classic LSTM model, referred to as LSTM Bidirectional. This model employed two separate models, one to learn the sequence from the input and the other to learn the reverse sequence. The third model implemented was the Stacked LSTM, which delivered marginally better results compared to LSTM Bidirectional and LSTM single layer. However, the computational cost associated with training the models is higher in comparison to the LSTM single layer.

Regarding the ConvLSTM, two algorithms were implemented: CNN-LSTM and ConvLSTM2D. In the CNN-LSTM model, the initial phase involved the CNN processing of the data, and in the second phase, the one-dimensional result of the first phase is fed into an LSTM model. In its turn, ConvLSTM2D ConvLSTM2D involved convolutional transformations for both input and recurring processes. Then, we compare the behaviours of the CNN-LSTM model and ConvLSTM2D. The training result of these models is better than the training result of the three models of LSTM, although the computational cost for training was notably higher. Figure 46 presents the outcome of these algorithms.

Models	Val_Loss
LSTM	0.650
CONVLSTM2D	0.632
STACKED LSTM	0.648
LSTM BIDIRECTIONAL	0.649
CNN-LSTM	0.639

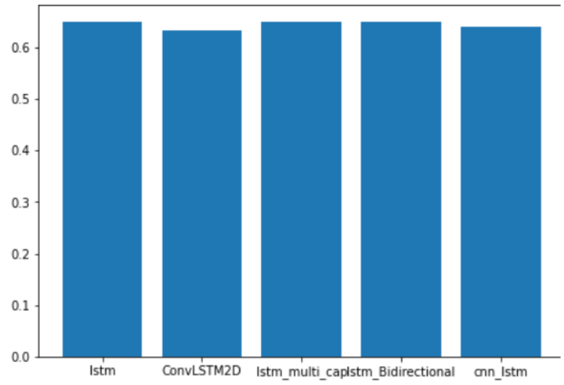

Figure 46: Comparison of the ML models.

The analysis that while there are no big differences between the models, there is a subtle distinction observed in the ConvLSTM models, showcasing a better performance. In scenarios where an expansive and flexible feature set is available and the computational training cost is not a limiting factor, the ConvLSTM2D model could be the preferred choice due to its lower loss value of 0.632. On the other hand, when confronted with a limited feature set and the necessity for the best model with the lower computational cost, opting for the LSTM single layer model with a loss value of 0.650 might be advantageous. This choice is bolstered by its relatively lower computational cost and only a slightly higher loss value. However, if we want a balance between the computational cost and the loss value, the CNN-LSTM model might be the middle ground. With a loss value of 0.639, it offers a compromise between lower computational costs compared to ConvLSTM2D and still has one of the better loss values within the model set.

6.1.2.2 Explainable Decision Support Prototype

Based on the results discussed in the previous section, we are going to implement a platform that facilitates the interpretation of the presented values. This endeavour is motivated by an intention to delve deeper into the black- box nature of Deep Neural Network (DNN) models and to provide an alternative way of explaining DNN behaviour. In other words, this work presents a new component in the dissemination of results extracted from forecasting algorithms, as compared to the prototype presented in Section 6.1.1. Therefore, with this work, we intend to demonstrate that a more objective interpretation of forecasts through the application of explainability models is possible and, ultimately benefiting stakeholders.

While our work deals with fully connected and CNN, it's worth noting that the Multivariate LSTM model, in particular, presents the most favourable value taking into account the overall computational effort required for the calculation (less storage memory). This model offers a representation directly equivalent to that of a Decision Tree (DT). In essence, it examines the transformation of the Multivariate LSTM model to a decision tree model by taking a sequence of nonlinear weights between them and transforming it into a new weight structure. The following image explains and combines the advantages of this DNN model and a DT.

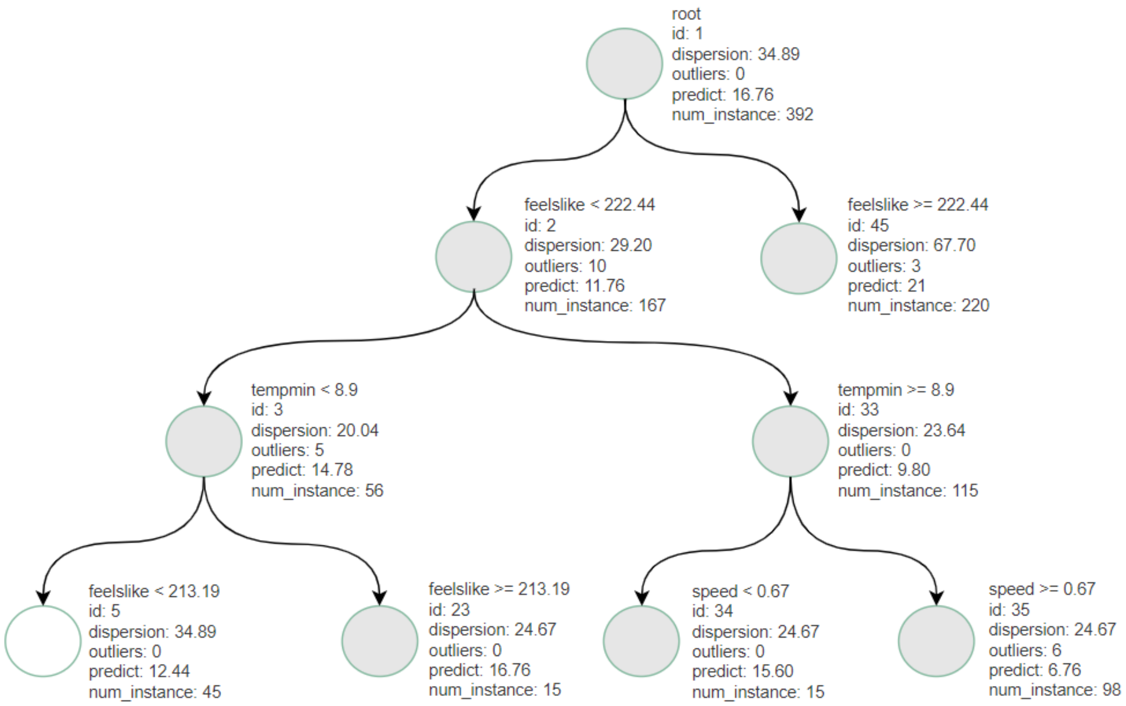

Figure 47: Distilling Multivariate LSTM model into Decision Tree model.

Figure 47 shows a sample forecast path from the root node to the leaf node generated via DT from the outcomes of the multivariate model. Beginning at the root, the DT predicts a value—in this case, 16—for the number of individuals. In the next forecast, with the $feelslike$ attribute equal or greater than 269.442, the DT predicts a count of 21. However, it substantiates this prediction by considering values of $feelslike$ below 269, resulting in a slight error. To model this specific issue, the DT employs 3 attributes—namely, $feelslike$, $tempmin$ and $speed$. Moreover, the DT performs automatic counter-factual analysis. For instance, when the value of the $tempmin$ attribute is below $7.4°C$, the tree gives a tree level and predicts 14 individuals at the center of NYC. In case of three level, the $feelslike$ attribute predicts 13 individuals; otherwise, it foresees 18 individuals. Additionally, when $tempmin$ is equal to or greater than $7.4°C$, it also impacts the census prediction, estimating 10 individuals. On child nodes, it should be noted that when the $speed$ value is below 0.54, the prediction for the number of individuals stands at 19. Conversely, when the $speed$ value is equal to or above 0.54, the forecast drops to 10 individuals. In summary, it should be noted that each decision rule within our trees is associated with a node that contains more than a dozen useful attributes, contributing to a comprehensive understanding of the data distribution.

6.1.3 Third-party API Integration Platform

This application was implemented to analyze human mobility during the Covid-19 pandemic. To facilitate and help the user, a recommendation system was developed to suggest safe places to move. Given the context of the Covid-19 situation, our focus was on recommending locations with specific criteria: less than 100 cases and at least 50% of the population vaccinated. With that in mind, we use information from disease data (such as cases and deaths), vaccination data (number of people fully vaccinated, number of people vaccinated with just one dose...) to mobility data. Therefore, to acquire this data we accessed various API sources: (1) Cases and Deaths: European Center was Deceased and Carings, (2) Vaccination: World in Data and (3) LinkNYC Kiosk [45, 129, 150].

6.1.3.1 Predictive Process Analytics

From integration with the third-party API services, the application's architecture is built on the Django server, organized following the Model-View-Template (MVT) style. While similar to the more common Model-View-Controller (MVC) model, in the MVT model, a user requests a resource from Django, which acts as a controller, verifying if the requested resource is available at the Uniform Resource Locator (URL). Then, the database is the source of two modules: one for processing of Deep Learning (DL) models and the other for the application's own API. Within this framework, the processing of DL models involves two key tasks. Firstly, a LSTM model is used to predict whether the number of Covid-19 cases would go up or down in a given week (binary categorical dependent variable). Secondly, another LSTM model is employed to predict the number of individual movements within each of five different boroughs for the upcoming five weeks.

To reach the expected results, the LSTM models were trained with cross-validation and the lowest MSE metric. The choice of hyper-parameters in training was also done experimentally. The training data is stored in the PostgreSQL database at the end of each training session. Additionally, the model forecasts are also saved in the database. This stored data is accessed by an API to generate graphs that display the future forecasts for the five boroughs. Thus, the application's API receives a diverse amount of data already processed and ready to be presented to the user.

6.1.3.2 Explainable Decision Support Prototype

In Figure 48, a DT (again with $max\_depth = 3$) is presented, serving to elucidate the used ML model's behaviour. In this technique, we see a sample forecast path, starting from the root node and proceeding to the leaf node, all generated via DT using the multivariate model's outcome. In the root node, the DT predicts a count of 16 individuals in NYC's center. In the next forecast, when the $feelslike$ attribute is equal to or higher than 269.442, XAI applied to Smart Human Mobility DT predicts a count of 21 individuals. However, it provides reasoning for this forecast by considering the $feelslike$ values below 269, introducing a slight error. In this prediction step, three attributes, namely $feelslike$, $tempmin$, and $speed$, collectively contribute to modeling this specific problem's outcome.

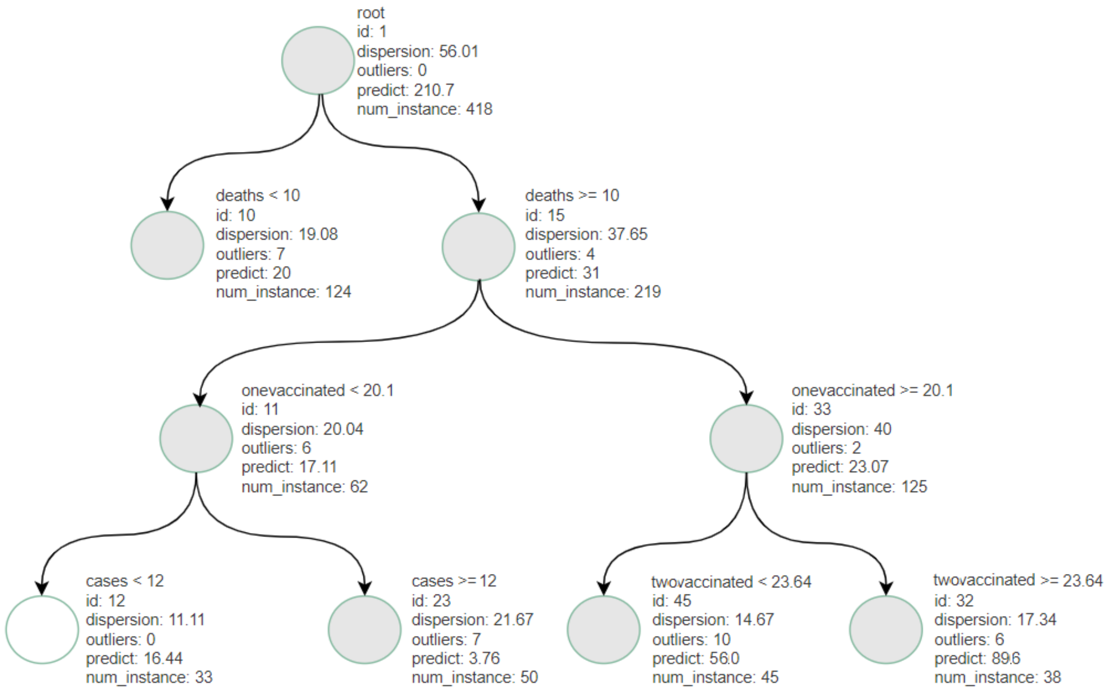

Figure 48: LSTM model into Decision Tree model.

Furthermore, this method generates automatic counterfactual analyses to provide additional insights. When the value of the $tempmin$ attribute is below $7.4°C$, the tree gives a tree level and predicts a count of 14 people in New York City's center. In the other hand, if we consider a tree level of three, the $feelslike$ attribute predicts 13 people; otherwise, it predicts 18. When the $tempmin$ value is equal to or above $7.4°C$, it also has an impact on census prediction, resulting in an estimation of 10 people. Moving to child nodes, it should be noted that when the $speed$ value is below 0.54, the prediction for the number of people is 19. Conversely, when the $speed$ value is above or equal to 0.54, the prediction drops to 10.

6.1.3.3 Extract Rules from Decision Tree

The rules extraction from the DT can help to better understand how samples propagate through the tree, of demonstrated in Section 6.1.3.2, during the prediction process. Moreover, DT is easy to move to any programming language due to its if-else statements. In this case, we move the Scikit-Learn Decision Trees to Python. Through this prototype, we illustrate the process of extracting decision rules from the DT (for a regression task). This transformation involves converting the complex Decision Tree structure into set of rules which are human-readable.

Currently, there isn't any built-in method for extracting if-else code rules from the Scikit-Learn tree [185]. Therefore, we developed a method to achieve this, but not without first extracting Human-readable rules. Initially, we eliminate unnecessary rule antecedents to simplify the rules, constructing contingency tables for each rule consisting of more than one antecedent. This step involves simplifying individual rules by eliminating antecedents that have no effect on the rule's outcome. Then, we eliminate unnecessary rules to simplify the rule set. Additionally, we replace those rules that share the most common consequence with a default rule. This default rule activates when no other rule is triggered. This process enhances the human-friendliness of the ruleset. For the sake of clarity, Listing 16 presents a Python code that exemplifies this process.

Listing 16: Extract human-readable rules.

```
def get_rules(tree, feature_names, class_names):
   tree_ = tree.tree_
   feature_name = [
      feature_names[i] if i != _tree.TREE_UNDEFINED else "undefined!"
      for i in tree_.feature
   ]

   paths = []
   path = []

   def recurse(node, path, paths):

      if tree_.feature[node] != _tree.TREE_UNDEFINED:
         name = feature_name[node]
         threshold = tree_.threshold[node]
         p1, p2 = list(path), list(path)
         p1 += [f"({name} <= {np.round(threshold, 3)})"]
         recurse(tree_.children_left[node], p1, paths)
         p2 += [f"({name} > {np.round(threshold, 3)})"]
         recurse(tree_.children_right[node], p2, paths)
      else:
         path += [(tree_.value[node], tree_.n_node_samples[node])]
         paths += [path]

   recurse(0, path, paths)

   # sort by samples count
   samples_count = [p[-1][1] for p in paths]
   ii = list(np.argsort(samples_count))
   paths = [paths[i] for i in reversed(ii)]

```

```
    rules = []
    for path in paths:
        rule = "if "

        for p in path[:-1]:
            if rule != "if ":
                rule += " and "
            rule += str(p)
        rule += " then "
        if class_names is None:
            rule += "response: "+str(np.round(path[-1][0][0][0],3))
        else:
            classes = path[-1][0][0]
            l = np.argmax(classes)
            rule += f"class: {class_names[l]} (proba: {np.round(100.0*classes[l]/np.sum(classes),2)}%)"
        rule += f" | based on {path[-1][1]:,} samples"
        rules += [rule]

    return rules
```

The DT is typically visualized as a graph, but it can also be converted into a text representation with a human-friendly format. Basically, we generate a rule set model nugget that represents the DT's structure as a set of rules that define the terminal branches of the tree. While this rule set retains most of the important information of a DT, it offers a simpler model. The most important difference is that in a rule set, more than one rule might apply for any particular record, or no rules could be applicable at all. Moreover, there are differences in the scoring of each rule when compared to scoring against the tree as each terminal branch in a tree is scored independently. Listing 17 presents the generated ruleset from the DT model.

Listing 17: Rules of Decision Tree.

```
if (deaths < 10) then response: 20 | based on 124 samples
if (deaths >= 10) and (onevaccinated < 20.1) and (cases < 12) then response: 16.44 | based on 33 samples
if (deaths >= 10) and (onevaccinated < 20.1) and (cases >= 12) then response: 3.76 | based on 50 samples
if (deaths >= 10) and (onevaccinated >= 20.1) and (twovaaccinated < 23.64) then response: 56.0 | based on 45 samples
if (deaths >= 10) and (onevaccinated >= 20.1) and (twovaaccinated >= 23.64) then response: 86.0 | based on 38 samples
```

At the conceptual level, to extract rules from a decision tree, one rule is created for each path from the root to a leaf node. Each splitting criterion along a given path is logically combined using "AND"operations to form the rule's antecedent ("IF"part). The leaf node holds the class prediction, forming the rule's consequent ("THEN"part). A disjunction (logical "OR") can be also implied between each of the extracted rules. Because the rules are extracted directly from the tree, they are mutually exclusive and exhaustive. Mutually exclusive means that rule conflicts are avoided here since no two rules will be triggered for the same tuple. Exhaustive signifies that, there is one rule for each possible attribute-value combination, eliminating the need for a default rule. Therefore, the order of the rules does not matter—they are inherently unordered.

As mentioned, we have explored the fundamental concepts of DT and approached not just building models but also interpreting them. DT is a powerful and interpretable tool for regression tasks, making them essential in this prototype. Finally, we also understand how the external Representational State Transfer (REST) servers are important elements in ML projects. The capabilities of Django programming language have enabled us to achieve an ecosystem completely assembled in Python, encompassing both the DL models and the application itself. As a result, we have developed a recommendation system that is able to process the most recent data and provide, as an output, valid and useful knowledge to the user.

6.1.4 SiteWhere platform supporting Crowdsensing

SiteWhere stands as an open-source Internet of Things (IoT) application with support for both on-premise and cloud deployment models. For the purpose of this prototype, it is set up and runs on a local server. Moreover, *SiteWhere* works with device integration to create point solutions for smart city problems.

The *SiteWhere* IoT platform was used to create the first *WalkingStreet* prototype. This preliminary phase of the *WalkingStreet* platform is an IoT solution that collects and integrates information sourced from a smartphone. This IoT solution allows data from a device to be collected, correlated and analyzed. The solution allows the user to better understand his behaviour activities (i.e., location and timestamp) and deliver information to end-users through SiteWhere's back-office.

According to the results, this prototype excels in three primary areas of the *SiteWhere* platform. Firstly, it brings security and extensibility to other types of devices, enabling a more comprehensive capture of human mobility phenomena. Secondly, it also supports multitenancy (i.e., deploying multiple apps across several tenants), and on-premises deployment, catering to intricate infrastructure requirements, such as those found in cloud environments. Lastly, this prototype allows real-time monitoring and high availability, empowering administrators to visually comprehend the collected data. Given these features, this testing phase can be an important element of the decentralized module of the *WalkingStreet* architecture.

6.1.4.1 Process Analytics

In this prototype, data acquisition was facilitated through the *SiteWhere* plaftorm. The structure data is similar to the model presented in Section 5.1.1. Our dataset is a product of the integration between smartphones and *SiteWhere*, showcasing attributes such as anonymized, precision-timestamped, and object data in high resolution. This dataset provides coverage of the population count within a specific geographical area. The primary dataset for this experiment is derived from the integration of smartphone and smartwatch data into the *SiteWhere* framework. This initial dataset comprises over 20,000 rows. Furthermore, the cumulative dataset encompasses around 50,000 rows.

To conduct the prediction process, we employed the LSTM model. As previously mentioned, this model plays a prominent role in the development of human mobility solutions. However, it's important to acknowledge that in this instance, the availability of high-quality real-world data is a significant prerequisite for the reliable development and deployment of such large-scale systems. In this particular context, the limitations in available human resources constrained the dataset, resulting in a relatively limited volume of data entries.

6.1.4.2 Decrypting Prediction from LIME

To interpret the LSTM, we used the Local Interpretable Model-Agnostic Explanation (LIME) model. In other words, we interpret the LSTM through LIME, focusing on a specific instance (in this example, the 20^{th} instance). With this in mind, from the $LimeTabularExplainer$ method, the chosen XAI model creates an explanation, and the $explain\_instance$ method evaluates these explanations related to the scenario described in Section 6.1.4. The $explain\_instance$ function requires three arguments: X_test, which specifies the testing segment of the first sequence *(X)*, predict function, which specifies the predictive behaviour using the linear model; and number features, indicating the maximum number of features present in explanation *(K)*. Additionally, we specify the number of features of the explanatory $num\_features$ variables. By applying the as_pyplot_figure to the object containing the explanation, we obtain a graphical presentation of the results. Figure 49 presents an output that utilizes the colours blue and orange depicting negative and positive associations, respectively.

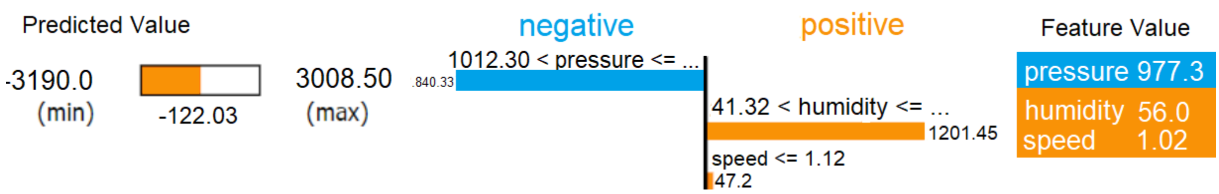

Figure 49: The LIME explanation.

Interpreting the aforementoned results, we can conclude that the relative census value (depicted by a bar on the left) attributed to the given test vector (X) can be linked to several factors: (1) the high value of the pressure feature, indicating a lower number of people in the NYC center, (2) the high value of the humidity feature, indicating the a higher number of people, and (3) the low value of the wind speed, suggesting an increase in the number of people in the center. Additionally, the LIME provides a localized interpretation for specific predictions, such as the Predicted Value Section. This interpretation includes both the direction (positive or negative) and the magnitude (weight) of the impact exerted by the most important features.

In the printed output, the *Predicted Value* computed value corresponds to the number given in the column labeled $model\_prediction$ in the printed output. In this case, the computed value is approximately 122 individuals in the NYC center. The prediction interval spans from a minimum value of -3190.0 to a maximum value of 3008.50. *Negative and Positive* sections show the features that contribute to decreasing and increasing the predicted value, respectively, along with their respective weights. Features with positive weights contribute to an increase in the prediction value, while features with negative weights contribute to a decrease. Considering the 20 instances, these features play a significant role in the prediction:

- *pressure > 1012.30* negatively affects the predicted value. This suggests that higher of pressure is associated with a lower median value of pedestrians in a specific area;
- *humidity > 41.2* and *speed >= 1.12* positively affect the predicted value. This suggests that a lower humidity percentage and a wind speed lower or equal to 1.12 are associated with a higher median value of pedestrians in a particular area.

The Feature Value or $model\_intercept$ column provides the value of the intercept. This value indicates which explanatory variables were given non-zero coefficients in the Linear Regression method. Additionally, it provides information about the values of the original explanatory variables for the observations in the calculation of the explanations. In the "Feature Value"section of the output, you can see the actual values of the top features for this instance (20^{th} instance). For example, the pressure value is 977.3, which is indeed greater than -3190.0, thereby negatively influencing the predicted value according to the model. Similarly, the humidity and speed values are variables that positively influence the prediction outcome.

6.1.4.3 Explain model prediction with interpretation of LIME

Creating a listing set using LIME provides users with a global view of a model's decision boundary, enabling effective human interaction with ML algorithms. Explaining individual predictions is fundamental for assessing trust and complimenting hold-out set validations in the selected model.

Listing 18: Select 20^{th} instance.

```
# Choose the 20th instance to predict the results
j = 20
exp = explainer.explain_instance(X_test.values[j], model.predict, num_features=3)

# Interpretation of results
Intercept 121.412345
Prediction_local [119.213230]
```

```
Right: 122.033745
```

While LIME doesn't offer direct export-to-dataframe capabilities, you can append predictions to a list and then transform it into a Dataframe. In Listing 18, the $explain\_instance$ needed to be adjusted to the arguments of the model, and the predicted value for the number of pedestrians is 122.03. However, more information is required than what the as_list provides. The explainer contains additional data. With this in mind, we ran an example (Listing 19) to see what else the explain instance function would retrieve. The feature values section contains the actual values of the top three variables. From this, it is evident that the pressure variable has a negative influence, while humidity and speed have a positive influence on the predicted number of pedestrians.

Listing 19: List the feature values.

```
# List the feature values
pd.DataFrame(exp.as_list())

#Influence level of variables on prediction
[(1012.30 <= "pressure" <= "...", -0.840330),
 (41.32 < "humidity" <= "...", 1201.346989),
 ("speed" <= 1.12, 47.152828)]
```

To conclude, with the help of LIME, we demonstrate the importance of explainable AI in a data-driven landscape. Nowadays, it is of the utmost importance to not only create highly accurate machine learning models but also to ensure their predictions are transparent, interpretable, and understandable. In this prototype, the inner workings of the XAI model can be understood, and meaningful explanations can be provided for their predictions. LIME allows us to comprehend the influence of individual features on specific predictions, offering a local, interpretable model of the original predictor. This local interpretation, although an approximation, provides valuable insights into the model's decision-making process, enhancing our trust and confidence in its predictions.

6.2 Scientific Dissemination

Regarding the dissemination of the results to the scientific community, we actively engaged in a variety of avenues such as documents and conferences. During his PhD, the author of this work faced the task of demonstrating the evolution and the respective preliminary results of the objectives outlined in the Thesis Planning phase. Notably, the Thesis Planning, conceived at the outset of the doctoral program for internal reference, served a pivotal role in elucidating the cutting-edge landscape related to the proposed endeavor. This process further helped in identifying pertinent requirements and challenges inherent in its execution. This document played a crucial role in garnering approval from a panel of academic peers for the thesis theme itself, thus providing the author with fresh perspectives and insights into the directions to be taken within the topics of this project.

Following the initial findings, we embarked on the dissemination of our results through participation in several national and international conferences. Following the chronology of events, we published the first developments at the symposium of Portuguese Conference on Artificial Intelligence (EPIA), hosted in Vila Real, Portugal. This symposium provided an informal environment to the author of this thesis during the early stages of their doctoral research, enabling the presentation of their main ideas and fostering discussions with fellow students and the senior scientific community. Notably, these exchanges delved into the exploration of permanent relationships between topics such as SC, IoT, ML, and Mobility. The submitted article explains the wide applicability of ML in projects involving mobility datasets. Additionally, it was developed using data analysis tools such as *KNIME Platform*.

Subsequently, we explored the current state of Mobile Networks and IoT infrastructures in the context of mobility; then, we analysed the workshops that fit the themes discussed. The article produced was submitted to the workshop "Social Media Analysis for Intelligent Environment"(SMAIE), in Spain. This relevant paper highlights the strong relationship between humans and devices associated with the IoT at the hardware and connectivity level (e.g., smartphones or smartwatches using Global Positioning System systems) that impact asset tracking applications. Further extending the impact research, our findings were distilled into another article submitted to the "Distributed Computing and Artificial Intelligence"(DCAI) conference, held in L'Aquila, Italy. In this paper, we address every mobile generation technology (ranging from 1G to 5G) and their diverse applications within SC projects. In summary, in both works, we explored the potential and repercussions inherent in the Human-Device relationship.

In 2020, we participated in "Citizen-Centric Smart Cities Services"(CCSCS). This conference considered SC as a new paradigm or a concept that is emerging as a necessary and unavoidable response to the constant growth in urban population across the globe. In response to the multifaceted technical, material, social, and organizational problems stemming from this urban expansion, SC has emerged as an imperative proposition. The aim of these cities is to improve the quality of life of their citizens, and to provide a more economic competitive, sustainable, and liveable city. We developed an important paper that promoted an IoT architecture tailored for pedestrians-oriented applications. This project approached several technologies that enables the creation of new possibilities and capabilities, fostering enhanced prospects for human mobility applications.

We elevated our research to another level by delving into the exploration of the role of open-source tools in capturing human mobility data. Presented at the "International Conference on Intelligent Data Engineering and Automated Learning"(IDEAL), this work combines microservices approach, real-time data, and an open-source IoT application, providing information from a heterogeneous mobile sensing platform. The architecture proposed in this paper, called Temperature Measurement System Architecture (TMSA), also demonstrated the potential of sensors in urban sensing context.

In this phase of our Ph.D. program, we revisited some topics already mentioned, such as SC, IoT, Big Data, and inclusive architectures. What emerged was a recognition of the interconnectedness among these domains, as the initiative of each one influences the functioning of the system they are a part of. Amidst these considerations, we recognized the imperative of delving into more detailed investigations involving citizens, all the while safeguarding their privacy. To this end, we directed our research to study the impact of comfort and well-being indicators on pedestrians within urban areas, using modelling and predicting methods. These indicators were subjected to a comparative forecasting performance between statistical and DL models. This article was an important step to confirm that ML models are the most appropriate models for analysing the human mobility phenomena.

Building upon the insights gained from our previously submitted paper, we published an article at the "4^{th} Workshop on Affective Computing and Context Awareness in Ambient Intelligence" (AfCAI). This article employs ML models to delve into sentiment analysis based on human mobility data, collected through LinkNYC devices. This work also provides extensive results in univariate and multivariate scenarios. Therefore, the collected data allows the modelling and prediction of sentiment of a citizen in relation to time and space properties.

Incorporating the time and space properties from the earlier paper, which are fundamental for representing human mobility patterns in SC, motivated us to submit another paper at the "1^{st} International Workshop on Social Media Analysis for Intelligent Environment" (SMAIE). Here, besides the LinkNYC Kiosks dataset, we used another dataset—311 Service Requests in New York City—to explore the potential of these two properties. Our analysis encompasses a range of dimensions in human mobility patterns tied to these attributes. We harnessed mathematical formulas individual and community level mobility analysis, while leveraging Gaussian distribution to explain population density within the city. This work culminates in the use of prediction algorithms, specifically ML, to further explore the provided data.

To consolidate the body of work developed thus far in this phase of the Doctoral Program and to be reviewed by appropriate peers, an article was created for the International Journal of Smart Cities (within the Urban Studies category). The choice of this journal was intentional, aiming to engage a diverse audience of researchers and policymakers interested in the field. This article provides essential guidelines for implementing a technological solution for human mobility. It emphasizes the necessity for authorities to invest in emerging, non-invasive devices that can swiftly interconnect with prediction algorithms. Throughout the document, many topics are addressed, culminating in insights for the next steps in building an architecture for a Smart human Mobility project.

Given our focus on the work submitted to the Journal of Smart Cities, in which we generally address the metrics that support researchers of the phenomena of human mobility, we decided to delve deeper into these scientific formulas. In this work, we explore the equations that make up the most common metrics used in the analysis of human mobility. Then, the experience went through the use of capturing data from two citizens from different sources in the city of New York. This heterogeneity in data sources allowed for a comparative analysis of metrics such as resilience, displacement, interval, or duration across different datasets. The discussion of the results focused on the comparison of mathematical spectrums associated with each metric.

At this stage of the thesis, we can finally present a prototype that combines all the elements that make up the works developed so far. In our submission to the "Conference on Data Engineering and Automated Learning"(IDEAL), we captured human mobility data through a mobile application called *WalkingStreet*. This application is supported by cloud tools that allow for large-scale storage. Subsequently, the stored data will be useful for metrics that investigate human mobility phenomena. The reader is also offered a nuanced comprehension of WalkingStreet's layers, its architectural framework, and the considerations shaping its design. The implementation and navigability of this application are two other topics that are discussed in this prototype.

Our recent engagements in scientific conferences have notably focused on the potential of XAI models. In an article submitted to the international workshops of Citizen- Centric Smart Cities Services (CCSCS), held in France in 2022, we elaborated on how a DT model provides a forecast accompanied by an explanation. Meanwhile, the paper from the "Distributed Computing and Artificial Intelligence"(DCAI) conference, in L'Aquila, Italy, launches a broad discussion on various XAI models. In a particular way, this work seeks to conduct a comparative study on the explainability performance of this type of models. However, transversally, these documents have demonstrated their significance as valuable scientific contributions to help solving the explainability challenges of human mobility phenomena, or at the very least, serve as a starting point for addressing them.

6.3 Summary

During the development of this work, we have gained an appreciation for the importance of the IoT and AI in supporting human mobility solutions within SC. We have also come to realize their extensive presence in the scientific community and the industry. These topics are getting communities and universities involved, alongside big companies and city authorities. We found that researchers have played an active role in this, consulting with citizens and working with city halls to promote cooperation between citizens and local institutions. This collaborative effort has helped shift the focus of smart city projects towards meeting the needs of people. Our literature review has shown relevant use cases that address urban challenges. In fact, while several approaches exist to tackle diverse city issues, only a few are focused on the main objective: enhancing smart human mobility. Researchers can still do more than what has already been done in this field.

The experiments conducted on human mobility have also detailed some problems, namely, capturing data and extracting knowledge. The collection of such geolocation data which, acquired from both individual and public devices as part of the identification of geographic scenes of telecommunications companies, presents legal concerns regarding the potential invasion of people's right to privacy. Even when focusing only on building sensorization infrastructures, some have difficulty interpreting the data generated by ML to exploit the patterns in historical data and predict mobility performance. The preliminary results showed that in AI there are many ML algorithms that can be used to learn from the extensive data collected through a smart city's infrastructure. However, the current state of the art offers an integrated view for addressing these several challenges in methodology research. This extends beyond merely predicting human mobility to the establishment of viable metrics capable of creating new services aimed at solving human mobility problems within SC.

Important works were developed, highlighting the built archetype to present the scope for smart human mobility. These efforts have shed a light on the limitations and capabilities of smart city projects. By employing techniques and paradigms to gather data and produce knowledge, we have laid the groundwork for improving the societal impact of human mobility solutions. Moreover, our findings underscore the imperative inclusion of both local and government authorities in such projects. The data generated by sensors of public-private entities amplifies the capability to fulfill the objectives of this work. In other words, the smart city infrastructure, with several data from public-private sources, ensured that it could boost efficacy in data collection contexts. Importantly, we have also succeeded in developing a service that provides a set of solutions within a highly secure and protected environment. This achievement serves to demonstrate that it is possible to analyze human mobility patterns while safeguarding individual privacy.

Once an archetype was developed in this Doctoral Thesis, the most notable triumph of this work is the ability to answer the research hypothesis defined in Section 1.3. The knowledge gleaned from the state-of-art review, previous laboratory experiments, and the creation of the *WalkingStreet* application provide comprehensive answers to these research questions. Additionally, the planning and execution of this application were meticulously guided by the definition of its scope and goals.

The first and second research questions find their resolution within the decentralized module of the *WalkingStreet* architecture. Positioned as the first phase of the data flow in the proposed architecture, this module effectively serves as the gateway for the influx of data that will permeate the rest of the architecture. Here, this module gathers a set of sensorization techniques and other open-source innovations to monitor and address human mobility data within Smart City environments. The use of an IoT platform enables the data management of data sources. For example, this platform acts as a bridge between device sensors and data networks, allowing the remote collection of human mobility data while maintaining secure connectivity. Moreover, the platform's software interface enables the search for information and provides visibility into operations, which enables better decision-making. Additionally, the IoT platform is the node of *WalkingStreet* architecture responsible for the first step of the data fusion process. It encompasses dedicated containers that manage the fluidity of data flow. Experiments in this field demonstrated that this platform is an innovative approach for data acquisition in mobile crowd-sensing systems.

The third research question is regarding a centralized system. This system connects directly one or more decentralized/server nodes, computing the ML and XAI algorithms as well as metrics to discover and describe human mobility patterns of urban space where these nodes are implemented. This centralized system emerges as a crucial component of the *WalkingStreet* architecture, wielding substantial influence over the decision-making process of its users. In essence, this component stands as the nexus where the outcomes of proposed AI, XAI algorithms, and metrics converge. In fact, this component enriches the proposed solution with premises such as DL models, interpretable AI and human mobility metrics taken from the literature review. The DL techniques were chosen for the predictive modelling of human mobility phenomena due to their adeptness in handling complex and demanding datasets with a large number of attributes and in managing available resources in the system. Therefore, a representation of human mobility is composed of an effective engine that estimates the next population movement within the urban space. This work also seeks to incorporate external factors for a better contextualization of where the user is embedded. Complementing this, human mobility metrics such as displacement assume a role in fortifying risk-mitigation strategies from a human mobility perspective. They provide critical information to decision-makers and responders during the course of human mobility phenomena. However, for the AI algorithms to be understandable, this platform uses the XAI models, bolstering transparency from a human-centric viewpoint. Experiments in this field demonstrated the platform is able to present the relative contributions of each feature in several human mobility dynamics.

The fourth research question refers to user behavioural modification through suggestive modules (i.e, predictive models and metrics) within the *WalkingStreet* platform. These modules, encompassing predictive models and metrics, find their fruition through a collaborative effort within the community and the support of a Decision-Support service, such as Mobility as a Service (MaaS). They show the different phenomena of human mobility behaviour in each prediction model or metric assessment. They also underscore the advantages of individual and community participation. In other words, the results produced by *WalkingStreet* consider not only the experiences of individual users but also collective patterns within a user cluster. Although not considered a gamification technique, the awareness that the participation of each user is essential for the common good demonstrates that there is possible to add other types of dynamics that encourage involvement with the proposed service and, consequently, contribute to the quality of collective mobility. Section 5.4 of this thesis shows the implementation of these scenarios in the *WalkingStreet* platform.

Chapter 7

Conclusions and Future Work

In this chapter concluding chapter, we intend to bring this document to a close by reflecting on the entirety of the work carried out and by outlining potential tasks that can be undertaken in the future. We believe that the proposed objectives have been successfully achieved over the course of this Doctoral Program. The culmination of our efforts has resulted in the proposal of an innovative architecture for human mobility. However, it is essential to acknowledge that Smart City (SC) projects are typically changeable, undergoing continuous improvements, driven not only by progressive technological sophistications but also by the increasing attention they attract from governmental and non-governmental authorities and citizens in general. Therefore, we also provide insights into what can still be improved in the proposed service. Finally, we highlight a series of initiatives that have been undertaken to disseminate the results of this work. Concurrently, these initiatives have served as platforms for supporting and/or collecting scientific contributions from students at the University of Minho. Therefore, we consider this Doctoral Program not only as an opportunity to create something new but also as a means to contribute to the ongoing evolution of the academic community.

7.1 Conclusions

Our core research focus has revolved around the exploration of human mobility within the context of smart cities. Multiple fields were considered to develop the proposed human mobility solution. In addition, the documented experiments presented at scientific conferences provide concrete evidence of the validity of our results. In our pursuit, we have employed strategies to actively engage citizens in the definition of the *WalkingStreet* archetype. Additionally, we've looked into the physical structure of cities as a valuable area for experimentation. The city's structure beneath the complex human mobility shows the inherent connection patterns within the city. Consequently, the data generated from these experiments provides valuable insights into using emerging data sources to reveal human mobility patterns, which will potentially aid in applying urban policies.

Throughout this research project, we have forged new insights that span the fields of science, technology, mobility, and governance. More specifically, our findings underscore the vital role of Artificial Intelligence (AI) in shaping the Mobility as a Service (MaaS) architecture, alongside the critical prerequisites for its construction. However, it's imperative to recognize that technology alone cannot bring about the radical changes that are necessary. Hence, innovative collaboration methods become essential. In Section 1.3, we framed our research with a set of Research Questions, which served as guides for the success of the *WalkingStreet* application. Alongside this, in Section 6.3, we summarized the answers to these questions.

The findings of our research demonstrate that the *WalkingStreet* application has the potential to serve as a MaaS solution, enhancing human mobility and providing tangible models for more effective urban planning. In essence, this service stands as a pivotal component in the ongoing evolution of SC, particularly in addressing issues linked to human mobility. Furthermore, this thesis is expected to contribute to the expanding body of research on mobility-based policies and promote their adoption for smart governance. In Section 7.2, we indicate the next steps as well as ethical principles to be implemented in the project.

7.2 Future Work

The development of our proposed solution is determined by scientific research. This research draws from the latest advancements in the field of human mobility, aiming to help in understanding the dominant topics and shed light on potential research areas of significance. Throughout this scientific exploration, a notable gap in the knowledge was unveiled: the absence of an architectural framework integrates several components, such as an Open Source Internet of Things (IoT) Platform running on-premises, to feed a cloud-based system or infrastructure to calculate the predictions and metrics. Recognizing this gap, there is a significant opportunity to further this research and develop a new concept: Mobility as a Service.

Although this service introduces a *"new concept"* for cities, unifying different fields and transcending traditional discipline boundaries to contribute to a fundamental prototype for human mobility, it also shows other distinct challenges and capabilities. First and foremost, the proposed smart city project extends its infrastructure from a local to a cloud-based element. This means that, for this proposed architecture to inspire and engage stakeholders (i.e., local authorities or governments) in addressing human mobility challenges within their cities, it requires continual investments in efficient data collection and the establishment of a highly secure and protected environment, with the integration of Machine Learning (ML) recommendation engine and a cloud element. In other words, it must demonstrate to ends-users that it is possible to participate in a sophisticated human mobility solution without compromising their privacy.

In essence, it is about the level of trust and reliability that the *WalkingStreet* platform instills to whoever is using it.

Aside from reliability, an extension of the proposed solution to the cloud allows more comprehensive, accessible and decentralized participation by citizens. Consequently, the stakeholders need to adopt the best procedures to manage system performance. The gamble on a fog computing type design is ideal in the implementation of this balancing of the performance. Additionally, although it is presented in a test version, *WalkingStreet* is prepared to adapt new devices or change the intermediate structure of the architecture, specifically the addition or removal of decentralized or centralized modules. In other words, a system that incorporates an agile architecture is essential to face extreme scenarios in data computation. Therefore, the *WalkingStreet* fulfils this requirement.

Another significant perspective to consider is that the effectiveness of this architecture is inherently tied to the context in which it is implemented. Considering this, we highlight the most apparent limitations of this architecture. At the data source level, this technology involves adding and modifying hardware personnel and some of these individual devices are too expensive for economically disadvantaged citizens, limiting the massive participation of citizens in a local, city, regional or national context. In terms of infrastructure costs, local authorities should hire specialized personnel and make substantial investments to build and maintain the proposed IoT infrastructure. This necessary investment can be an obstacle for less favoured cities that seek to allocate their scarce financial resources to more critical areas. However, it is worth noting that these challenges have been addressed throughout the development process to give a clearer understanding of the current state of the application and how these challenges may affect its implementation in specific contexts.

The fundamental principles of ethics that guide the implementation and deployment of this system should always be at the forefront for all stakeholders. These principles should be written down and formalized by different entities in order to ensure the integrity of the potential and reliability of this new solution. Among the very basic principles that both private and public entities should uphold including respecting and preserving the individual's integrity, ensuring security and transparency. These principles are at work in the fundamental elements of the *WalkingStreet* platform, such as the needed credentials to authorize any Application Programming Interface (API) operation, testing basic connectivity with a device and collecting and interpreting data. However, this author acknowledges that the proposed architecture will require a continuous adaptation of these principles depending on the scalability of its implementation. In other words, the application of these principles can vary significantly depending on factors such as the level of investment that the entities involved are willing to make, the number of modules they choose to implement (i.e., sensors or servers), or even the availability of other technological solutions on the market that are more profitable (some mentioned in the literature review) that have not been tested in the context of this work.

7.3 Discussion of Results

The Research Plan described in Section 1.5 reflects the dissemination and engagement plan often seen in grant applications. Along the course of this PhD project, this strategy was particularly relevant, encompassing activities such as publishing in journals, presenting at conferences, creating prototypes, and undertaking other scientific dissemination tasks. Moreover, in a research project, the need arises for a slightly more nuanced approach to knowledge consolidation of the project and its findings. It is important to note that this journey also involved actively supervising and monitoring master's students. These students had a crucial role in the engagement and development of various subtopics that leveraged this thesis. Lastly, active participation in scientific meetings and events also had tremendous importance in sharing and disclosing the findings of this PhD work. In summary, this section provides a comprehensive reviews of the outcomes resulting from the extensive work carried out during the course of this project.

7.3.1 Publications

To maximize its overall impact and trigger effects across its targeted stakeholders and communities, this work was disseminated through publications in an international journal, book chapters, and conference proceedings.

7.3.1.1 International Journals

The work has been documented in an international scientific journal. Here is the information for the published journal article:

- Rosa, L., Silva, F., & Analide, C. (2021). Mobile Networks and Internet of Things Infrastructures to Characterize Smart Human Mobility. International Journal of Smart Cities (Vol 4, pp. 894–918). DOI: `https://doi.org/10.3390/smartcities4020046`, URI: `https://hdl.handle.net/1822/74138`

7.3.1.2 Book Chapters

Further articles were developed for scientific publication purposes and are currently presented in the following book chapters:

- Rosa, L., Silva, F., & Analide, C. (2020). Representing Human Mobility Patterns in Urban Spaces. In: Fernández, C., Novella J., Ricci, A., Pinto, D., Roman, D (Eds.), Workshop Proceedings of the 16^{th} International Conference on Intelligent Environments, Intelligent Environments 2020 (pp. 410–420). IOS Press. DOI: `https://doi.org/10.3233/AISE200067`, URI: `https://hdl.handle.net/1822/83659`.

- Rosa, L., Silva, F., & Analide, C. (2020). IoT Architecture Proposal from a Survey of Pedestrian-Oriented Applications. Ambient Intelligence and Smart Environments, Intelligent Environments 2020 (Vol. 28, pp. 177–186). IOS Press. DOI: `https://doi.org/10.3233/AISE200040`, URI: `https://hdl.handle.net/1822/83690`.

- Rosa, L., Silva, F., & Analide, C. (2020). TMSA: Participatory Sensing Based on Mobile Phones in Urban Spaces. In: Analide. In: C., Novais, P., Camacho, D., Yin, H. (Eds.), Intelligent Data Engineering and Automated Learning 2020. 21^{th} International Conference. Lecture Notes in Computer Science (Vol 12489, pp. 257–267). Springer, Cham. DOI: `https://doi.org/10.1007/978-3-030-62362-3_23`, URI: `https://hdl.handle.net/1822/83692`.

- Rosa, L., Silva, F., & Analide, C. (2021). Mobile Networks and Internet of Things: Contributions to Smart Human Mobility. In: Dong, Y., Herrera-Viedma, E., Matsui, K., Omatsu, S., González Briones, A., Rodríguez González, S. (Eds.) Distributed Computing and Artificial Intelligence, 17^{th} International Conference. DCAI 2020. Advances in Intelligent Systems and Computing (Vol 1237, pp.168–178). Springer, Cham. DOI: `https://doi.org/10.1007/978-3-030-53036-5_18`, URI: `https://hdl.handle.net/1822/83696`.

- Rosa, L., Silva, F., & Analide, C. (2021). WalkingStreet: Understanding Human Mobility Phenomena Through a Mobile Application. In: et al. Intelligent Data Engineering and Automated Learning 2021. Lecture Notes in Computer Science (Vol 13113, pp. 599–610). Springer, Cham. DOI: `https://doi.org/10.1007/978-3-030-91608-4_58`, URI: `https://hdl.handle.net/1822/83694`.

- Rosa, L., Silva, F., & Analide, C. (2021). Urban Human Mobility Modelling and Prediction: Impact of Comfort and Well-Being Indicators. In: Marreiros, G., Melo, F.S., Lau, N., Lopes Cardoso, H., Reis, L.P. (Eds.) Progress in Artificial Intelligence. EPIA 2021. Lecture Notes in Computer Science (Vol 12981, pp. 278–289). Springer, Cham. DOI: `https://doi.org/10.1007/978-3-030-86230-5_22`, URI: `https://hdl.handle.net/1822/83699`.

- Rosa, L., Silva, F., & Analide, C. (2022). Analyzing Metrics to Understand Human Mobility Phenomena: Challenges and Solutions. In: , et al. Distributed Computing and Artificial Intelligence, Volume 2: Special Sessions 18^{th} International Conference. DCAI 2021. Lecture Notes in Networks and Systems (Vol 332, pp. 161–170). Springer, Cham. DOI: `https://doi.org/10.1007/978-3-030-86887-1_15`, URI: `https://hdl.handle.net/1822/83698`.

- Rosa, L., Faria, H., Tabrizi, R., Gonçalves, S., Silva, F., & Analide, C. (2022). Sentiment Analysis Based on Smart Human Mobility: A Comparative Study of ML Models. In: Ferrández Vicente, J.M., Álvarez-Sánchez, J.R., de la Paz López, F., Adeli, H. (Eds.) Bio-inspired Systems and Applications: from Robotics to Ambient Intelligence. IWINAC 2022. Lecture Notes in Computer Science (Vol

13259, pp. 55–64). Springer, Cham. DOI: `https://doi.org/10.1007/978-3-031-06527-9_6`, URI: `https://hdl.handle.net/1822/83703`.

- Rosa, L., Guimarães, M., Carneiro, D., Silva, F., & Analide, C. (2022). Explainable Decision Tree on Smart Human Mobility. Ambient Intelligence and Smart Environments. Workshops at 18^{th} International Conference on Intelligent Environments 2022 (Vol 31, pp. 325–334). IOS Press. DOI: `https://doi.org/10.3233/AISE220059`, URI: `https://hdl.handle.net/1822/85067`.
- Rosa, L., Silva, F., & Analide, C. (2023). Explainable Artificial Intelligence on Smart Human Mobility: a comparative study approach. In: Machado, J., Chamoso, P., Hernández, G., Bocewicz, G., Loukanova, R. Distributed Computing and Artificial Intelligence 2022, Special Sessions, 19^{th} International Conference. DOI: `https://doi.org/10.1007/978-3-031-23210-7_10`, URI: `https://hdl.handle.net/1822/83658`.

7.3.1.3 Conference Proceedings

Participation in conferences was also another scientific dissemination tool for communicating project- related information and findings to a wide research and scientific community. It also provided an opportunity to gather valuable input and feedback from key stakeholders and experts. The following list presents the conference proceedings published during this work:

- Rosa, L., Silva, F., & Analide, C. (2019). Smart Human Mobility in Smart Cities. Doctorate Symposium on Artificial Intelligence of 19^{th} Conference on Artificial Intelligence. EPIA 2019, University of Trás-os-Montes and Alto Douro (UTAD), Vila Real, Portugal. URI: `https://hdl.handle.net/1822/83674`.

7.3.2 Scientific Dissemination in Society

To promote the exchange of ideas, raise public awareness, facilitate knowledge sharing, and advance progress in this doctoral programme, the author of this thesis participated in various scientific and informal activities. These are the events and workshops attended:

- **Data Science Portugal Conference:** An exclusive event for students who have conducted work in the field of Data Science, held in Porto, Portugal. This conference provided a fantastic opportunity to present the ongoing work in the area of human mobility using ML algorithms.
- **Data Science Portugal Workshop:** An event open to students, researchers, and professionals, focusing on opening the Black Box and protecting it against Adversarial Attacks. Participation in this workshop allowed an understanding of how companies handle black boxes and how this knowl- edge

can later be replicated in the development of this work. Moreover, this course delved into the topic of explainability in ML, exploring Shapley Additive Explanations (SHAP) as a tool for explainability. This workshop took place in Braga, Portugal.

- **Data Science Portugal Talk:** Participation in a talk about relational databases to Big Data. In this initiative, we focused on the architecture designed for a Big Data stack, challenges encountered in the process, and potential solutions for the future. Key considerations for applications with years on the market versus newly developed ones, as well as on-premises versus cloud solutions were addressed. This event was held in Braga, Portugal.
- **Iberian Congress "The Bicycle and the City":** This congress was dedicated to the theme "The era of infrastructure", and focused on active mobility. Organized by the Portuguese Federation of Cycle Tourism and Bicycle Users, participation in this event provided insights into policies and good practices in the field of cycling and human mobility that are being designed and implemented in the cities. This congress took place in Barcelos, Portugal.
- **International Forum for Smart and Sustainable Communities:** Held in Braga, this forum offered an integrated and transversal vision of the latest advances in smart, sustainable, and inclusive regional and municipal development. It also highlighted global partnerships between companies and strategic partners, promoting synergies and dynamics of the business for SC.
- **Technological Challenges of Cities:** This conference, organized by the University of Minho in Braga, Portugal, provided valuable insights into the upcoming challenges faced by municipalities and regions regarding mobility. The author gained knowledge through exchange and reciprocal appreciation of the challenges ahead.

7.3.3 Invited Presentations

This work was also disseminated through invited tutorials and talk demonstrations. This is the detailed list of the oral presentations made during the doctoral programme:

- **Doctoral Program on Informatics:** Multiple presentations to Informatics Master's Students, showing the potential of ML models in solving SC problems and expanding their knowledge in this field.
- **Human Mobility and Smart Cities:** Presentation to master's students, suggesting subtopics for theses as potential areas of study within the context of graduate master's programs.

- **Summer on Campus Program:** Presentation to middle and high school students. This program, developed by the University of Minho in Portugal, aims to promote culture, science, art, and literature among young people and help students in their decision-making process when choosing a career path in Higher Education.

7.3.4 Supervision of Students

During the doctoral programme, the candidate co-supervised Master's students. Attending the Integrated Project curricular unit for the Master's degree in Computer Engineering or writing a Dissertation for the Master's degree in Management at the University of Minho, these students made an essential contribution to developed work in the context of this PhD thesis. In group or in a particular way, these students worked on various projects that are presented in the following list:

7.3.4.1 Recommendation system about human mobility during Covid-19

A recommendation system platform to understand human mobility during the Covid-19 pandemic period was built by Daniel Pereira and Luís Freitas. Primarily, its goal was to educate and help understanding the Covid-19 disease. To achieve this, the platform was designed to analyze the current pandemic situation and to understand its evolution. It incorporated data visualizations, predictive models, and a recommendation system. The implementation required sourcing data from various outlets, including disease data (i.e., cases, deaths, etc.), vaccination data (number of people fully vaccinated, number of people vaccinated with just one dose, etc.), and mobility data. To obtain this data, we accessed the following sources (1) Cases and Deaths: European Center for Deceased and Control, and (2) Vaccination: Our World in Data.

Co-supervised with: Professor Cesar Analide

Year of completion: 2020

7.3.4.2 Emotion prediction based on external factors

This project was developed by Diogo Barroso and Tiago Gonçalves. Using ML algorithms, this project's objective was to create an Long Short-Term Memory (LSTM) time series prediction model capable of predicting the emotions that people attending specific places in the city of Porto might experience over the next few hours. Only places from the city of Porto were selected. Factors considered in these predictions included weather conditions, event occurrences at the location, and historical emotional data for the place over the previous year. It's important to note that this project was primarily a Proof of Concept (PoC), designed to assess the feasibility of predicting subjective emotions in such settings. Therefore, some values, particularly those related to emotion, were fictional.

Co-supervised with: Professor Cesar Analide

Year of completion: 2021

7.3.4.3 Human mobility prediction in New York City using Third-party APIs

In this project, three master's students—Simão Gonçalves, Hugo Alves de Faria and Mohammad Tabrizi—implemented an analysis and prediction system based on crowd density in New York City . The project proposed a solution for SC to develop innovative responses to enhance the lives of the city's inhabitants. It emphasized placing the citizen at the center of urban development and planning through technology and communication. By applying various Machine Learning tools, the project aimed to predict population density in the city for future hours, providing significant advantages for city residents and tourists alike.

Co-supervised with: Professor Cesar Analide

Year of completion: 2022

7.3.4.4 A deep study about human mobility metrics

Eduardo Cardoso is currently working on this project, focusing on the use of AI models in the context of human mobility to optimize and predict its functioning. The research involves studying, understanding, and modelling metrics related to human mobility. The ultimate goal is to predict prob- lematic situations, identify safe places, and provide recommendations based on individual feelings. ML algorithms, such as LSTM and CNN, are employed for this purpose. The project includes several key objectives, including defining and implementing a data model to facilitate building a relational database, identifying the most relevant variables for the case under study, applying DL algorithms, and providing comprehensive explanations of the ML and Metrics steps. These explanations encompass data pre-processing, model training, prediction, and include relevant graphs and values for a thorough analysis of results. Additionally, scientific dissemination of the findings is planned.

Co-supervised with: Professor Cesar Analide

Year of completion: Currently ongoing

Bibliography

[1] D. M. Abbas. *Sensing-as-a-Service: The New Internet of Things (IOT) Business Model*. 2014. url: https://www.slideshare.net/mazlan1/sensing-asaservice-the-new-internet-of-things-iot-business-model (visited on 10/05/2023).

[2] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. E. Mohamed, and H. Arshad. *State-of-the-art in artificial neural network applications: A survey*. 2018. doi: 10.1016/j.heliyon.2018.e00938. url: https://www.sciencedirect.com/science/article/pii/S2405844018332067.

[3] M. Abrahamson. *Urban Sociology: A Global Introduction*. 2013, p. 282. isbn: 978-0-521-13923-6. url: http://www.amazon.com/Urban-Sociology-A-Global-Introduction/dp/0521139236.

[4] F. Agostinho, P. Pierucci, T. Fonseca, C. M. Almeida, and B. F. Giannetti. "What Makes Cities Sustainable? Empirical Evidence From a Brazilian Context." In: *Frontiers in Sustainable Cities* 4 (2022), p. 71. issn: 26249634. doi: 10.3389/frsc.2022.862956.

[5] R. Agrawal, T. Imielinski, and A. Swami. "Database Mining: A Performance Perspective." In: *IEEE Transactions on Knowledge and Data Engineering* 5.6 (1993), pp. 914 –925. doi: 10.1109/69.250074.

[6] K. Ahmad, M. Maabreh, M. Ghaly, K. Khan, J. Qadir, and A. Al-Fuqaha. "Developing Future Human-Centered Smart Cities: Critical Analysis of Smart City Security, Interpretability, and Ethical Challenges." In: *Computer Science Review* 43.C (2022), p. 76. doi: https://doi.org/10.1016/j.cosrev.2021.100452. arXiv: 2012.09110.

[7] A. K. Alamoudi, R. B. Abidoye, and T. Y. Lam. "The Impact of Citizens' Participation Level on Smart Sustainable Cities Outcomes: Evidence from Saudi Arabia." In: *Buildings* 13.2 (2023), p. 343. issn: 20755309. doi: 10.3390/buildings13020343. url: https://www.mdpi.com/2075-5309/13/2/343.

[8] E. C. Alexopoulos. "Introduction to multivariate regression analysis." In: *Hippokratia* 14.Suppl 1 (2010), pp. 23–8. issn: 1790-8019. url: http://www.ncbi.nlm.nih.gov/pubmed/21487487.

[9] A. A. Alkhatib, K. A. Maria, S. AlZu'bi, and E. A. Maria. "Smart Traffic Scheduling for Crowded Cities Road Networks." In: *Egyptian Informatics Journal* 23.4 (2022), pp. 163–176. issn: 11108665. doi: 10.1016/j.eij.2022.10.002.

[10] S. Ameer, M. A. Shah, A. Khan, H. Song, C. Maple, S. U. Islam, and M. N. Asghar. "Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities." In: *IEEE Access* 7 (2019), pp. 128325–128338. issn: 21693536. doi: 10.1109/ACCESS.2019.2925082.

[11] D. Andreola Serraglio. "The LFDD–Human Mobility Nexus in Africa, Latin America and the Caribbean." In: *die-gdi.de* (2020). doi: 10.23661/dp22.2020. url: www.die-gdi.dehttps://www.die-gdi.de/uploads/media/DP_22.2020.pdf.

[12] S. Angelo, M. Agriesti, R.-M. Soe, and M. A. Saif. "Framework for connecting the mobility challenges in low density areas to smart mobility solutions: the case study of Estonian municipalities." In: *European Transport Research Review 2022 14:1* 14.1 (2022), pp. 1–24. issn: 1866-8887. doi: 10.1186/S12544-022-00557-Y. url: https://etrr.springeropen.com/articles/10.1186/s12544-022-00557-y.

[13] E. Artur and R. Minghim. "A novel visual approach for enhanced attribute analysis and selection." In: *Computers and Graphics (Pergamon)* 84 (2019), pp. 160–172. issn: 00978493. doi: 10.1016/j.cag.2019.08.015.

[14] A. Asthana. *The 2022 API Platform Landscape: Trends and Challenges*. 2022. url: https://blog.postman.com/2022-api-platform-landscape-trends-and-challenges/ (visited on 04/25/2023).

[15] A. Athanasopoulou, T. Deijkers, B. Ozkan, and O. Turetken. "MaaS platform features: An exploration of their relationship and importance from supply and demand perspective." In: *Journal of Urban Mobility* 2 (2022), p. 100028. issn: 26670917. doi: 10.1016/j.urbmob.2022.100028.

[16] H. Attaran, N. Kheibari, and D. Bahrepour. "Toward integrated smart city: a new model for implementation and design challenges." In: *GeoJournal* 87.4 (2022), pp. 511–526. issn: 15729893. doi: 10.1007/s10708-021-10560-w. url: https://link.springer.com/article/10.1007/s10708-021-10560-w.

[17] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini. *Human mobility: Models and applications*. 2018. doi: 10.1016/j.physrep.2018.01.001. arXiv: 1710.00004.

[18] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, M. Tomasini, and H. Barbosa-Filho. "Human mobility: Models and applications." In: *Physics Reports* 734.734 (2018), pp. 1–74. doi: 10.1016/j.physrep.2018.01.001ï. arXiv: 1710.00004v1. url: https://shs.hal.science/halshs-01813382.

[19] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera. "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI." In: *Information Fusion* 58 (2020), pp. 82–115. issn: 15662535. doi: 10.1016/j.inffus.2019.12.012. arXiv: 1910.10045.

[20] A. Bartschat, M. Reischl, and R. Mikut. "Data mining tools." In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 9.4 (2019). issn: 19424795. doi: 10.1002/widm.1309. url: https://www.educba.com/data-mining-tool/.

[21] N. Baucells Aletà. "Smart Mobility in Smart Cities." In: Valencia, Spain, 2016, pp. 1209–1219. doi: 10.4995/cit2016.2016.3485. url: https://core.ac.uk/download/pdf/148684891.pdf.

[22] R. Bauer, J. Buss, and L. Chacko. *Mobility as a Service: A sum of parts present today*. Tech. rep. 2020. url: https://www.oliverwyman.com/content/dam/oliver-wyman/global/en/images/insights/automotive/2020/Mobility_as_a_service.pdf.

[23] S. Baxter, A. Barnes, C. Lee, R. Mead, and M. Clowes. "Increasing public participation and influence in local decision-making to address social determinants of health: a systematic review examining initiatives and theories." In: *Local Government Studies* (2022). issn: 17439388. doi: 10.1080/03003930.2022.2081551. url: https://www.tandfonline.com/doi/abs/10.1080/03003930.2022.2081551.

[24] A. Bentahar, A. Meraoumia, L. Bradji, and H. Bendjenna. "Sensing as a service in Internet of Things: Efficient authentication and key agreement scheme." In: *Journal of King Saud University - Computer and Information Sciences* 34.8 (2022), pp. 5493–5509. issn: 22131248. doi: 10.1016/j.jksuci.2021.06.007.

[25] Bharani. *What is IoT (Internet of Things)? IoT Architecture Explained*. 2018. url: https://www.edureka.co/blog/what-is-iot/ (visited on 03/13/2019).

[26] Big Data LDN. *Improving Big Data Analytics with Machine Learning | BIG DATA LDN*. 2016. url: http://bigdataldn.com/improving-big-data-analytics-with-machine-learning/?utm_source=datafloq&utm_medium=ref&utm_campaign=datafloq (visited on 03/13/2019).

[27] B. Bonthron, J. Bornstein, T. Fishman Dovey, and W. Goodall. "The rise of mobility as a service - Reshaping how urbanites get around." In: *Deloitte Review* 20 (2017), pp. 111–130. url: https://www2.deloitte.com/nl/nl/pages/consumer-industrial-products/articles/the-rise-of-mobility-as-a-service.html.

[28] K. Borsekova, A. Vanova, and K. Vitalisova. “The Power of Communities in Smart Urban Development.” In: *Procedia - Social and Behavioral Sciences* 223 (2016), pp. 51–57. issn: 18770428. doi: 10.1016/j.sbspro.2016.05.289.

[29] F. Botta. “Quantifying the differences in call detail records.” In: *Royal Society Open Science* 8.6 (2021). issn: 20545703. doi: 10.1098/rsos.201443. url: https://royalsocietypublishing.org/doi/10.1098/rsos.201443.

[30] H. Bradbury, S. Waddell, K. O’ Brien, M. Apgar, B. Teehankee, and I. Fazey. *A call to Action Research for Transformations: The times demand it*. 2019. doi: 10.1177/1476750319829633. url: https://www.researchgate.net/publication/331795580_A_call_to_Action_Research_for_Transformations_The_times_demand_it.

[31] H. Bradley. *Bringing History to LinkNYC Kiosks*. 2018. url: https://medium.com/urban-archive/bringing-history-to-linknyc-kiosks-17f7d63a4129 (visited on 04/20/2019).

[32] M. Brockerhoff and U. Nations. *World Urbanization Prospects: The 1996 Revision*. Tech. rep. 4. New York: United Nations, 2006, p. 883. doi: 10.2307/2808041. arXiv: arXiv:1011.1669v3. url: http://www.un.org/en/development/desa/population/publications/pdf/urbanization/WUP2011_Report.pdf.

[33] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira, and C. Ratti. “Understanding individual mobility patterns from urban sensing data: A mobile phone trace example.” In: *Transportation Research Part C: Emerging Technologies* 26 (2013), pp. 301–313. issn: 0968090X. doi: 10.1016/j.trc.2012.09.009. url: https://www.sciencedirect.com/science/article/pii/S0968090X12001192.

[34] F. Calderón and E. J. Miller. “A literature review of mobility services: definitions, modelling state-of-the-art, and key considerations for a conceptual modelling framework.” In: *Transport Reviews* 40.3 (2020), pp. 312–332. issn: 14645327. doi: 10.1080/01441647.2019.1704916.

[35] S. Callegaro, R. Alves Palacios, and E. Mezzomo Luciano. “A discussion about Action Research studies and their variations in Smart Cities and the challenges in Latin America.” In: *Revista Latinoamericana Y Del Caribe De La Associacion De Sistemas De Informacion* 14.1 (2022), pp. 1–15. doi: 10.17705/1relc.00088. url: https://aisel.aisnet.org/relcasi.

[36] I. Calzada, M. Pérez-Batlle, and J. Batlle-Montserrat. “People-Centered Smart Cities: An exploratory action research on the Cities’ Coalition for Digital Rights.” In: *Journal of Urban Affairs* (2021). issn: 14679906. doi: 10.1080/07352166.2021.1994861. url: https://www.tandfonline.com/doi/abs/10.1080/07352166.2021.1994861.

[37] A. Caragliu, C. del Bo, and P. Nijkamp. “Smart cities in Europe.” In: *Journal of Urban Technology* 18.2 (2011), pp. 65–82. issn: 10630732. doi: 10.1080/10630732.2011.601117. arXiv: arXiv:1011.1669v3.

[38] D. Carneiro, F. Silva, M. Guimarães, D. Sousa, and P. Novais. “Explainable intelligent environments.” In: *Advances in Intelligent Systems and Computing*. Vol. 1239 AISC. Springer Science and Business Media Deutschland GmbH, 2021, pp. 34–43. isbn: 9783030583552. doi: 10.1007/978-3-030-58356-9_4.

[39] C. B. Casady. “Customer-led mobility: A research agenda for Mobility-as-a-Service (MaaS) enablement.” In: *Case Studies on Transport Policy* 8.4 (2020), pp. 1451–1457. issn: 22136258. doi: 10.1016/j.cstp.2020.10.009. url: https://www.researchgate.net/publication/345578565_Customer-Led_Mobility_A_Research_Agenda_for_Mobility-as-a-Service_MaaS_Enablement.

[40] O. Cats. “Identifying Human Mobility Patterns using Smart Card Data.” In: (2022). doi: 10.48550/arxiv.2208.05352. arXiv: 2208.05352. url: https://arxiv.org/abs/2208.05352v1.

[41] W. Chatti and M. T. Majeed. “Information communication technology (ICT), smart urbanization, and environmental quality: Evidence from a panel of developing and developed economies.” In: *Journal of Cleaner Production* 366 (2022), p. 132925. issn: 09596526. doi: 10.1016/j.jclepro.2022.132925.

[42] M. Chen, S. Mao, and Y. Liu. “Big data: A survey.” In: *Mobile Networks and Applications*. Vol. 19. 2. 2014, pp. 171–209. isbn: 1383-469X. doi: 10.1007/s11036-013-0489-0.

[43] C. Ciavarella and N. M. Ferguson. “Deriving fine-scale models of human mobility from aggregated origin-destination flow data.” In: *PLoS Computational Biology* 17.2 (2021), e1008588. issn: 15537358. doi: 10.1371/JOURNAL.PCBI.1008588. url: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008588.

[44] E. Commission. *Smart cities*. 2022. url: https://ec.europa.eu/info/es-regionu-ir-miestu-pletra/temos/miestai-ir-miestu-pletra/miestu-iniciatyvos/smart-cities_en (visited on 08/30/2022).

[45] E. C. f. D. P. and Control. *Surveillance data from public online national reports on COVID-19 in long-term care facilities*. 2021. url: https://www.ecdc.europa.eu/en/all-topics-z/coronavirus/threats-and-outbreaks/covid-19/prevention-and-control/LTCF-data (visited on 01/12/2023).

[46] A. Corsi, A. Corsi, R. Pagani, T. B. R. e. Cruz, F. F. de Souza, and J. L. Kovaleski. “Smart Sustainable Cities: Characterization and Impacts for Sustainable Development Goals.” In: *Revista de Gestão Ambiental e Sustentabilidade* 11.1 (2022), e20750. issn: 2316-9834. doi: `10.5585/geas.v11i1.20750`. url: `https://periodicos.uninove.br/geas/article/view/20750`.

[47] B. Culp. *Cisco Survey Reveals Close to Three-Fourths of IoT Projects Are Failing*. 2017. url: `https://newsroom.cisco.com/press-release-content?articleId=1847422` (visited on 07/07/2019).

[48] C. D. Osborne. *The definitive list of smart cities projects taking the world by storm*. 2016. url: `https://www.iotworldtoday.com/2017/09/28/definitive-list-smart-cities-projects-taking-world-storm/` (visited on 03/13/2019).

[49] B. Dahiya. “Southeast Asia and Sustainable Urbanization.” In: *Global Asia* 9.3 (2014), pp. 84–91. doi: `https://www.researchgate.net/publication/283463332_Southeast_Asia_and_Sustainable_Urbanization`.

[50] Dairenkon Majime. *Best Practices When Working With Docker for Machine Learning*. 2022. url: `https://neptune.ai/blog/best-practices-docker-for-machine-learning` (visited on 11/09/2022).

[51] B. Das. “An overview on big data: characteristics, security and applications.” In: *International Journal of Computer Applications* 154.10 (2020), pp. 29–35. url: `https://www.researchgate.net/publication/350075342_An_overview_on_big_data_characteristics_security_and_applications`.

[52] H. M. Data. *The Use of Human Mobility Data in Public Health Emergencies*. 2021. url: `https://www.crisisready.io/2022/the-use-of-mobility-data-in-public-health-emergencies/` (visited on 09/08/2022).

[53] O. Data Mining Concepts. *Regression*. 2011. url: `https://docs.oracle.com/cd/E24693_01/datamine.11203/e16808/regress.htm` (visited on 06/22/2019).

[54] DataGrip. *Connect to Google Cloud SQL instances*. 2022. url: `https://www.jetbrains.com/help/datagrip/connect-to-google-cloud-sql-instances.html#configuring-google-cloud-sdk` (visited on 11/15/2022).

[55] Datareportal. *Digital 2022: Time Spent Using Connected Tech Continues to Rise — DataReportal – Global Digital Insights*. 2022. url: `https://datareportal.com/reports/digital-2022-time-spent-with-connected-tech` (visited on 05/24/2022).

[56] A. De Mauro, M. Greco, and M. Grimaldi. "What is big data? A consensual definition and a review of key research topics." In: *AIP Conference Proceedings*. Vol. 1644. 2015, pp. 97–104. isbn: 9780735412835. doi: 10.1063/1.4907823. arXiv: arXiv:1804.02998v1.

[57] A. De Mauro, M. Greco, and M. Grimaldi. "A formal definition of Big Data based on its essential features." In: *Library Review* 65.3 (2016), pp. 122–135. issn: 00242535. doi: 10.1108/LR-06-2015-0061. arXiv: aimag.v17i3.1230.

[58] Deloitte. *Smart Cities: The importance of a smart ICT infrastructure for smart cities*. 2017. url: https://www.stokab.se/Documents/Nyheterbilagor/SmartCityInfraEn.pdf (visited on 04/06/2019).

[59] I. Digel, D. Mussabalina, M. Urdabayev, N. Nurmukhametov, and A. Akparova. "Evaluating development prospects of smart cities: Cluster analysis of Kazakhstan's regions." In: *Problems and Perspectives in Management* 20.4 (2022), pp. 76–87. issn: 18105467. doi: 10.21511/ppm.20(4).2022.07. url: http://dx.doi.org/10.21511/ppm.20.

[60] E. Dumbill. "Planning for Big Data: A CIO's Handbook to the Changing Data Landscape." In: *O'Reilly Media*. O'Reilly Media, 2012, p. 83. isbn: 978-1-4493-2963-1.

[61] U. Ehsan, Q. V. Liao, M. Muller, M. O. Riedl, and J. D. Weisz. "Expanding explainability: Towards social transparency in ai systems." In: *Conference on Human Factors in Computing Systems - Proceedings*. 2021. isbn: 9781450380966. doi: 10.1145/3411764.3445188. arXiv: 2101.04719. url: https://doi.org/10.1145/3411764.3445188.

[62] Y. Ennali and T. Van Engers. "Data-driven AI development: An integrated and iterative bias mitigation approach." In: *CEUR Workshop Proceedings*. Vol. 2891. 2021.

[63] Euro Cities. *Smart cities and digital transformation*. 2021. url: https://eurocities.eu/goals/smart-cities-digital-transformation/ (visited on 09/13/2022).

[64] European Commission. *Question and Answer: Europe Urban Mobility Framework*. 2021. url: https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_6729 (visited on 08/30/2022).

[65] Y. Fang and Z. Shan. "How to Promote a Smart City Effectively? An Evaluation Model and Efficiency Analysis of Smart Cities in China." In: *Sustainability (Switzerland)* 14.11 (2022), p. 6512. issn: 20711050. doi: 10.3390/su14116512. url: https://www.mdpi.com/2071-1050/14/11/6512.

[66] Z. Fang, S. L. Shaw, B. Yang, P. Santi, and W. Tu. "Integrated environmental and human observations for smart cities." In: *Environment and Planning B: Urban Analytics and City Science* 48.6 (2021), pp. 1375–1379. issn: 23998091. doi: 10.1177/23998083211023296. url: https://journals.sagepub.com/doi/full/10.1177/23998083211023296.

[67] J. Ferreira, J. Fonseca, D. Gomes, J. Barraca, B. Fernandes, J. Rufino, J. Almeida, and R. Aguiar. "PASMO: An open living lab for cooperative ITS and smart regions." In: *2017 International Smart Cities Conference, ISC2 2017*. Institute of Electrical and Electronics Engineers Inc., 2017. isbn: 9781538625231. doi: 10.1109/ISC2.2017.8090866. url: https://www.researchgate.net/publication/320831045_PASMO_An_open_living_lab_for_cooperative_ITS_and_smart_regions.

[68] E. Fidje, M. Haddara, and M. Langseth. "Sustainable Smart Waste Management Adoption Challenges in Developing Countries." In: *Procedia Computer Science* 219 (2023), pp. 905–914. issn: 1877-0509. doi: 10.1016/J.PROCS.2023.01.366. url: https://linkinghub.elsevier.com/retrieve/pii/S1877050923003757.

[69] G. A. Fink, T. W. Edgar, T. R. Rice, D. G. MacDonald, and C. E. Crawford. *Security and Privacy in Cyber-Physical Systems*. 2016, pp. 129–141. isbn: 9780128038741. doi: 10.1016/B978-0-12-803801-7.00009-2. url: http://dx.doi.org/10.1016/B978-0-12-803801-7.00009-2.

[70] D. Fonseca, M. Sanchez-Sepulveda, S. Necchi, and E. Peña. "Towards smart city governance. Case study: Improving the interpretation of quantitative traffic measurement data through citizen participation." In: *Sensors* 21.16 (2021), p. 5321. issn: 14248220. doi: 10.3390/s21165321. url: https://www.mdpi.com/1424-8220/21/16/5321.

[71] C. Forsyth. *For Big Data Analytics There's No Such Thing as Too Big: The Compelling Economics and Technology of Big Data Computing*. Tech. rep. March. Cisco, 2012, pp. 1–20. url: http://www.cisco.com/c/dam/en/us/solutions/data-center-virtualization/big_data_wp.pdf.

[72] I. T. Forum. *Leveraging digital technology and data for human-centric smart cities: the case of smart mobility*. 2020. url: http://www.oecd.org/termsandconditions. (visited on 04/21/2023).

[73] R. Gall. *Machine Learning Explainability vs Interpretability: Two concepts that could help restore trust in AI*. 2018. url: https://www.kdnuggets.com/2018/12/machine-learning-explainability-interpretability-ai.html (visited on 10/03/2022).

[74] J. Gantz and D. Reinsel. *Extracting Value from Chaos*. Tech. rep. IDC IVIEW, 2011, p. 12. doi: 10.1007/s10916-016-0565-7. arXiv: 1312.4722.

[75] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera. "Big data preprocessing: methods and prospects." In: *Big Data Analytics* 1.1 (2016), p. 9. issn: 2058-6345. doi: 10.1186/s41044-016-0014-0. url: http://bdataanalytics.biomedcentral.com/articles/10.1186/s41044-016-0014-0.

[76] F. J. García-Peñalvo. "Avoiding the dark side of digital transformation in teaching. an institutional reference framework for eLearning in higher education." In: *Sustainability (Switzerland)* 13.4 (2021), pp. 1–17. issn: 20711050. doi: 10.3390/su13042023. url: https://ror.org/02f40zc51.

[77] P. Giganti and P. M. Falcone. *Strategic Niche Management for Sustainability: A Systematic Literature Review*. 2022. doi: 10.3390/su14031680.

[78] Global Footprint Network. *Open Data Platform Services*. 2021. url: https://data.footprintnetwork.org (visited on 04/25/2023).

[79] S. Goethals, D. Martens, and T. Evgeniou. "The non-linear nature of the cost of comprehensibility." In: *Journal of Big Data* 9.1 (2022), pp. 1–23. issn: 21961115. doi: 10.1186/s40537-022-00579-2. url: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00579-2.

[80] Google Developers. *What is Pub/Sub?* 2020. url: https://cloud.google.com/pubsub/docs/overview (visited on 12/13/2022).

[81] J. Grimond. *A Survey of Cities: The World goes to Town*. 2007. url: https://www.economist.com/special-report/2007/05/03/the-world-goes-to-town (visited on 03/13/2019).

[82] D. Gunning and D. W. Aha. "DARPA's explainable artificial intelligence (XAI) program." In: *AI Magazine* 40.2 (2019), pp. 44–58. issn: 07384602. doi: 10.1609/aimag.v40i2.2850. url: https://dl.acm.org/doi/10.1145/3301275.3308446.

[83] G. H. R. and P. S. Aithal. "How to Choose an Appropriate Research Data Collection Method and Method Choice among Various Research Data Collection Methods and Method Choices during Ph.D. Program in India?" In: *SSRN Electronic Journal* (2022). doi: 10.2139/ssrn.4275696. url: https://papers.ssrn.com/abstract=4275696.

[84] K. Haresamudram, S. Larsson, and F. Heintz. "Three Levels of AI Transparency." In: *Computer* 56.2 (2023), pp. 93–100. issn: 15580814. doi: 10.1109/MC.2022.3213181.

[85] J. Hartley. "Innovation in governance and public services: Past and present." In: *Public Money and Management* 25.1 (2005), pp. 27–34. issn: 14679302. doi: 10.1111/j.1467-9302.2005.00447.x.

[86] D. A. Hensher and C. Mulley. "Hensher, D.A. and Mulley, C. Mobility bundling and cultural tribalism - Might passenger mobility plans through MaaS remain niche or are they truly scalable?" In: *Transport Policy* 100 (2021), pp. 172–175. issn: 1879310X. doi: 10.1016/j.tranpol.2020.11.003.

[87] H Herdiansyah. “Smart city based on community empowerment, social capital, and public trust in urban areas.” In: *Global Journal of Environmental Science and Management* 9.1 (2023), pp. 113–128. issn: 23833866. doi: 10.22034/gjesm.2023.01.09.

[88] J. C. Herrera, D. B. Work, R. Herring, X. J. Ban, Q. Jacobson, and A. M. Bayen. “Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment.” In: *Transportation Research Part C: Emerging Technologies* 18.4 (2010), pp. 568–583. issn: 0968090X. doi: 10.1016/j.trc.2009.10.006. url: https://linkinghub.elsevier.com/retrieve/pii/S0968090X09001430.

[89] K. Holländer, A. Krüger, and A. Butz. “Save the smombies: App-assisted street crossing.” In: *Conference Proceedings - 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services: Expanding the Horizon of Mobile Interaction, MobileHCI 2020*. Vol. 11. ACM, 2020. isbn: 9781450375160. doi: 10.1145/3379503.3403547. url: https://doi.org/10.1145/3379503.3403547.

[90] R. G. Hollands. “Will the real smart city please stand up? : Intelligent, progressive or entrepreneurial?” In: *The Routledge Companion to Smart Cities* 1.112 (2020), pp. 179–199. doi: 10.4324/9781315178387-13.

[91] I. Hong, W. S. Jung, and H. H. Jo. “Gravity model explained by the radiation model on a population landscape.” In: *PLoS ONE* 14.6 (2019), e0218028. issn: 19326203. doi: 10.1371/journal.pone.0218028. arXiv: 1803.09067. url: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0218028.

[92] M. Hosseinzadeh, A. Hemmati, and A. M. Rahmani. “Clustering for smart cities in the internet of things: a review.” In: *Cluster Computing* 25.6 (2022), pp. 4097–4127. issn: 15737543. doi: 10.1007/s10586-022-03646-8. url: https://www.researchgate.net/publication/361577082_Clustering_for_smart_cities_in_the_internet_of_things_a_review.

[93] H. Huang, X. A. Yao, J. M. Krisp, and B. Jiang. “Analytics of location-based big data for smart cities: Opportunities, challenges, and future directions.” In: *Computers, Environment and Urban Systems* 90 (2021), p. 101712. issn: 01989715. doi: 10.1016/j.compenvurbsys.2021.101712.

[94] A. Huss, J. Beekhuizen, H. Kromhout, and R. Vermeulen. “Using GPS-derived speed patterns for recognition of transport modes in adults.” In: *International Journal of Health Geographics* 13.1 (2014), p. 40. issn: 1476072X. doi: 10.1186/1476-072X-13-40. url: http://www.ncbi.nlm.nih.gov/pubmed/25304171.

[95] O. o. T. Innovation. *What LinkNYC kiosks offer*. url: https://www.nyc.gov/content/oti/pages/linknychttps://www.nyc.gov/content/oti/pages/ (visited on 03/10/2023).

[96] O. D. Institute. *Challenges for open data companies*. 2016. url: https://theodi.org/article/research-open-data-means-business-pg4/ (visited on 04/25/2023).

[97] W. Institute. *State of the World 2007: Our Urban Future*. Ed. by Linda Starke. New York, London: W. W. Norton and Company, 2007, pp. 48–65. isbn: 039306557X.

[98] InteliLIGHT. *Smart lighting paves the way for smart cities all over the world*. 2021. url: https://intelilight.eu/smart-street-lighting-paves-the-way-for-smart-cities-all-over-the-world/ (visited on 04/17/2023).

[99] R. V. Ionescu, M. L. Zlati, and V. M. Antohi. "Smart cities from low cost to expensive solutions under an optimal analysis." In: *Financial Innovation* 9.1 (2023), pp. 1–34. issn: 21994730. doi: 10.1186/s40854-023-00448-8. url: https://jfin-swufe.springeropen.com/articles/10.1186/s40854-023-00448-8.

[100] Istio. *Istio*. 2022. url: https://istio.io/ (visited on 10/31/2022).

[101] ITU. *ITU-T focus group on smart sustainable cities.* Tech. rep. 2016. url: https://www.itu.int/dms_pub/itu-t/oth/0b/04/T0B0400004F2C01PDFE.pdf.

[102] A. R. Javed, W. Ahmed, S. Pandya, P. K. R. Maddikunta, M. Alazab, and T. R. Gadekallu. "A Survey of Explainable Artificial Intelligence for Smart Cities." In: *Electronics 2023, Vol. 12, Page 1020* 12.4 (2023), p. 1020. issn: 2079-9292. doi: 10.3390/ELECTRONICS12041020. url: https://www.mdpi.com/2079-9292/12/4/1020https://www.mdpi.com/2079-9292/12/4/1020/htm.

[103] M. H. Kabir, K. F. Hasan, M. K. Hasan, and K. Ansari. "Explainable Artificial Intelligence for Smart City Application: A Secure and Trusted Platform." In: *Studies in Computational Intelligence*. Vol. 1025. 2022, pp. 241–263. doi: 10.1007/978-3-030-96630-0_11. arXiv: 2111.00601. url: https://www.researchgate.net/publication/355843171_Explainable_Artificial_Intelligence_for_Smart_City_Application_A_Secure_and_Trusted_Platform.

[104] U. Kamath and J. Liu. *Explainable Artificial Intelligence: An Introduction to Interpretable Machine Learning*. Springer International Publishing, 2021. doi: 10.1007/978-3-030-83356-5.

[105] H. Kang, G. Lee, and J. Han. "Obstacle detection and alert system for smartphone AR users." In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*. Vol. 11. Association for Computing Machinery, 2019. isbn: 9781450370011. doi: 10.1145/3359996.3364256. url: https://dl.acm.org/doi/10.1145/3359996.3364256.

[106] D. Karamshuk, C. Boldrini, M. Conti, and A. Passarella. "SPoT: Representing the social, spatial, and temporal dimensions of human mobility with a unifying framework." In: *Pervasive and Mobile Computing* 11 (2014), pp. 19–40. issn: 15741192. doi: 10.1016/j.pmcj.2013.07.011.

[107] H. Kim, H. J. Kim, J. Park, J. K. Ryu, and S. C. Kim. “Recognition of fine-grained walking patterns using a smartwatch with deep attentive neural networks.” In: *Sensors* 21.19 (2021), p. 6393. issn: 14248220. doi: 10.3390/s21196393. url: https://www.mdpi.com/1424-8220/21/19/6393.

[108] C. E. Kontokosta, N. Johnson, B. Bonczak, B. Hong, A. Malik, P. Mandiola, and C. Blankinship. *Using High Resolution Mobility Data to Understand Human Behavior and Population Dynamics in Cities*. Tech. rep. 2020, p. 54.

[109] S Kumar. “The McKinsey 7S Model helps in Strategy implementation: A Theoretical Foundation.” In: *Tecnia Journal of Management Studies* 14.1 (2021), pp. 7–12. url: https://www.researchgate.net/publication/357352792_The_McKinsey_7S_Model_helps_in_Strategy_implementation_A_Theoretical_Foundation.

[110] A. KUZIOR and B. SOBOTKA. “THE ROLE OF SOCIAL CAPITAL IN THE DEVELOPMENT OF SMART CITIES.” In: *Scientific Papers of Silesian University of Technology. Organization and Management Series* 2019.134 (2019), pp. 109–119. issn: 16413466. doi: 10.29119/1641-3466.2019.134.9.

[111] C. Lamsfus, D. Martín, A. Alzua-Sorzabal, and E. Torres-Manzanera. “Smart Tourism Destinations: An Extended Conception of Smart Cities Focusing on Human Mobility.” In: *Information and Communication Technologies in Tourism 2015*. Springer International Publishing, 2015, pp. 363–375. doi: 10.1007/978-3-319-14343-9_27.

[112] C. Lechner, M. Rumpler, M. C. Dorley, Y. Li, A. Ingram, and H. Fryman. “Developing an Online Dashboard to Visualize Performance Data—Tennessee Newborn Screening Experience.” In: *International Journal of Neonatal Screening* 8.3 (2022), p. 49. issn: 2409515X. doi: 10.3390/ijns8030049. url: https://www.mdpi.com/2409-515X/8/3/49.

[113] J. Lee, D. Kim, and J. Park. “A Machine Learning and Computer Vision Study of the Environmental Characteristics of Streetscapes That Affect Pedestrian Satisfaction.” In: *Sustainability (Switzerland)* 14.9 (2022). issn: 20711050. doi: 10.3390/su14095730. url: https://www.researchgate.net/publication/360496396_A_Machine_Learning_and_Computer_Vision_Study_of_the_Environmental_Characteristics_of_Streetscapes_That_Affect_Pedestrian_Satisfaction.

[114] Y. Liao, S. Yeh, and G. S. Jeuken. “From individual to collective behaviours: exploring population heterogeneity of human mobility based on social media data.” In: *EPJ Data Science* 8.1 (2019). issn: 21931127. doi: 10.1140/epjds/s13688-019-0212-x.

[115] E. J. Liu and X. Y. Yan. *Research advances in intervening opportunity class models for predicting human mobility*. 2020. doi: 10.7498/aps.69.20201119. url: https://wulixb.iphy.ac.cn/en/article/doi/10.7498/aps.69.20201119.

[116] J. Liu, R. Chen, Y. Chen, L. Pei, and L. Chen. "iParking: An intelligent indoor location-based smartphone parking service." In: *Sensors (Switzerland)* 12.11 (2012), pp. 14612–14629. issn: 14248220. doi: 10.3390/s121114612. url: http://www.mdpi.com/1424-8220/12/11/14612.

[117] T. Liu, Z. Yang, Y. Zhao, C. Wu, Z. Zhou, and Y. Liu. "Temporal understanding of human mobility: A multi-time scale analysis." In: *PLoS ONE* 13.11 (2018). issn: 19326203. doi: 10.1371/journal.pone.0207697. url: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6258540/.

[118] I. C. c. Q. C. S. LLC. *LinkNYC Kiosks: Free super fast Wi-Fi and that's just the beginning.* 2021. url: https://www.link.nyc/ (visited on 02/24/2021).

[119] P. Lombardi, S. Giordano, H. Farouh, and W. Yousef. *Modelling the smart city performance*. 2012. doi: 10.1080/13511610.2012.660325. arXiv: arXiv:1011.1669v3.

[120] C. A. Luís Rosa, Fábio Silva. "IoT Architecture Proposal from a Survey of Pedestrian-Oriented Applications." In: *Intelligent Environments 2020: Workshop Proceedings of the 16th International Conference on Intelligent Environments*. Ed. by C. A. Iglesias, J. I. Moreno Novella, A. Ricci, D. Rivera Pinto, and D. Roman. Vol. 28. IOS Press, 2020, p. 177. isbn: 978-1-64368-090-3. url: https://dialnet.unirioja.es/servlet/articulo?codigo=7770779.

[121] S. M. Lundberg and S. I. Lee. "A unified approach to interpreting model predictions." In: *Advances in Neural Information Processing Systems*. Vol. 2017-Decem. Neural information processing systems foundation, 2017, pp. 4766–4775. doi: 10.48550/arxiv.1705.07874. arXiv: 1705.07874.

[122] R. Machlev, L. Heistrene, M. Perl, K. Y. Levy, J. Belikov, S. Mannor, and Y. Levron. "Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities." In: *Energy and AI* 9 (2022), p. 100169. issn: 26665468. doi: 10.1016/j.egyai.2022.100169.

[123] E. Macias, A. Suarez, and J. Lloret. *Mobile sensing systems*. 2013. doi: 10.3390/s131217292. url: http://www.mdpi.com/1424-8220/13/12/17292.

[124] S. Mahajan, M. K. Chung, J. Martinez, Y. Olaya, D. Helbing, and L. J. Chen. *Translating citizen-generated air quality data into evidence for shaping policy*. 2022. doi: 10.1057/s41599-022-01135-2. url: https://www.nature.com/articles/s41599-022-01135-2.

[125] A. Mahizhnan. *Smart cities*. Vol. 16. 1. 1999, pp. 13–18. isbn: 0042098013494. doi: 10.1016/S0264-2751(98)00050-X. arXiv: arXiv:1307.8198v1. url: http://linkinghub.elsevier.com/retrieve/pii/S026427519800050X.

[126] Z. J. Makieła, M. M. Stuss, K. Mucha-Kus, G. Kinelski, M. Budzinski, J. Michałek, K. Mucha-Kus, G. Kinelski, M. Budzinski, and J. Michałek. "Smart City 4.0: Sustainable Urban Development in the Metropolis GZM." In: *Sustainability (Switzerland)* 14.6 (2022), p. 3516. issn: 20711050. doi: 10.3390/su14063516. url: https://www.mdpi.com/2071-1050/14/6/3516.

[127] C. Mandl and T. Kirschner. "European Smart City Initiative Assessment of Best Practices to Stimulate Market-/ Demand-Pull." In: *BEHAVE Energy Conference 2014*. September. 2014. url: https://www.researchgate.net/publication/266208340_European_Smart_City_Initiative_Assessment_of_Best_Practices_to_Stimulate_Market-_Demand-Pull.

[128] H. Martin, Y. Hong, N. Wiedemann, D. Bucher, and M. Raubal. "Trackintel: An open-source Python library for human mobility analysis." In: *Computers, Environment and Urban Systems* 101 (2023), p. 101938. issn: 01989715. doi: 10.1016/j.compenvurbsys.2023.101938. arXiv: 2206.03593.

[129] E. Mathieu, H. Ritchie, E. Ortiz-Ospina, M. Roser, J. Hasell, C. Appel, C. Giattino, and L. Rodés-Guirao. "A global database of COVID-19 vaccinations." In: *Nature Human Behaviour* 5.7 (2021), pp. 947–953. issn: 23973374. doi: 10.1038/s41562-021-01122-8. url: https://ourworldindata.org/covid-vaccinations.

[130] M. Matyas and M. Kamargianni. "The potential of mobility as a service bundles as a mobility management tool." In: *Transportation* 46.5 (2019), pp. 1951–1968. issn: 15729435. doi: 10.1007/s11116-018-9913-4.

[131] Melody Williams. *Smart Waste Management for Cleaner, More Livable Cities*. 2021. url: https://nmbtc.com/blog/intelligent-waste-management-for-cleaner-more-livable-cities/.

[132] D. Miguel-Santiago, M. E. Rivero-Angeles, L. I. Garay-Jiménez, I. Y. Orea-Flores, and B. Tovar-Corona. "Teletraffic analysis of a mobile crowdsensing system: The pedestrian-to-vehicle scenario." In: *International Journal of Distributed Sensor Networks* 18.11 (2022). issn: 15501477. doi: 10.1177/15501329221133291. url: https://journals.sagepub.com/doi/full/10.1177/15501329221133291.

[133] N. Mileu, N. M. Costa, E. M. Costa, A. Alves, M. Schickler, F. Beierle, J. Schobel, N. Mileu, N. M. Costa, E. M. Costa, and A. Alves. "Mobility and Dissemination of COVID-19 in Portugal: Correlations and Estimates from Google's Mobility Data." In: *Data* 7.8 (2022), p. 107. issn: 23065729. doi: 10.3390/data7080107. url: https://www.mdpi.com/2306-5729/7/8/107/htm.

[134] C. Millette and P. Hosein. "A consumer focused open data platform." In: *2016 3rd MEC International Conference on Big Data and Smart City, ICBDSC 2016*. 2016, pp. 101–106. isbn: 9781509013654. doi: 10.1109/ICBDSC.2016.7460350. arXiv: 1606.05832.

[135] Miniwatts Marketing Group. *World Internet Users and Population Stats*. 2012. url: http://www.internetworldstats.com/stats.htm (visited on 03/13/2019).

[136] L. Mitropoulos, A. Kortsari, V. Mizaras, and G. Ayfantopoulou. "Mobility as a Service (MaaS) Planning and Implementation: Challenges and Lessons Learned." In: *Future Transportation 2023, Vol. 3, Pages 498-518* 3.2 (2023), pp. 498–518. issn: 2673-7590. doi: 10.3390/FUTURETRANSP3020029. url: https://www.mdpi.com/2673-7590/3/2/29.

[137] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani. *Deep learning for IoT big data and streaming analytics: A survey*. 2018. doi: 10.1109/COMST.2018.2844341. arXiv: 1712.04301.

[138] S. P. Mohanty, U. Choppali, and E. Kougianos. "Everything you wanted to know about smart cities." In: *IEEE Consumer Electronics Magazine* 5.3 (2016), pp. 60–70. issn: 21622256. doi: 10.1109/MCE.2016.2556879.

[139] A. Montanari. "Human mobility, global change and local development." In: *BELGEO* 1-2 (2005), pp. 7–18. issn: 13772368. doi: 10.4000/belgeo.12391. url: http://journals.openedition.org/belgeo/12391.

[140] K. Mwita. "Factors to consider when choosing data collection methods." In: *International Journal of Research in Business and Social Science (2147- 4478)* 11.5 (2022), pp. 532–538. issn: 2147-4478. doi: 10.20525/ijrbs.v11i5.1842. url: https://www.ssbfnet.com/ojs/index.php/ijrbs/article/view/1842.

[141] S. Myeong, J. Park, and M. Lee. "Research Models and Methodologies on the Smart City: A Systematic Literature Review." In: *Sustainability (Switzerland)* 14.3 (2022), p. 1687. issn: 20711050. doi: 10.3390/su14031687. url: https://www.mdpi.com/2071-1050/14/3/1687/htm.

[142] S. Nakano and A. Washizu. "Will smart cities enhance the social capital of residents? The importance of smart neighborhood management." In: *Cities* 115 (2021), p. 103244. issn: 02642751. doi: 10.1016/j.cities.2021.103244.

[143] U. Nations. *Urbanization trends | Knowledge for policy*. 2014. url: https://knowledge4policy.ec.europa.eu/node/9722_pt (visited on 06/26/2022).

[144] Q. Q. Ngo, F. L. Dennig, D. A. Keim, and M. Sedlmair. "Machine learning meets visualization-Experiences and lessons learned." In: *IT - Information Technology* 64.4-5 (2022), pp. 169–180. issn: 21967032. doi: 10.1515/itit-2022-0034. url: https://www.researchgate.net/publication/363208052_Machine_learning_meets_visualization_-_Experiences_and_lessons_learned.

[145] T. N. Nguyen and S. Zeadally. "Mobile Crowd-sensing Applications: Data Redundancies, Challenges, and Solutions." In: *ACM Transactions on Internet Technology* 22.2 (2022). issn: 15576051. doi: 10.1145/3431502. url: https://dl.acm.org/doi/abs/10.1145/3431502.

[146] P. Nitsche, P. Widhalm, S. Breuss, and P. Maurer. "A Strategy on How to Utilize Smartphones for Automatically Reconstructing Trips in Travel Surveys." In: *Procedia - Social and Behavioral Sciences* 48 (2012), pp. 1033–1046. issn: 18770428. doi: 10.1016/j.sbspro.2012.06.1080. url: https://linkinghub.elsevier.com/retrieve/pii/S1877042812028169.

[147] Nominet. *List of Smart City Projects*. 2016. url: https://www.nominet.uk/list-smart-city-projects/ (visited on 06/28/2019).

[148] Nominet. *Smart city projects showcase: A global snapshot of great uses of IoT in urban contexts*. Tech. rep. 2016: Nominet, 2016, p. 1. url: https://nominet-prod.s3.amazonaws.com/wp-content/uploads/2016/09/nominet_smartcities_analysis.pdf.

[149] NYC Open Data. *2014 Yellow Taxi Trip Data*. 2014. url: https://data.cityofnewyork.us/Transportation/2018-Yellow-Taxi-Trip-Data/t29m-gskq (visited on 01/13/2023).

[150] NYC Open Data. *LinkNYC Kiosk Status*. 2019. url: https://data.cityofnewyork.us/City-Government/LinkNYC-Kiosk-Status/n6c5-95xh (visited on 04/20/2019).

[151] NYC Open Data. *nynta | NYC Open Data*. 2019. url: https://data.cityofnewyork.us/dataset/nynta/q2z5-ai38 (visited on 07/06/2022).

[152] P. Oemkiosks. *A tecnologia debate-se na Aveiro Tech Week, com a contribuição dos mupis digitais da Partteam e Oemkiosks*. 2020. url: https://partteams.com/?page=news&news=aveiro-tech-week-com-a-contribuicao-dos-mupis-digitais-da-partteam-oemkiosks (visited on 03/11/2023).

[153] S. Paiva. *Precision Positioning with Commercial Smartphones in Urban Environments*. 2021, p. 181. isbn: 9783030712877. doi: 10.1007/978-3-030-71288-4.

[154] A. Pando. *Smart Cities: Can Data Account For Population Growth?* 2018. url: https://www.forbes.com/sites/forbestechcouncil/2018/05/10/smart-cities-can-data-account-for-population-growth/#4c71f8c93437 (visited on 03/13/2019).

[155] F. Peng, B. Altieri, T. Hutchinson, A. J. Harris, and D. McLean. "Design for Social Innovation: A Systemic Design Approach in Creative Higher Education toward Sustainability." In: *Sustainability (Switzerland)* 14.13 (2022), p. 8075. issn: 20711050. doi: 10.3390/su14138075. url: https://www.mdpi.com/2071-1050/14/13/8075.

[156] C. Perera, M. Barhamgi, S. De, T. Baarslag, M. Vecchio, and K.-K. R. Choo. "Designing the Sensing as a Service Ecosystem for the Internet of Things." In: *IEEE Internet of Things Magazine* 1.2 (2019), pp. 18–23. issn: 2576-3180. doi: 10.1109/iotm.2019.1800023. arXiv: 1904.05023. url: https://arxiv.org/abs/1904.05023v1.

[157] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos. "Sensing as a service model for smart cities supported by Internet of Things." In: *Transactions on Emerging Telecommunications Technologies* 25.1 (2014), pp. 81–93. issn: 21613915. doi: 10.1002/ett.2704. arXiv: arXiv:1307.8198v1.

[158] Pooja G, Sundar R, Harshini R, Arjuna S, and Ram Kumar C. "Recent Trends and Challenges in Smart Cities." In: *EAI Endorsed Transactions on Smart Cities* 6.3 (2022), e4. doi: 10.4108/eetsc.v6i3.2273.

[159] F. Portela. *Data science and knowledge discovery*. 2021. doi: 10.3390/fi13070178. url: https://doi.org/10.3390/fi13070178.

[160] G. R. Portuguesa. *Portal de dados abertos da Administração Pública*. 2020. url: https://dados.gov.pt/pt/ (visited on 01/14/2023).

[161] PTI. *Digital universe to touch 44 trillion gigabytes by 2020: Report*. 2014. url: https://economictimes.indiatimes.com/tech/internet/digital-universe-to-touch-44-trillion-gigabytes-by-2020-report/articleshow/33747368.cms (visited on 03/13/2019).

[162] A. Quaglieri, J. Bloom, A. Triantafyllidis, B. Green, M. R. Condina, P. B. Ngov, G. Infusini, and A. I. Webb. "Mass Dynamics 2.0: An improved modular web-based platform for accelerated proteomics insight generation and decision making." In: *bioRxiv* (2022), p. 2022.12.12.517480. doi: 10.1101/2022.12.12.517480.

[163] A. P. F. de Queiroz, D. S. Guimarães Júnior, A. M. Nascimento, and F. J. C. de Melo. "Overview of Urban Mobility in Smart Cities." In: *Research, Society and Development* 10.9 (2021), e18210917830. issn: 2525-3409. doi: 10.33448/rsd-v10i9.17830. url: https://rsdjournal.org/index.php/rsd/article/view/17830.

[164] A. Ray, · Chandreyee Chowdhury, S. Bhattacharya, and S. Roy. "A survey of mobile crowdsensing and crowdsourcing strategies for smart mobile device users." In: *CCF Transactions on Pervasive Computing and Interaction 2022* (2022), pp. 1–26. issn: 2524-5228. doi: 10.1007/S42486-022-00110-9.

[165] D. J. Reck, D. A. Hensher, and C. Q. Ho. “MaaS bundle design.” In: *Transportation Research Part A: Policy and Practice* 141 (2020), pp. 485–501. issn: 09658564. doi: 10.1016/j.tra.2020.09.021.

[166] C. Reiff, S. Oechsle, F. Eger, and A. Verl. “Web-based Platform for Data Analysis and Monitoring.” In: *Procedia CIRP* 86 (2019), pp. 31–36. doi: 10.1016/J.PROCIR.2020.01.009.

[167] Reliable Prognosis-5. *Weather for 243 countries*. 2021. url: https://rp5.ru/http://rp5.am/Weather_in_the_world (visited on 01/14/2023).

[168] M. T. Ribeiro, S. Singh, and C. Guestrin. “”why should i trust you?”explaining the predictions of any classifier.” In: *NAACL-HLT 2016 - 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Demonstrations Session*. Association for Computational Linguistics (ACL), 2016, pp. 97–101. isbn: 9781450342322. doi: 10.18653/v1/n16-3020. arXiv: 1602.04938.

[169] L. Rosa, H. Faria, R. Tabrizi, S. Gonçalves, F. Silva, and C. Analide. “Sentiment Analysis Based on Smart Human Mobility: A Comparative Study of ML Models.” In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 13259 LNCS. Springer Science and Business Media Deutschland GmbH, 2022, pp. 55–64. isbn: 9783031065262. doi: 10.1007/978-3-031-06527-9_6.

[170] L. Rosa, M. Guimarães, D. Carneiro, F. Silva, and C. Analide. “Explainable Decision Tree on Smart Human Mobility.” In: IOS Press, 2022, pp. 325–334. doi: 10.3233/aise220059. url: https://ebooks.iospress.nl/doi/10.3233/AISE220059.

[171] L. Rosa, F. Silva, and C. Analide. “TMSA: Participatory Sensing Based on Mobile Phones in Urban Spaces.” In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 12489 LNCS. Springer Science and Business Media Deutschland GmbH, 2020, pp. 257–267. isbn: 9783030623616. doi: 10.1007/978-3-030-62362-3_23. url: https://link.springer.com/chapter/10.1007/978-3-030-62362-3_23.

[172] L. Rosa, F. Silva, and C. Analide. “Mobile networks and internet of things infrastructures to characterize smart human mobility.” In: *Smart Cities* 4.2 (2021), pp. 894–918. issn: 26246511. doi: 10.3390/smartcities4020046. url: https://www.mdpi.com/2624-6511/4/2/46.

[173] L. Rosa, F. Silva, and C. Analide. “Urban Human Mobility Modelling and Prediction: Impact of Comfort and Well-Being Indicators.” In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 12981 LNAI. Springer Science and Business Media Deutschland GmbH, 2021, pp. 278–289. isbn: 9783030862299. doi: 10.1007/978-3-030-86230-5_22. url: https://www.researchgate.net/

publication/354444595_Urban_Human_Mobility_Modelling_and_Prediction_Impact_of_Comfort_and_Well-Being_Indicators.

[174] L. Rosa, F. Silva, and C. Analide. "Analyzing Metrics to Understand Human Mobility Phenomena: Challenges and Solutions." In: *Lecture Notes in Networks and Systems*. Vol. 332. Springer Science and Business Media Deutschland GmbH, 2022, pp. 161–170. isbn: 9783030868864. doi: 10.1007/978-3-030-86887-1_15. url: https://link.springer.com/chapter/10.1007/978-3-030-86887-1_15.

[175] L. Rosa, F. Silva, C. Analide, L. Silva, and C. Analide. "Representing Human Mobility Patterns in Urban Spaces." In: *Intelligent Environments 2020* 28 (2020). Ed. by C. A. Iglesias, J. I. Moreno Novella, A. Ricci, D. Rivera Pinto, and D. Roman, pp. 177–186. doi: 10.3233/AISE200067. url: https://hdl.handle.net/1822/83659.

[176] A. Rosenfeld. "Better metrics for evaluating explainable artificial intelligence." In: *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*. Vol. 1. 2021, pp. 45–50. isbn: 9781713832621. url: www.ifaamas.org.

[177] K. C. Roy, M. Cebrian, and S. Hasan. "Quantifying human mobility resilience to extreme events using geo-located social media data." In: *EPJ Data Science* 8.1 (2019). issn: 21931127. doi: 10.1140/epjds/s13688-019-0196-6. url: https://doi.org/10.1140/epjds/s13688-019-0196-6.

[178] P. Saariluoma. "Human Technology." In: *Human Technology* 5.2 (2009). url: https://ht.csr-pub.eu/index.php/ht.

[179] G. Sagl, B. Resch, and T. Blaschke. "Contextual sensing: Integrating contextual information with human and technical geo-sensor information for smart cities." In: *Sensors (Switzerland)* 15.7 (2015), pp. 17013–17035. issn: 14248220. doi: 10.3390/s150717013. url: https://www.researchgate.net/figure/Model-of-smart-city-interactions-between-humans-the-environment-and-technology-The_fig1_280316988.

[180] M. K. Sahu and S. Patil. "Linear Regression based Power Optimization of Wireless Sensor Network in Smart City." In: *IOP Conference Series: Materials Science and Engineering* 1085.1 (2021), p. 012005. issn: 1757-8981. doi: 10.1088/1757-899x/1085/1/012005. url: https://iopscience.iop.org/article/10.1088/1757-899X/1085/1/012005.

[181] K. Sallah, R. Giorgi, L. Bengtsson, X. Lu, E. Wetter, P. Adrien, S. Rebaudet, R. Piarroux, and J. Gaudart. "Mathematical models for predicting human mobility in the context of infectious disease spread: Introducing the impedance model." In: *International Journal of Health Geographics* 16.1 (2017), pp. 1–11. issn: 1476072X. doi: 10.1186/s12942-017-0115-7. url: https://ij-healthgeographics.biomedcentral.com/articles/10.1186/s12942-017-0115-7.

[182] C. Santana, F. Botta, H. Barbosa, F. Privitera, R. Menezes, and R. Di Clemente. “Changes in the time-space dimension of human mobility during the COVID-19 pandemic.” In: (2022). arXiv: 2201.06527. url: http://arxiv.org/abs/2201.06527.

[183] R. Schapire, D. A. Kusumajati, R. P. Chairiyani, and N. T. Martoredjo. “The Influence of Social Media on Human Behavior in Adolescents (Case Study of Bina Nusantara University Students).” In: *ACM International Conference Proceeding Series*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–6. isbn: 9781450376839. doi: 10.1145/3378904.3378917. url: https://www.cs.princeton.edu/courses/archive/spr08/cos511/scribe_notes/0204.pdf.

[184] M. Schläpfer, L. M. Bettencourt, S. Grauwin, M. Raschke, R. Claxton, Z. Smoreda, G. B. West, and C. Ratti. “The scaling of human interactions with city size.” In: *Journal of the Royal Society Interface* 11.98 (2014). issn: 17425662. doi: 10.1098/rsif.2013.0789. arXiv: 1210.5215. url: https://royalsocietypublishing.org/doi/10.1098/rsif.2013.0789.

[185] Scikit-learn. *1.10. Decision Trees — scikit-learn 1.0.2 documentation*. 2021. url: https://scikit-learn.org/stable/modules/tree.html (visited on 03/10/2022).

[186] Sergios Karagiannakos. *Introduction to Kubernetes with Google Cloud: Deploy your Deep Learning model effortlessly*. 2020. url: https://theaisummer.com/kubernetes/ (visited on 12/28/2022).

[187] L. Shi, L. Wang, C. Long, S. Zhou, F. Zheng, N. Zheng, and G. Hua. “Social Interpretable Tree for Pedestrian Trajectory Prediction.” In: *Proceedings of the AAAI Conference on Artificial Intelligence* 36.2 (2022), pp. 2235–2243. issn: 2159-5399. doi: 10.1609/aaai.v36i2.20121. arXiv: 2205.13296. url: https://ojs.aaai.org/index.php/AAAI/article/view/20121.

[188] P. Siirtola and J. Röning. *Comparison of regression and classification models for user-independent and personal stress detection*. 2020. doi: 10.3390/s20164402. url: https://www.mdpi.com/1424-8220/20/16/4402.

[189] A. Silva, M. Schrum, E. Hedlund-Botti, N. Gopalan, and M. Gombolay. “Explainable Artificial Intelligence: Evaluating the Objective and Subjective Impacts of xAI on Human-Agent Interaction.” In: *International Journal of Human-Computer Interaction* (2022). issn: 15327590. doi: 10.1080/10447318.2022.2101698. url: https://www.tandfonline.com/doi/abs/10.1080/10447318.2022.2101698.

[190] F. Simini, G. Barlacchi, M. Luca, and L. Pappalardo. “A Deep Gravity model for mobility flows generation.” In: *Nature Communications* 12.1 (2021), pp. 1–13. issn: 20411723. doi: 10.1038/s41467-021-26752-4. url: https://www.nature.com/articles/s41467-021-26752-4.

[191] M. Singh, J. S. Bhatia, and D. Malhotra. "Big Data: The Future of Data Storage." In: *IOSR Journal of Computer Engineering* 16.5 (2014), pp. 130–136. issn: 22788727. doi: 10.9790/0661-1654130136. url: www.iosrjournals.org.

[192] H. Sinky, B. Khalfi, B. Hamdaoui, and A. Rayes. "Responsive Content-Centric Delivery in Large Urban Communication Networks: A LinkNYC Use-Case." In: *IEEE Transactions on Wireless Communications* 17.3 (2018), pp. 1688–1699. issn: 15361276. doi: 10.1109/TWC.2017.2784433. arXiv: 1805.12268v1. url: https://azure.microsoft.com/en-us/documentation/articles/cdn-pop-locations.

[193] SiteWhere. *System Architecture | SiteWhere CE 2.1.0 Documentation*. 2022. url: https://sitewhere.io/docs/2.1.0/platform/architecture.html#kubernetes-orchestration (visited on 10/31/2022).

[194] D. Slah. *Data Mining Tools*. 2017. url: https://towardsdatascience.com/data-mining-tools-f701645e0f4c (visited on 04/10/2019).

[195] G. Smith, J. Sochor, and I. C. A. Karlsson. "Adopting Mobility-as-a-Service: An empirical analysis of end-users' experiences." In: *Travel Behaviour and Society* 28 (2022), pp. 237–248. issn: 2214367X. doi: 10.1016/j.tbs.2022.04.001.

[196] K. Smolak, K. Siła-Nowicka, J. C. Delvenne, M. Wierzbinski, and W. Rohm. "The impact of human mobility data scales and processing on movement predictability." In: *Scientific Reports* 11.1 (2021), pp. 1–10. issn: 20452322. doi: 10.1038/s41598-021-94102-x. url: https://www.nature.com/articles/s41598-021-94102-x.

[197] C. Stephanidis, G. Salvendy, M. Antona, J. Y. Chen, J. Dong, V. G. Duffy, X. Fang, C. Fidopiastis, G. Fragomeni, L. P. Fu, Y. Guo, D. Harris, A. Ioannou, K. a. K. Jeong, S. Konomi, H. Krömker, M. Kurosu, J. R. Lewis, A. Marcus, G. Meiselwitz, A. Moallem, H. Mori, F. Fui-Hoon Nah, S. Ntoa, P. L. P. Rau, D. Schmorrow, K. Siau, N. Streitz, W. Wang, S. Yamamoto, P. Zaphiris, and J. Zhou. *Seven HCI Grand Challenges*. 2019. doi: 10.1080/10447318.2019.1619259.

[198] Sustainable Development. *IEC-ISO-ITU Joint Smart Cities Task Force (J-SCTF) Forum on "Strengthening IEC, ISO and ITU collaboration for Smart Cities"*. 2021. url: https://www.itu.int/en/ITU-T/Workshops-and-Seminars/2021/0621/Pages/default.aspx (visited on 08/30/2022).

[199] Sydney Stone. *Key Challenges of Smart Cities and How to Overcome Them*. 2018. url: https://ubidots.com/blog/the-key-challenges-for-smart-cities/ (visited on 04/21/2023).

[200] T. Tamagusko and A. Ferreira. "Data-driven approach to understand the mobility patterns of the portuguese population during the covid-19 pandemic." In: *Sustainability (Switzerland)* 12.22 (2020), pp. 1–12. issn: 20711050. doi: `10.3390/su12229775`. url: `www.mdpi.com/journal/sustainability`.

[201] M. E. Taylor. "Intelligibility is a key component to trust in machine learning." In: *Borealis AI* (2019). url: `https://www.borealisai.com/en/blog/intelligibility-key-component-trust-machine-learning/`.

[202] D. d. C. Teixeira, J. M. Almeida, A. C. Viana, D. Do Couto Teixeira, J. M. Almeida, and A. C. Viana. "On estimating the predictability of human mobility: the role of routine." In: *EPJ Data Science* 10.1 (2021), pp. 1–30. issn: 21931127. doi: `10.1140/epjds/s13688-021-00304-8`. url: `https://doi.org/10.1140/epjds/s13688-021-00304-8`.

[203] T. (TLC) and L. Commission. *2015 Green Taxi Trip Data*. 2017. url: `https://data.cityofnewyork.us/Transportation/2016-Green-Taxi-Trip-Data/hvrh-b6nb` (visited on 01/13/2023).

[204] UNFPA. "The state of the world population 2007: unleashing the potential of urban growth." In: *Indian pediatrics* (2008). issn: 0019-6061. doi: `ISBN978-0-89714-807-8`.

[205] M. Uras, R. Cossu, E. Ferrara, A. Liotta, and L. Atzori. "PmA: A real-world system for people mobility monitoring and analysis based on Wi-Fi probes." In: *Journal of Cleaner Production* 270 (2020), p. 122084. issn: 09596526. doi: `10.1016/j.jclepro.2020.122084`.

[206] S. A. Veenstra, T. Thomas, and S. I. Tutert. "Trip distribution for limited destinations: A case study for grocery shopping trips in the Netherlands." In: *Transportation* 37.4 (2010), pp. 663–676. issn: 00494488. doi: `10.1007/s11116-010-9274-0`. url: `https://link.springer.com/article/10.1007/s11116-010-9274-0`.

[207] R. Venkateswaran, T. Divagar, T. Luo, H. P. Tan, R. Venkat, T. Divagar, T. Luo, and H. P. Tan. "Participatory Sensing for Government-Centric Applications: A Singapore Case Study." In: (2015). arXiv: `1508.05021`. url: `http://arxiv.org/abs/1508.05021`.

[208] A. Verma. *Internet of Things and Big Data - Better Together*. 2018. url: `https://www.whizlabs.com/blog/iot-and-big-data/` (visited on 03/13/2019).

[209] T. M. Vinod Kumar and B. Dahiya. "Smart Economy in Smart Cities." In: 2016, pp. 3–76. isbn: 978-981-10-1608-0. doi: `10.1007/978-981-10-1610-3_1`.

[210] J. S. M. Visit. *Back-End Developer Roadmap*. 2022. url: `https://www.geeksforgeeks.org/back-end-developer-roadmap-2022/` (visited on 09/11/2023).

[211] A. Wang, A. Zhang, E. H. W. Chan, W. Shi, X. Zhou, and Z. Liu. "A review of human mobility research based on big data and its implication for smart city development." In: *ISPRS International Journal of Geo-Information* 10.1 (2021), p. 13. issn: 22209964. doi: 10.3390/ijgi10010013. url: https://www.mdpi.com/2220-9964/10/1/13.

[212] Y. Wang, A. Yalcin, and C. VandeWeerd. "An entropy-based approach to the study of human mobility and behavior in private homes." In: *PLoS ONE* 15.12 December (2020), e0243503. issn: 19326203. doi: 10.1371/journal.pone.0243503. url: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0243503.

[213] K. S. Willis and C. Nold. "Sense and the city: An Emotion Data Framework for smart city governance." In: *Journal of Urban Management* 11.2 (2022), pp. 142–152. issn: 25890360. doi: 10.1016/j.jum.2022.05.009.

[214] R. Wolff. *10 Best Data Mining Tools in 2022*. 2020. url: https://medium.com/datatobiz/25-best-data-mining-tools-in-2022-65e77c905a2b (visited on 04/25/2023).

[215] Y. Z. Wong and D. A. Hensher. "Delivering mobility as a service (MaaS) through a broker/aggregator business model." In: *Transportation* 48.4 (2021), pp. 1837–1863. issn: 15729435. doi: 10.1007/s11116-020-10113-z. url: https://link.springer.com/article/10.1007/s11116-020-10113-z.

[216] Y. Z. Wong, D. A. Hensher, and C. Mulley. "Mobility as a service (MaaS): Charting a future context." In: *Transportation Research Part A: Policy and Practice* 131 (2020), pp. 5–19. issn: 09658564. doi: 10.1016/j.tra.2019.09.030.

[217] World Economic Forum. *Our Alliance is creating smart city governance*. 2022. url: https://www.weforum.org/impact/smart-cities-governance-alliance/ (visited on 09/13/2022).

[218] M. Wu and Q. Huang. "Human movement patterns of different racial-ethnic and economic groups in U.S. top 50 populated cities: What can social media tell us about isolation?" In: *Annals of GIS* 28.2 (2022), pp. 161–183. issn: 19475691. doi: 10.1080/19475683.2022.2026471. url: https://www.tandfonline.com/doi/abs/10.1080/19475683.2022.2026471.

[219] F. Xia, J. Wang, X. Kong, Z. Wang, J. Li, and C. Liu. "Exploring Human Mobility Patterns in Urban Scenarios: A Trajectory Data Perspective." In: *IEEE Communications Magazine* 56.3 (2018), pp. 142–149. issn: 01636804. doi: 10.1109/MCOM.2018.1700242.

[220] H. Xiong, K. Xie, L. Ma, F. Yuan, and R. Shen. "Exploring the Citywide Human Mobility Patterns of Taxi Trips through a Topic-Modeling Analysis." In: *Journal of Advanced Transportation* 2021 (2021). issn: 20423195. doi: 10.1155/2021/6697827.

[221] J. Yang, Y. Sun, B. Shang, L. Wang, and J. Zhu. "Understanding collective human mobility spatiotemporal patterns on weekdays from taxi origin-destination point data." In: *Sensors (Switzerland)* 19.12 (2019), p. 2812. issn: 14248220. doi: 10.3390/s19122812. url: https://www.mdpi.com/1424-8220/19/12/2812/htm.

[222] S. C.-H. Yang, T. Folke, and P. Shafto. "A Psychological Theory of Explainability." In: (2022), p. 14. doi: 10.48550/arxiv.2205.08452. arXiv: 2205.08452.

[223] X. Yang, Z. Pu, and L. Zhang. "Designing Data Visualization Using Artificial Intelligence for Urban Intelligent Transportation Scenarios." In: *Mobile Information Systems* 2022 (2022). issn: 1875905X. doi: 10.1155/2022/3967267.

[224] M. Yeom. *Gentrification: An Introduction, Overview, and Application*. 2019. url: https://www.researchgate.net/publication/337200128_Gentrification_An_Introduction_Overview_and_Application (visited on 05/28/2022).

[225] S. Zhang, F. Zhen, Y. Kong, T. Lobsang, and S. Zou. "Towards a 15-minute city: A network-based evaluation framework." In: *Environment and Planning B: Urban Analytics and City Science* 50.2 (2022), p. 239980832211185. issn: 2399-8083. doi: 10.1177/23998083221118570.

[226] Z. Zhao, H. N. Koutsopoulos, and J. Zhao. "Identifying hidden visits from sparse call detail record data." In: *Transactions in Urban Data, Science, and Technology* 1.3-4 (2022), pp. 121–141. issn: 2754-1231. doi: 10.1177/27541231221124164. arXiv: 2106.12885. url: https://www.researchgate.net/publication/353069019_Identifying_Hidden_Visits_from_Sparse_Call_Detail_Record_Data.

[227] Z. Zhao, S. L. Shaw, L. Yin, Z. Fang, X. Yang, F. Zhang, and S. Wu. "The effect of temporal sampling intervals on typical human mobility indicators obtained from mobile phone location data." In: *International Journal of Geographical Information Science* 33.7 (2019), pp. 1471–1495. issn: 13623087. doi: 10.1080/13658816.2019.1584805. url: https://www.researchgate.net/publication/332424900_The_effect_of_temporal_sampling_intervals_on_typical_human_mobility_indicators_obtained_from_mobile_phone_location_data.