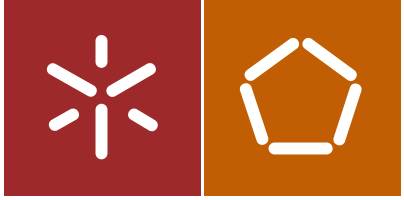
Universidade do Minho
Escola de Engenharia

Nasim Sadat Mosavi

Intelligent Decision Support System for Precision Medicine

April 2024

Nasim Sadat Mosavi Intelligent Decision Support System for Precision Medicine
UMinho | 2024



Universidade do Minho
Escola de Engenharia

Nasim Sadat Mosavi

Intelligent Decision Support System for Precision Medicine

Doctoral Program in Information Systems and Technology

Thesis performed under the supervision of
Professor Doutor Manuel Filipe Santos

April 2024

COPYRIGHT

ACKNOWLEDGMENTS

I would like to begin by expressing my deepest gratitude to Professor Manuel Filipe Santos, my esteemed PhD supervisor, for his unwavering guidance, expertise, and mentorship throughout my doctoral journey. His insightful feedback, constructive criticism, and steadfast support have played a pivotal role in shaping my research and academic development. I am profoundly grateful for the opportunity to learn from him, and his mentorship has been instrumental in my academic growth.

To my beloved parents, whose constant support has been the cornerstone of my journey, I owe an immeasurable debt of gratitude. Their endless encouragement, sage advice, and strong belief in my abilities have been a constant source of strength and inspiration. Their sacrifices and dedication to my success have shaped me into the person I am today, and for that, I am eternally thankful.

I am also deeply appreciative of my son Kourosh and my daughter Elena for their deep understanding and steadfast support. Their presence in my life has brought immeasurable joy and motivation, and their belief in my capabilities has propelled me forward during challenging times. Their unwavering faith in me has been a guiding light, driving my determination and resilience.

Special recognition is extended to my husband, Roozbeh, whose unconditional love, support, and understanding have been my rock throughout this journey. His persistent encouragement and belief in my potential have provided me with the strength and confidence to overcome obstacles and achieve success. His steadfast support has created a nurturing environment that allowed me to focus wholeheartedly on my research and academic pursuits.

Lastly, I extend my heartfelt gratitude to all my family members for their encouragement, and belief in my potential. Their continuous support has been the wind beneath my wings, propelling me toward my academic goals. I am truly fortunate to have such incredible motivations on my side, and I am deeply grateful for their presence in my life.

DECLARATION 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 University of Minho.

RESUMO

Sistema de Suporte à Decisão Inteligente para Medicina de Precisão

Sistema de Suporte à Decisão Inteligente para Medicina de Precisão

Este trabalho de doutoramento corresponde a uma exploração abrangente da integração de Sistemas de Suporte à Decisão Inteligente (SSDI) no contexto da medicina de precisão. O capítulo introdutório estabelece o cenário oferecendo uma visão detalhada da investigação, delineando motivações, objetivos e a importância do trabalho. Motivada pela necessidade imperativa de ir além da abordagem tradicional baseada em sintomas na tomada de decisão clínica, a pesquisa procura abordar questões fundamentais relacionadas à adaptação de estratégias de tratamento das características individuais do paciente através da abordagem baseada em dados.

Uma análise crítica aos protocolos de tomada de decisão clínica revela limitações significativas na adequação dos tratamentos às circunstâncias individuais dos pacientes. O paradigma da medicina de precisão destaca a necessidade de uma introspeção baseada nos dados por forma a aprimorar os processos decisórios. Apesar dos avanços nos Sistemas de Suporte à Decisão Clínica (SSDC), desafios como a qualidade dos dados e interoperabilidade persistem, exigindo uma abordagem transformadora. Com base em dois projetos distintos, o estudo desenvolve-se em domínios interdisciplinares, preenchendo lacunas numa perspetiva baseada em dados, para aperfeiçoar os SSDI no âmbito da Medicina de Precisão (IDSS4PM). Ao explorar conjuntos de dados oriundos da psicoterapia e da medicina intensiva, a investigação centra-se no processamento de dados, capacidade analítica e na modelação preditiva, estabelecendo as bases para aprimorar os protocolos de tomada de decisão clínica. O enquadramento teórico foi inspirado no modelo de tomada de decisão de Simon para orientação do desenho do IDSS4PM. Nos capítulos subsequentes encontra-se a definição da metodologia de pesquisa e um conjunto de capítulos onde são apresentados os resultados experimentais que contribuem para o aperfeiçoamento dos SSDI. O estudo visa fazer uma contribuição significativa para a medicina de precisão, tirando partido de perceções baseadas em dados. Nosso objetivo é promover uma abordagem precisa e eficaz à saúde e orientar mudanças transformadoras nas práticas de tomada de decisão clínica. Na conclusão, é feito um conjunto recomendações para futuros trabalhos de pesquisa, reconhecendo as limitações do trabalho atual.

Palavras-chave: Inteligência Artificial, Medicina de Precisão, Sistemas de Suporte à Decisão Inteligente Baseados em Dados, Tomada de Decisão Clínica,.

ABSTRACT

Intelligent Decision Support System for Precision Medicine

This doctoral thesis embarks on a comprehensive exploration into the integration of Intelligent Decision Support Systems (IDSS) within the evolving landscape of Precision Medicine (PM). The introductory chapter sets the stage by offering a detailed overview of the research, outlining motivations, objectives, and the significance of the study. Motivated by the imperative need to move beyond the traditional symptom-driven approach in clinical decision-making, the research seeks to address fundamental questions surrounding the tailoring of treatment strategies to individual patient characteristics via a data-driven approach.

A critical analysis of existing clinical decision-making protocols reveals significant limitations, particularly in the realm of tailoring treatments to individual patient's unique circumstances. The emergence of Precision Medicine as a paradigm shift underscores the necessity for data-driven insights to inform decision-making processes effectively. Despite the advancements in Clinical Decision Support Systems (CDSS), challenges such as data quality, interoperability, and fragmented data generation persist, necessitating a transformative approach.

Through two distinct projects, the study navigates interdisciplinary domains, bridging gaps in data-driven perspectives to advance the framework of IDSS for Precision Medicine (IDSS4PM). By exploring psychotherapy and Intensive Care Unit (ICU) datasets, the research delves into data processing, analytical insights, and predictive modeling, laying the groundwork for enhancing clinical decision-making protocols. The theoretical framework draws inspiration from Simon's model of decision-making, guiding the proposal of IDSS4PM. Chapters unfold to elucidate the research methodology, practical phases, and contributions toward refining the IDSS framework.

The study aimed to make a meaningful contribution to PM by leveraging data-driven insights. Our goal is to foster a precise and effective approach to healthcare and guide transformative shifts in clinical decision-making practices.

In conclusion, the thesis offers exhaustive recommendations for future research endeavors, acknowledging limitations.

Keywords: Artificial Intelligence, Clinical Decision Making, Data Driven, Intelligent Decision Support Systems, Precision Medicine.

INDEX

LIST OF ABBREVIATIONS/ACRONYMS

ABI	Adaptive BusinessIntelligence
AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC	Area Under Curve
BI	Business Intelligence
BA	Business Analytics
CC	Cognitive Computing
CDSS	Clinical Decision Support System
CRISPDM	Cross-Industry Standard Process for Data Mining
CV	Cross Validation
CEid	Clinical Event identity
DE	Distinct Episod number
DP	Distinct Process number
DSR	Design Science Research
DSS	Decision Support System
DD_ICUM_DSS	Data-Driven ICU Medicine Decision Support
DSS_PT_DSS	Data-Driven Psychotherapy Decision Support
DM	Data Mining
DT	Decision Tree
EDA	Electrodermal Activity
EIS	Executive Information System
EHR	Electronic Health Record
EMR	Electronic Medical Record
ETL	Extract Load tRANSform
GDSS	Group Decision Support System
HR	Heart Rate
ICU	Intensive Care Unit
IOT	Internet of Things
ID	Identity
IDSS	Intelligent Decision Support System

IDSS4PM	Intelligent Decision Support System for Precision Medicine
IDSSK	Intelligent Decision Support System based on Knowledge Discovery
3IDSS	Intelligent, Interactive, and Integrated Decision Support System
KNN	K-Nearest Neighbours
KDD	Knowledge Discovery from Database
ML	Machine Learning
MSE	Mean Square Error
MPI	Master Patient Index
NLP	Natural Language Processing
NRC	National Research Council
ODSS	Organization Decision Support System
OLAP	Online Analytical Processing
PDSS	Personal Decision Support System
PM	Precision Medicine
R	Rows
RF	Random Forest
ROC	Receiver Operating Characteristic
R^2	R Squared Coefficient
SOAP	Subject Object Plan
SNOMED	systematized Nomenclature of Medicine
SVR	Support Vector Regressor
TA	Therapeutic Alliance
TDA	Topological Data Analysis
V	Variables
WCSS	Within-Cluster Sum of Squares
WAI	Working Alliance Inventory

List of Figures

LIST OF EQUATIONS

LIST OF TABLES

1. INTRODUCTION

The introduction chapter serves as the gateway to this PhD thesis, offering a comprehensive overview of our research. Here, we provide the backdrop and context of the study, elucidate our motivations, and outline our objectives. Additionally, we define the scope of the study, establish the theoretical foundation, and emphasize the significance of our contributions. Finally, we outline the structure of the document, guiding readers through the subsequent chapters. Our goal with this introduction is to provide readers with a clear understanding of the journey ahead and the importance of our research within its domain.

1.3 Motivation

The concept of tailoring treatment to individual patients is not novel. Dating back to 1946, when Austin Bradford Hill successfully demonstrated the efficacy of streptomycin in treating tuberculosis, significant methodological advancements have shaped the landscape of clinical trials (Kosorok & Laber, 2019b). However, despite these advancements, the integration of data, physician expertise, and experience, coupled with clinical judgment, often led to a medical decision-making process reliant on trial and error. This symptom-driven model, founded on empirical evidence, tends to generalize solutions and treat all patients with similar symptoms uniformly, (Banappagoudar et al., 2023; S. Liu et al., 2010a). For decades, this evidence-based medicine has been the primary approach for all patients regardless of the individual patient's variabilities and disease conditions; this approach views patients with a single disease and multiple conditions from one lens only and has resulted in over-treatments, costly, inefficient, and complex medication protocols (Beckmann & Lew, 2016), (Pencina & Peterson, 2016). In addition, delays in disease diagnosis and uncertainties regarding treatment response significantly impact disease progression and treatment outcomes (Kosorok & Laber, 2019a), (Naithani et al., 2021), (Gopal et al., 2019). Moreover, Recent literature underscores the prevalence of medication errors and delayed treatments. Despite these figures, the analysis of fatalities resulting from incorrect medication remains inadequate (Alqenae et al., 2020). Estimates indicate that approximately 44 million individuals worldwide are currently living with dementia. As the global population continues to age, projections suggest that this number will more than triple by the year 2050. With this demographic shift, the economic burden of dementia is also expected to escalate significantly. In the United States alone, experts anticipate that the

annual cost associated with dementia may surpass an astonishing $600 billion by 2050 (Alzheimer's and Dementia, 2017).

In addition, the surge in consumerism, coupled with heightened demand for enhanced services, has propelled the healthcare industry to adapt (Shrank, 2017). The escalating number of individuals grappling with intricate, multiple health conditions further underscores the need for a healthcare system that goes beyond a one-size-fits-all approach (Kizer, 2001). In response, there has been a shift towards ensuring that people receive personalized care tailored to their unique circumstances, moving away from a rigid adherence to evidence-based practices alone. This evolution reflects a commitment to addressing the diverse and nuanced needs of individuals in the pursuit of optimal healthcare outcomes (König et al., 2017). Consequently, fundamental questions persist: Why do certain drugs work effectively for some patients but not others? What factors contribute to medication side effects in specific individuals? Why do some individuals develop cancer while others do not? Addressing these questions necessitates a paradigm shift toward tailoring treatment strategies to each patient's unique characteristics and circumstances(Mosavi & Santos, 2020),(Gopal et al., 2019), (Bonkhoff & Grefkes, 2022), (Ahmed et al., 2020).

The rise of Precision Medicine (PM) gained significant momentum following its endorsement by Barack Obama in 2015 and has been further bolstered by extensive scientific research from institutions such as the National Library of Medicine and the National Health Service (NHS). This movement has not only showcased remarkable technological advancements but has also brought to the forefront the inherent challenges in traditional clinical decision-making protocols(Borovska et al., 2018) (Bonkhoff & Grefkes, 2022) , (Haque et al., 2020), (Mosavi et al., 2023).

Despite the development of Clinical Decision Support Systems (CDSS), which have been in use since the 1980s and their advancements, CDSSs face several limitations and challenges. Issues such as data quality, data processing and management, transportability/interoperability, and integration with other hospitals or systems pose significant hurdles. Moreover, the disrupted/fragmented workflow associated with CDSSs can make them inefficient, hindering the dissemination and scaling of otherwise high-quality systems (Sutton et al., 2020), (Hee Lee & Yoon, 2021), (Bekbolatova et al., 2024),(Mosavi et al., 2023). Additionally, they often fall short in addressing requirements related to genetic/genomic and biological data, such as interpreting genetic test results and understanding their implications for family members and future generations in medical decision-making (Sadat Mosavi & Filipe Santos, 2021), (Bonkhoff & Grefkes, 2022).

In the early twenty-first century, the limitations of traditional clinical decision-making models became evident, exacerbated by the vast amount of healthcare big data available, which often overwhelmed human cognition in making timely decisions(Mesko, 2017a). This period also witnessed transformative changes in the healthcare landscape, driven by technological advancements such as Cognitive Computing (CC), Artificial Intelligence (AI), and Big Data Analytics(Gopal et al., 2019), (Sriram & Subrahmanian, 2020).

Recognizing the inherent complexity of patient heterogeneity in treatment assessment, particularly highlighted in the late twentieth century, alongside the rise of data analytics around 2010 and subsequent advancements in AI and CC from 2015 to 2020, has emphasized the necessity for a data-driven approach. This paradigm shift prioritizes precise and timely decision-making tailored to individual variables (Watson, 2017b, 2017a).

This shift aims to optimize the utilization of patient clinical data and harness the power of AI and CC to tailor precise treatment pathways (K. B. Johnson et al., 2021), (Jabbar et al., 2018). Through deep research and extensive literature studies, PM has evolved from a theoretical concept to a practical approach, to maximize treatment efficacy and patient outcomes(Tarassoli, 2019),(Leone et al., 2021).

In 2011, the concept of PM was introduced by the National Research Council (NRC), emphasizing the tailoring of medical interventions to individual patient characteristics. This approach involves categorizing patients based on their susceptibilities to diseases and responses to treatments. PM plays a pivotal role in guiding healthcare decisions, thereby enhancing the quality of care while minimizing unnecessary tests and therapies. As highlighted by Geoffrey S. Ginsburg and Kathryn A. Phillips in 2018, this method optimizes interventions, directing resources where they yield the greatest benefits and mitigating costs and potential side effects for patients (Maglaveras et al., 2016).

According to the U.S. National Library of Medicine, PM represents an emerging paradigm that considers individual differences, including genes, environment, and lifestyle, to prevent and treat specific diseases for everyone (Francis S. Collins, 2015; Mesko, 2017; Haque et al., 2020). Additionally, the U.S. National Cancer Institute defines PM as "a form of medicine that uses information about a person's genes, proteins, and environment to prevent, diagnose, and treat disease" (Haque et al., 2020). This progressive methodology seeks to replace the antiquated "one size fits all" model with a patient-centric "patient like me" approach. Its core mission is to address issues like inefficient treatments and medical errors, ultimately reducing the burden of overtreatment and hospitalization while saving more lives(Glen H. Murata, Albuquerque, 2014). Moreover, this approach represents a pivotal shift in medical science, where the core objectives include predicting the likelihood of developing a disease, achieving precise diagnoses,

and optimizing the most effective treatment for individual patients (Awwalu et al., 2015). The overarching goal is to usher in a new era of healthcare that is not only more personalized but also more effective in addressing the unique needs of each patient.

PM is commonly defined as a methodology that customizes treatments according to the distinct needs of individual patients. This customization is rooted in considerations such as their genetic composition, biomarkers, phenotypic traits, or psychosocial characteristics, which set apart one patient from another despite sharing highlighted clinical symptoms(König et al., 2017), (Jameson & Longo, 2015).

Widely embraced definitions of PM span several dimensions. Alternatively, PM is perceived as an all-encompassing, data-driven paradigm that underscores the significance of integrating diverse types of data, incorporating clinical information and other pertinent data sources. This data integration enables the categorization of patients into distinct subgroups, with the expectation that these subgroups share a common foundation of disease susceptibility and presentation. Ultimately, this data-driven approach aims to facilitate more accurate and personalized therapeutic solutions. These dual aspects of PM collectively underscore its commitment to tailoring medical care to individual patients while leveraging advanced data analysis and integration techniques for enhanced diagnosis and treatment (Grady, 2016)

While a precise definition of PM may remain elusive, it is widely acknowledged that the PM approach transcends the traditional symptom-based method in clinical decision-making, emphasizing the consideration of individual variabilities. As a result, this groundbreaking concept embraces a diverse range of individual data, including biomarkers, lifestyle factors, environmental influences, and genetic information, thereby underlining the comprehensive and personalized nature of PM (Sanchez-Pinto et al., 2023), (König et al., 2017), (Pelter & Druz, 2024).

Inspired by the evolving landscape of data-driven decision-making in PM,this study stands as a dedicated endeavor to contribute significantly to this dynamic protocol. Going beyond exploration, our focus is to intricately detail how data-driven insights can not only shape the framework but also drive the development of applications in PM. We envision this research as a catalyst, actively advancing the integration of data-driven methodologies, thereby fostering a transformative shift in clinical decision-making for the betterment of precision healthcare.

The personal connection and motivation behind choosing this topic for this PhD journey stem from a profound experience during my childhood. At that time, my health was jeopardized due to the potential risks associated with an incorrect medical prescription. Despite being young, the memories of that challenging period, which not only impacted my health but also profoundly affected my family, have served as a powerful inspiration.

This personal experience ignited a passion within me to embark on an academic journey. The motivation lies in leveraging the advancements in technology and utilizing the available tools to contribute to this area. The goal is to enhance precision in clinical decision-making, ensuring that others do not face the same risks and challenges that I encountered in my early years. This journey is a testament to my commitment to making a meaningful and positive impact on healthcare practices, drawing strength from my narrative.

2.3 Research Question And Objectives

In the ever-evolving landscape of healthcare, the integration of Intelligent Decision Support Systems (IDSS) has become increasingly pivotal. This research seeks to address the fundamental question: "How can data-driven insights shape the framework of Intelligent Decision Support Systems and advance optimal clinical decision-making?" To achieve this overarching goal, a set of interconnected objectives guided the investigation. Firstly, an exploration of existing IDSS frameworks will be conducted to discern their strengths and limitations. Subsequently, the research will delve into the realm of data-driven insights, investigating methodologies for extracting meaningful information from clinical data. Building upon this, the study will propose methods for seamlessly integrating these insights into the design and functionality of an IDSS.

Furthermore, objective includes the dissemination of the study's outcomes within academic platforms, aiming to validate both the performance and scientific merit of our work. By sharing our findings in scholarly settings, we seek to contribute to the academic discourse, foster collaboration and invite constructive feedback from the scientific community. This commitment to transparent communication ensures the robustness and credibility of our research, aligning with the rigorous standards of scholarly inquiry.

In conclusion, this research endeavors to furnish exhaustive recommendations for future works, accompanied by a thoughtful discussion on identified limitations. Through this research, we aim to contribute to the ongoing discourse surrounding the optimization of healthcare decision-support systems and their transformative potential.

Based on the points, our summarized objectives are delineated as follows:

a. To Investigate Current Practices

Objective: To analyze existing clinical decision-making protocols in precision medicine.

Rationale: Understanding the current landscape is crucial for identifying gaps and areas for improvement.

b. To Evaluate Data-Driven Technologies

Objective: Assess the effectiveness and limitations of data-driven technologies in supporting clinical decision-making.

Rationale: Evaluating the current technologies provides insights into their practical applications and potential enhancements.

c. To Incorporate with IDSS

Objective: Propose data-driven insights more effectively.

Rationale: Identifying areas for improvement in the current framework helps contribute to the evolution of clinical decision-making practices.

d. To obtain scientific validity

Objective: Attain scientific validity for our research.

To achieve this, we have strategically communicated our study's findings on academic platforms, including publication in reputable journals and presentation at conferences. By actively engaging with the scholarly community, we intended to subject our research to rigorous scrutiny, peer review, and dissemination, thus ensuring its recognition, impact, and contribution to the broader scientific discourse.

3.3 Scope of The Study

This research project aims to pioneer an innovative paradigm termed “Intelligent Decision Support for Precision Medicine." The primary objective is to propose a data-driven framework grounded in the concept of a "Decision Support System," strategically addressing identified challenges within the interdisciplinary field. This model carries the transformative potential to revolutionize the feasibility and applicability, facilitating a departure from the traditional "One-Size-Fits-All" medical decision-making model towards an avant-garde approach known as "Patient-Like-Me."

Envisioned as a transformative shift, this framework aims to enhance decision-making processes in PM. With a keen focus on tailoring the approach to each patient and considering their unique circumstances, the research aims to offer insights that guide the formulation of personalized and effective courses of action. Through these efforts, the project aspires to significantly contribute to the progress of precision healthcare.

4.3 Significance And Contributions

The contributions to IDSS4PM encompassed two distinct projects. Project I, the Data-Driven Psychotherapy Decision Support (DD-PT-DS), conducted exploratory analyses using data from the psychotherapy domain. Project II, the Data-Driven ICU Medicine Decision Support (DD-ICUM-DS), focused on enhancing the concept of an IDSS4PM by addressing data processing and integration challenges.

The initial phase of Project I involved the exploration of physiological and psychological data, offering valuable insights into the emerging field of data-driven models in clinical decision-making. Aligned with our research objectives, this exploratory phase aimed to uncover hidden patterns, relationships, and trends within the data, thereby deepening our understanding of Therapeutic Alliance (TA) and its implications for therapeutic practices.

TA is a relational factor that is considered effective and a robust predictor of therapy outcomes across different psychotherapy approaches and mental health diagnoses indicated by several meta-analyses(Del Re et al., 2021; Flückiger et al., 2020).

We have set out to determine:

How can data mining and machine learning techniques be utilized to delve into and uncover insights regarding the therapy factors influencing therapeutic alliance? This aims to facilitate precise predictions of the therapeutic alliance, leveraging diverse datasets that encompass both physiological and psychological factors from both clients and therapists.

From a data-driven perspective, our research sheds light on the integration of data mining and machine learning (ML) techniques in psychotherapy, which hold significant promise for informing clinical decision-making by considering diverse patient characteristics and treatment histories. While these technologies have been underutilized in the context of the therapeutic alliance, the study demonstrates their effectiveness in enhancing it. ML models can be trained to predict alliance levels based on intricate data patterns, incorporating biological and psychological factors, thereby improving the ability to understand and forecast outcomes. By tailoring predictions to individual clients and therapists, ML facilitates personalized interventions, ultimately enhancing the effectiveness of the therapeutic process. Additionally, ML algorithms excel at navigating the complexity and non-linear relationships inherent in such datasets. Finally, At the application level, findings highlight the essential connection between therapeutic alliance and therapy outcomes, emphasizing the importance of comprehending the multifaceted nature of interactions between clients and therapists.

Furthermore, our exploration extended to a second dataset from the Intensive Care Unit (ICU), presenting golden opportunities for analytical insights, predictive modeling, and advanced optimization in clinical decision-making.

In Project II, "Data-Driven ICU Medicine Decision Support" (DD_ICUM_DS), our commitment lies in conscientiously addressing the limitations and challenges identified in the background of IDSS4PM. We have actively bridged the identified gaps in data-driven perspectives, particularly in terms of data processing. The design and proposal of our framework allowed us a unique opportunity to contribute to the PM concept, leveraging two distinct datasets.

This study, which utilized diverse datasets aligned with the research question, aimed to address crucial aspects of data processing, particularly focusing on data integration and standardization. These efforts greatly facilitated our predictive analysis process, enabling us to identify patterns and similarities in data behavior. As a result, this study supports decision-making processes for tailoring treatments at the application level.

Both studies represent interdisciplinary efforts, making noteworthy contributions in diverse areas to achieve key significance within the domain of IDSS. Through our work, we systematically identified challenges and limitations, addressing critical aspects from a data-driven perspective, including data processing and management. The integration of predictive analysis and analytical insights has significantly advanced the task of decision-making. At the application level, our research illuminates critical aspects such as patient monitoring and health profiling, enabling the identification of at-risk patients, evidence-based treatment, early detection, cutting the cost of over-treatment, resource optimization, and increasing patient engagement. Ultimately, this leads to more precise decision-making and enhances the overall quality of clinical services.

1.5 Theoretical Framework

This study draws inspiration from Simon's model of decision-making, particularly referencing the intelligence-design-choice model. We utilize this model as a benchmark to validate the proposed framework, IDSS4PM, aiming to support a novel approach in clinical decision-making, specifically within the realm of precision medicine. Our objective is to maximize the utilization of data and advanced technologies, including AI, and analytics, to enhance the decision-making process. As a result, the findings of this study contribute to the advancement of optimal clinical decision-making practices.

1.6 Structure of Thesis

The thesis structure is as follows: Chapter two provides a comprehensive examination of two significant domains: the development of DSSs and PM. The first section offers an overview of various types of DSSs, with a particular emphasis on IDSSs. Following this, the second part delves into PM as an emerging approach in clinical decision-making, elucidating its evolution and underlying catalysts. Subsequently, the chapter focuses on IDSSs tailored for PM, with an emphasis on identifying gaps and limitations from a data perspective.

Furthermore, chapter three elucidates the methodologies and literature review strategy employed in this study. Here, we delineate our research process and its associated phases using two methodologies: CRISP-DM and Design Science Research for Information Systems (DSR).

In Chapters 4 and 5, we present an extensive examination of the practical phases involved in contributions to IDSS4PM. The exploration of defined objectives unfolds through the lens of two distinct research projects. This dual approach has provided with a unique opportunity to operate within a multidisciplinary platform, traversing the domains of both psychotherapy and Intensive Care Units (ICU).

Chapter 4 delves into the technical process of contributing to IDSS4PM using psychotherapy data. The technical exploration encompasses key aspects such as data exploration, analytical insights through interrelated clinical events, and predictive modeling. Each subsection highlights significant contributions to the research question and objectives.

In Chapter 5, we delve into the technical phases aimed at enhancing the framework of IDSS4PM using ICU data. This includes the exploration of data, followed by the creation of the Clinical Event Identity (CEid). Subsequently, the next section addresses analytical insights, and thereafter, cluster analysis is performed in two parts. The first part involves identifying similar data behaviors through cluster analysis and pinpointing cluster references and patterns. In the second part, we map cluster references onto the initial dataset and analyze the insights of each cluster. This chapter concludes with a classification technique utilized to predict the cluster and ascertain the significance of each predictor in predicting the cluster number. Furthermore, we employ metrics to evaluate the performance of the enhanced algorithm. As It was mentioned above, after each chapter (4, 5), we discuss the major contributions to IDSS4PM.

In Chapter 6, we delve into the significant contributions made towards the creation of an IDSS4PM. We thoroughly review the key performance indicators and elucidate the significance of our work in the context of the new approach to clinical decision-making. Within this section, we highlight the data-driven contributions and briefly discuss the implications of the refined results at the application level. Drawing from the Simon

decision-making model, acknowledged as a suboptimal protocol in decision-making, we explore and analyze our framework with this theoretical reference. This presentation underscores how the study holds the potential to advance precise decision-making.

Finally, in the conclusion, we provide a brief overview of the research work and highlight how we enhanced the results using two different projects to address the research question. Additionally, we discuss potential avenues for future research and major limitations encountered during the study.

2. STATE OF THE ART

In this chapter, we provide an overview of the state-of-the-art in two key aspects. Firstly, we review the development of DSSs, discussing various types of DSS and the integration of AI in these systems. Secondly, we delve into PM as a novel approach to clinical decision-making, exploring the concept and the driving forces behind this evolving paradigm. Furthermore, we extensively present the state-of-the-art IDSS4PM, highlighting both their potential and limitations from a data-driven perspective.

2.1 Development of Decision Support Systems

Scott Morton, in the 1970s, pioneered the introduction of the "Decision Support System" (DSS) concept. His conceptualization was informed by two pivotal research studies: the first delved into the theoretical aspects of organizational decision-making, conducted within a research stream at the Carnegie Institute of Technology. Simultaneously, the second study focused on the technical examination of interactive computer systems and was carried out at the Massachusetts Institute of Technology (Fentahun Moges Kasie, Glen Bright, 2017).

The term "DSS" first surfaced in 1970 when Gorry and Scott Morton introduced the article titled "A Framework for Management Information Systems." In their seminal work, Gorry and Morton (1971) categorized decisions into three levels: structured, semi-structured, and unstructured. Structured decisions involve known variables with measurable criteria, whereas unstructured decisions deal with variables that cannot be quantified. Semi-structured decisions fall between the structured and unstructured domains(Kirs et al., 2006),(Asemi et al., 2011),(Dulcic et al., 2012)(Phillips-Wren et al., 2017).

The integration of Anthony's (1965) delineation of management activities—such as strategic planning, management control, and operational control—and Simon's (1960s) classifications of decision types, programmed and non-programmed, led to an evolved version of the Decision Support System (DSS). Programmed decisions adhere to established routines, standards, and management guidelines, while non-programmed decisions lack predetermined rules or agreed-upon solutions (Asemi et al., 2011).

Simon's contributions extended further as he provided a basis for developing a predictive model in decision-making. In 1950, Herbert Simon and James March introduced a novel framework for decision-making that acknowledged human limitations, considering all alternatives in the decision-making process. Simon's "Bounded Rationality" theory emphasizes that environmental and psychological factors influence

decision-makers, leading to satisfying decisions rather than optimizing ones. Simon addressed human limitations affecting decision-making accuracy and argued for a more practical model to achieve more rational decisions (Jean-Charles Pomerol, 1997), (Kalantari, 2010),(Michael A. Eierman, Fred Niederman, 1995).

Simon also made a pivotal distinction between enterprise-wide DSS and desktop DSS. An enterprise DSS accesses a data repository, making it accessible to multiple individuals, while a desktop DSS is a standalone system (Felsberger et al., 2017).

Since the inception of Decision Support Systems (DSS), researchers have adopted various approaches to delve into the diverse types of DSS, each embodying distinct philosophies of support and potential. Notably, in the 1970s, Personal Decision Support Systems (PDSS) and Group Decision Support Systems (GDSS) gained prominence. PDSSs, being the earliest type of DSS, focused on addressing singular decision tasks for individual managers or independent users. However, as the 1980s unfolded, the focus shifted towards Group-oriented DSS (GDSS) and Organizational Decision Support Systems (ODSS), where multiple individuals collaboratively contribute to decision-making processes. Executive Information Systems (EIS), a data-oriented DSS providing organizational reports for management, Online Analytical Processing Systems (OLAP) for multidimensional queries, data warehousing (a repository of raw data for decision-making), and Business Intelligence (BI), encompassing applications for data collection, categorization, analysis, and presentation as DSS inputs, are additional categories within the DSS landscape(Arnott & Pervan, 2008), (Dulcic et al., 2012), (P. Bernus, P. Nemes, 1946).

Furthermore, in the mid-1980s, the integration of expert systems into DSS marked a significant development. Keen and Morton (1978) identified AI as a potential support to enhance decision-making. The fusion of AI and DSS has given rise to knowledge-based decision support systems, often referred to as Intelligent DSS (IDSS). The marriage of AI and decision science emerged from a vast literature of problem-solving methods, with the advent of computers in the 1940s facilitating automated reasoning (Horvitz et al., 1988).

In the 1990s, the globalized world economy prompted organizations to navigate integrated business platforms, emphasizing the generation of data to support managerial decision-making (Arnott & Pervan, 2008). As the twenty-first century unfolded, Business Intelligence (BI), communication, and knowledge management (KM) became pivotal contributors to Decision Support Systems (DSS) through web platforms, ushering in a new framework for solving organizational problems(Abubakar et al., 2019),(D. J. Power, 2000, 2002), highlighted the significance of communication driven DSS, acknowledging its role in considering alternatives beyond operational scopes, standards, and existing domains for problem-

solving(Fentahun Moges Kasie, Glen Bright, 2017), (Shim J. P. et al., 2002). This shift broadened the perspective, accommodating innovative approaches to decision-making outside the conventional boundaries (Asemi et al., 2011),(Felsberger et al., 2017), (Reza Kheirandish, 2019).

Moreover, the evolution from traditional databases to the implementation of data warehousing and online analytical processing (OLAP), coupled with the shift from mainframe to client-server architecture, presented both an opportunity and a necessity. This transition compelled the exploration of Business Intelligence (BI) and Intelligent Decision Support System (IDSS) solutions to effectively navigate the increasing complexity of business operations (Kirs et al., 2006), (S. Liu et al., 2010b), (Shim J. P. et al., 2002).

Figure 1, the literature review map, provides an overview of the evolution of Decision Support Systems (DSS) within a specific timeframe.

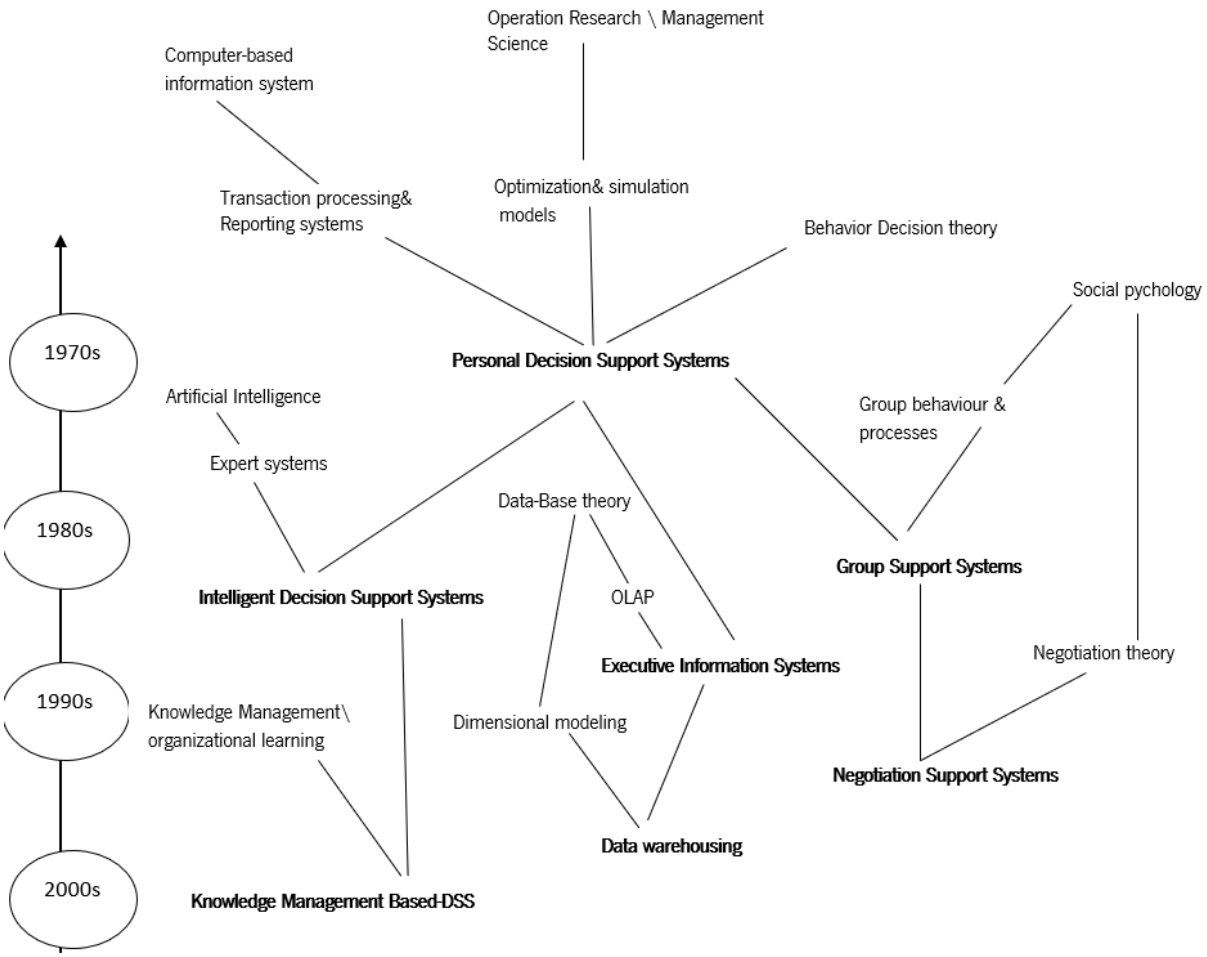


Figure 1: Literature-Map 1960s to 2000s

Adapted from (Arnott & Pervan, 2005)

According to Figure 2, The development phases of Decision Support Systems (DSS) from 2000 to 2019 can be summarized as follows:

2000-2005:

- Integration of data warehousing and business intelligence technologies.
- Emergence of web-based DSS, enabling remote access and collaboration.
- Advancements in data mining techniques for predictive analytics.

2005-2010:

- Growth of real-time decision support systems, facilitating faster decision-making.
- Expansion of DSS into mobile platforms, enhancing accessibility.
- Integration of social media data for decision support purposes.

2010-2015:

- Adoption of cloud computing for DSS, enabling scalability and cost-effectiveness.
- Focus on big data analytics, handling large volumes of structured and unstructured data.
- Emphasis on personalized decision support, utilizing machine learning algorithms.

2015-2019:

- Rise of prescriptive analytics in DSS, providing recommendations for optimal actions.
- Integration of artificial intelligence and cognitive computing technologies.
- Enhancement of visualization techniques for better data representation and interpretation.

Throughout these phases, DSS continued to evolve to meet the changing needs of organizations, leveraging advancements in technology and data analytics to provide more powerful and effective decision-support capabilities (Delen, 2020),(Sprague, 1980),

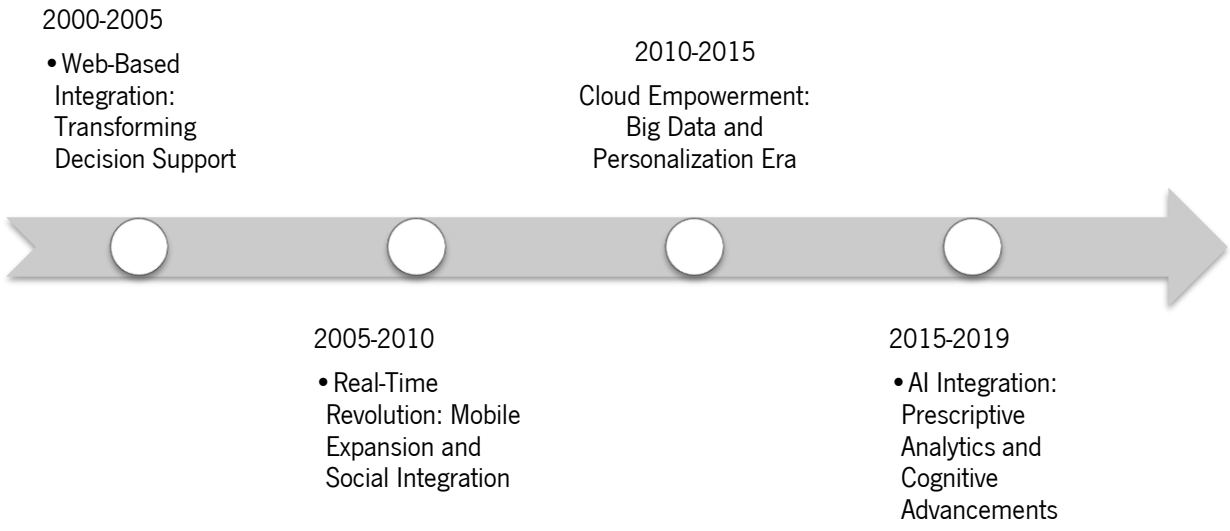


Figure 2: Literature-Map 2000s to 2019s

In summary, in the subsequent sections, we explore the significant evolution of Decision Support Systems (DSS) spanning the years 1980 to 2020, as depicted in Figure 3. The progression began with the inception of DSS in the 1980s, followed by the emergence of Enterprise Data Warehousing in the 1990s. The early 2000s saw the rise of real-time data warehousing, which continued to evolve until 2010.

From 2010 to 2019, a notable shift occurred with the introduction of "Adaptive Business Intelligence" (ABI), challenging traditional paradigms of "Business Intelligence" (BI). Concurrently, the concept of "Business Analytics" (BA) emerged, ushering in a transformative approach to decision-making.

The landscape of problem-solving, heavily influenced by analytics, has prompted a shift in terminology. Terms like "intelligence," "mining," and "discovery" are gradually being replaced by the more comprehensive term "analytics." This linguistic evolution is evident in the transition from "Business Intelligence" to the broader scope of "Business Analytics," and from "Knowledge Discovery" to the encompassing term "Data Analytics" (Delen, 2020). This reflects the evolving nature of decision support systems and their reliance on advanced analytical methodologies for nuanced insights and informed decision-making (Praful Bharadiya, 2023)

Finally, in 2019, we witnessed the advancement of cognitive computing, which seamlessly integrated with the adoption of AI in decision-making processes, further enhancing the capabilities of DSS (Lytras et al., 2020)

Data analysis by humans can be a labor-intensive process, consuming significant time and resources. Utilizing sophisticated cognitive systems can alleviate this burden by efficiently processing vast amounts of data. Cognitive computing presents a promising solution to mitigate the challenges associated with handling big data (S. Gupta et al., 2018).

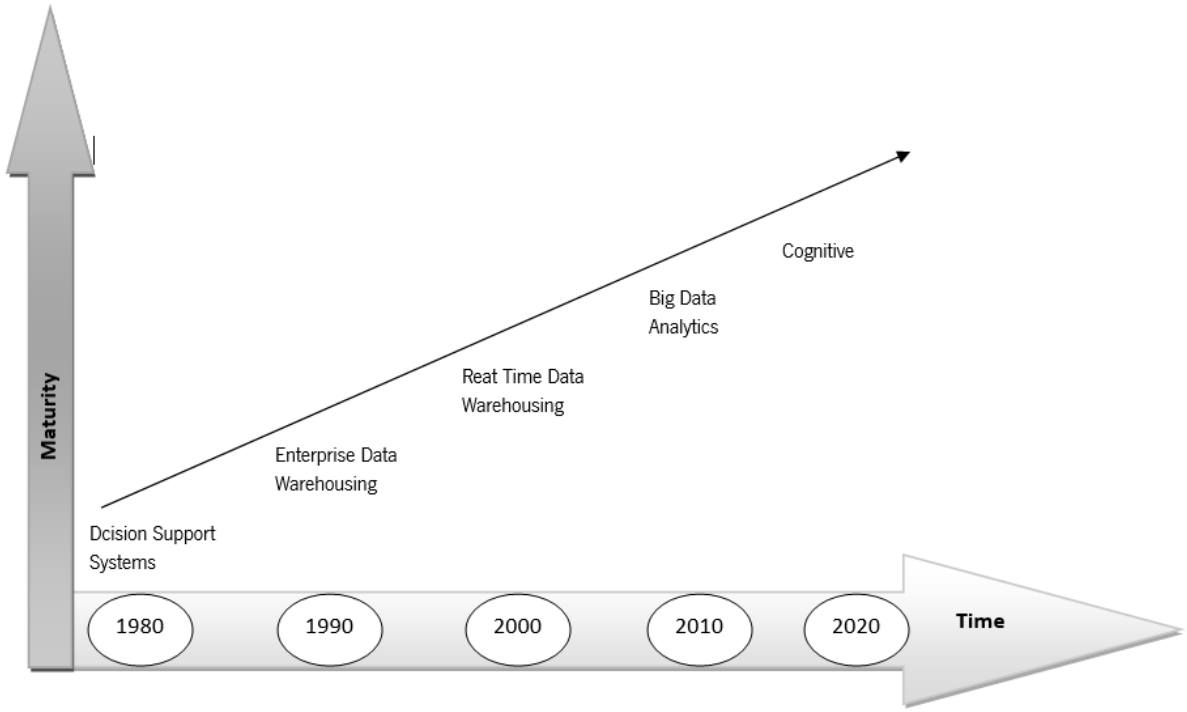


Figure 3: Maturity model of DSS

Adapted from (Watson, 2017b)

The cognitive aspect involves the integration of cognitive computing technologies into DSS, significantly influencing its development during this period in several ways, including:

- Cognitive computing involves the use of AI techniques, such as natural language processing (NLP), machine learning, and neural networks, to mimic human-like cognitive abilities in processing and analyzing data (S. Gupta et al., 2018; Lytras et al., 2020).

- AI-powered DSS can interpret and understand unstructured data sources, such as text and images, enabling more comprehensive decision support capabilities.

Enhanced Data Interpretation:

- Cognitive DSS can analyze complex datasets more effectively, extracting meaningful insights and patterns that may not be apparent through traditional analytics methods.

- AI algorithms in DSS can learn from past data and adapt to changing environments, improving decision-making accuracy over time (Behera et al., 2019).

- Cognitive computing enables DSS to provide personalized decision support tailored to individual user preferences and behaviors.

- By analyzing user interactions and feedback, cognitive DSS can refine recommendations and insights to better meet user needs.

Real-time Decision Making

- Cognitive DSS can process and analyze data in real time, enabling organizations to make faster and more informed decisions(Tarafdar et al., 2017).
- By continuously monitoring data streams and detecting patterns or anomalies, cognitive DSS can alert users to potential opportunities or risks in real time (Hugh J. Watson,2028).

Overall, the integration of cognitive computing technologies into DSS has significantly advanced the field, empowering organizations to harness the full potential of their data for better decision-making(Baxter et al., 2008; Bini, 2018; Srivani et al., 2023).

2.1.1 Types of Decision Support Systems

The taxonomy of DSS delineates a classification based on the inherent operations it performs, independent of variables such as the type of problem, function, or situation. DSS can assume varied orientations, being either completely model-oriented or data-oriented. Due to the inherent heterogeneity within the DSS category, as noted by Steven Alter in 1977, there hasn't been a universally agreed-upon taxonomy for DSS.

Haettenschwiler introduced a categorization scheme, classifying DSS into passive, active, and cooperative types. Passive DSS cannot provide suggestions, while active DSS can proactively propose solutions. Cooperative DSS leverages feedback mechanisms to enhance the decision-making process. Additionally, Sprague and Carlson in 1982 characterized DSS as a system that collaborates with other components of the information system to facilitate decision-making tasks. While no specific classification system for DSS exists, Power suggests a categorization based on variables like the frequency of decision-making and the complexity of the situation, distinguishing between structured, semi-structured, and unstructured problems.

Figure 4 illustrates this classification, showcasing six categories of DSS for semi-structured and unstructured decision-making situations. These categories include Communication-Driven, Data-Driven, Document-Driven, Knowledge-Driven, and Model-Driven DSS.

In particular, represents a form of Group Decision Support System (GDSS) centered around facilitating communication, information sharing, and coordination among two or more individuals. An exemplary instance of this type is web conferencing, illustrating the collaborative nature of Communication-Driven

DSS (Stanek et al., 2014).

Data-Driven Decision Support Systems (DSS) are characterized by their reliance on business analytics, leveraging substantial datasets for analysis. Within this category, Business Intelligence (BI) systems stand out, encompassing both operational and strategic functionalities (D. J. Power, 2008),(Wong & Wang, 2003).

In addition, Document-Driven DSS, exemplified by Correspondence or Document Management Systems (DMS), falls under the umbrella of Knowledge Management Systems (KMS). These systems utilize document processing technologies, including retrieval, versioning, and analysis, to support the decision-making process.

Knowledge-driven systems represent another type of DSS designed to assist managers in decision-making through recommendations and suggestions. These systems are typically structured around Artificial Intelligence and statistical techniques, shaping their architectural framework.

Lastly, Model-Driven systems employ analytical tools rooted in statistical, mathematical, optimization, and simulation models to dissect and understand the situation. In essence, quantitative models play a pivotal role in shaping the functionality of these DSS.

Moreover, automated decision-making is most suitable for well-structured situations and frequent decision-making scenarios. In contrast, computerized special studies are designed for rational decision-making, utilizing analytical processes when faced with unstructured and infrequent situations. Automated decision processes in this context are devoid of human involvement and rely on programmed algorithms, AI, or statistical and mathematical modeling for execution (D. Power, 2017), (Felsberger et al., 2017), (Prakash & Sarkar, 2015), (Reza Kheirandish, 2019),(Richard & Averweg, n.d.).

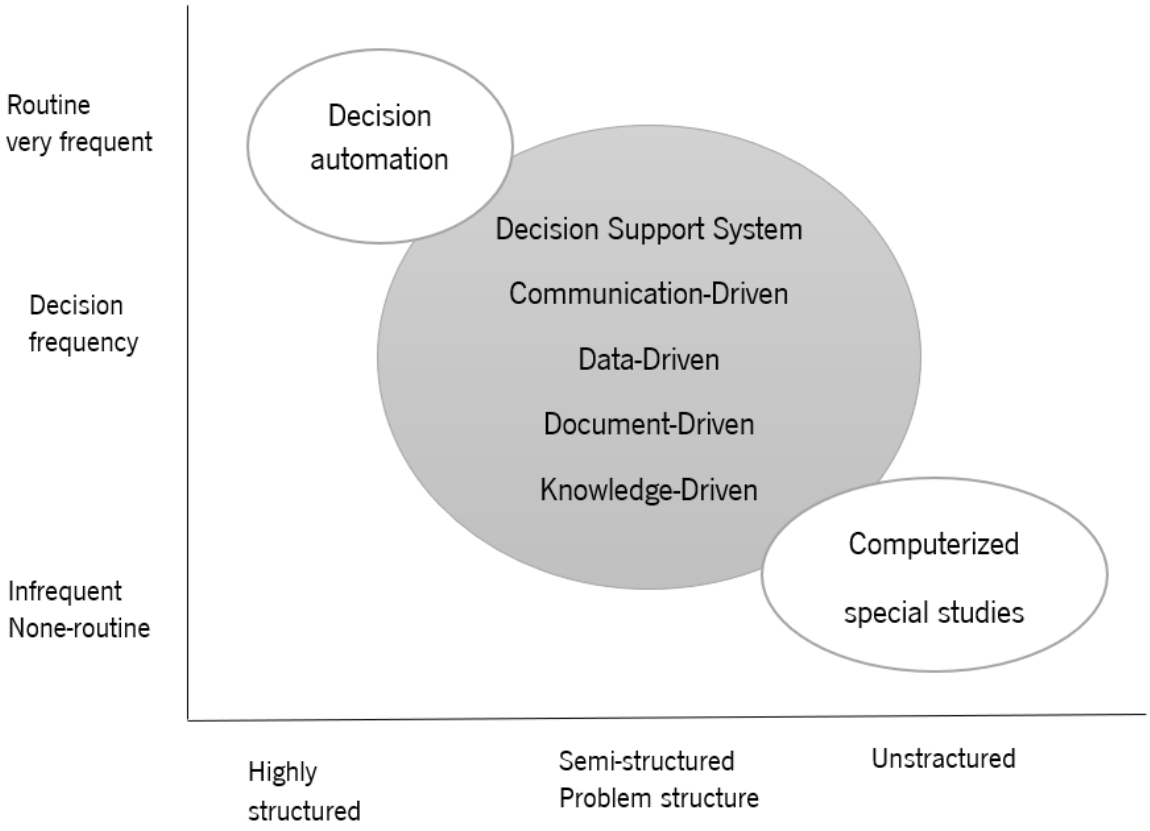


Figure 4: Types of DSS
Adapted from (Ciara Heavin, Daniel J. Power,2017)

2.1.2 Intelligent Decision Support Systems

The architecture of an IDSS follows a structured process encompassing input, processing, and output, all integrated with a feedback loop. Input sources may include data, knowledge, advice, algorithms, or models. Processing involves tasks such as forecasting, generating recommendations, or providing explanations based on the organized input, leading to the generation of output—the outcome of the processing. Importantly, the system utilizes this output as the new input for subsequent analyses (PHILLIPS-WREN, 2012).

In essence, an intelligent system derives the most logical actions by thoroughly analyzing data and available information. Expert systems, for instance, represent a category of intelligent systems that rely on a knowledge base as input and store rules for decision-making. Additionally, IDSS leverages data mining techniques to acquire intelligence and knowledge critical to the decision-making process. The application of data mining technology serves as a knowledge discovery tool, extracting trends, patterns,

and insights from databases to inform decision-making through recommendations and predictions (Aristodemou & Tietze, 2018), (Phillips-Wren, 2008), Jean-(Jean-Charles Pomerol, 1997), S. Liu, Duffy, (S. Liu et al., 2010a), (Skulimowski, 2016).

IDSS exemplifies intelligent behavior, encompassing self-learning from experience, swift and adept responses to novel situations, knowledge processing, operational functions based on probabilistic data, value generation, filtering, reasoning for problem-solving, and the ability to generate answers with incomplete input variables(Kuceba & Kieltyka, 2009), (PHILLIPS-WREN, 2012). The multifaceted nature of IDSS has led researchers to conceptualize it in various ways, each emphasizing specific aspects.

For instance, the concept of an Intelligent, Interactive, and Integrated Decision Support System (3IDSS) is particularly beneficial for large-scale enterprises dealing with economic and social development strategies. In scenarios where high levels of management involvement and intricate environmental complexities are present, the integration of all influential factors becomes crucial. In such cases, a DSS with a heightened level of machine-human interaction proves to be essential.

Another notable example is the Intelligent Decision Support System based on Natural Language Understanding. This variant of IDSS operates on natural language understanding, addressing scenarios where a substantial amount of unstructured information is presented in text format. Consequently, this type of IDSS employs intelligent techniques to navigate and extract insights from natural language text efficiently. Furthermore,(Zhou et al., 2008) the IDSS based on Knowledge Discovery (IDSSKD) plays a pivotal role in facilitating decision-making processes in commerce and finance. This form of intelligent support relies on predictive analytics and in-depth analysis to enhance decision-making outcomes (Zhou, Yang, Li, & Chen, 2008).

2.2 Precision Medicine as a New Approach in Clinical Decision Making

2.2.1 Precision Medicine: an Evolving Paradigm in Clinical Decision-Making

In 2011, the National Research Council, through its Toward Precision Medicine initiative, elucidated precision medicine as the personalized tailoring of medical interventions to the unique characteristics of each patient. This involves classifying individuals into distinct subpopulations based on varying susceptibilities to specific diseases or responses to treatments. The essence of precision medicine lies in its ability to guide healthcare decisions, directing them toward the most effective treatments for individual patients. Doing so, not only enhances the overall quality of care but also diminishes the need for

unnecessary diagnostic tests and therapies. This strategic approach, as highlighted by Geoffrey S. Ginsburg and Kathryn A. Phillips in 2018, ensures that preventative or therapeutic interventions are concentrated where they will be most beneficial, thus sparing both expenses and potential side effects for those who would not derive significant benefits.

On January 20, 2015, former President Barack Obama heralded a groundbreaking advancement in healthcare with the introduction of the "Precision Medicine" initiative. This initiative, a transformative leap in medical practice, was envisioned to significantly enhance public health. President Obama emphasized its potential, stating, "Tonight, I'm launching a new Precision Medicine Initiative to bring us closer to curing diseases like cancer and diabetes and to give all of us access to the personalized information we need to keep ourselves and our families healthier" (Maglaveras et al., 2016).

The innovative approach to medical practice, as outlined by Francis S. Collins in 2015, encompasses two key components: a short-term focus on cancer diseases and a long-term perspective addressing a spectrum of other illnesses. According to the U.S. National Library of Medicine, "precision medicine" is an emerging paradigm that accounts for individual differences, including genes, environment, and lifestyle, to prevent and treat specific diseases for everyone (Mesko, 2017b), (Francis S. Collins, 2015),(Haque et al., 2020b).

Additionally, the U.S. National Cancer Institute defines Precision Medicine as "a form of medicine that uses information about a person's genes, proteins, and environment to prevent, diagnose, and treat disease" (Haque et al., 2020b).

This progressive methodology seeks to replace the antiquated "one size fits all" model with a patient-centric "patient like me" approach. Its core mission is to address issues like inefficient treatments and medical errors, ultimately reducing the burden of overtreatment and hospitalization while saving more lives (Glen H. Murata, Albuquerque, 2014). Moreover, this approach represents a pivotal shift in medical science, where the core objectives include predicting the likelihood of developing a disease, achieving precise diagnoses, and optimizing the most effective treatment for individual patients(Awwalu et al., 2015). The overarching goal is to usher in a new era of healthcare that is not only more personalized but also more effective in addressing the unique needs of each patient.

For over a decade, Hood and colleagues have championed the concept of P4 Medicine—Personalized, Preventive, Predictive, and Participatory (Chen & Snyder, 2013), (Mesko, 2017c).This forward-looking paradigm encompasses "Predictive" medicine, leveraging longitudinal biological information for comprehensive health insights. The "Preventive" aspect focuses on early disease detection, contributing significantly to population health. "Personalized Medicine," a widely used term until 2015, saw a transition

with former President Obama's introduction of "Precision Medicine" (Hasenfuß & Vogelmeier, 2019) emphasizing evidence-based approaches to therapeutics and diagnostics. The "Participatory" element advocates for wellness through consideration of environmental indicators, requiring active population engagement in clinical practice.

It's crucial to discern between "Personalized Medicine" and "Precision Medicine." Despite the widespread interchangeability in literature, these terms represent distinct concepts. The U.S. National Library of Medicine clarifies this distinction, stating that while "Personalized Medicine" and "Precision Medicine" were initially used interchangeably, they have diverged in meaning. The National Research Council, expressing concerns about potential misinterpretation, favored the term "Precision Medicine." In this context, precision medicine focuses on identifying effective approaches for patients based on genetic, environmental, and lifestyle factors, steering away from the notion that treatments are uniquely tailored to each individual (Mosavi & Santos, n.d., 2022a).

An alternative perspective on precision medicine underscores the significance of Tailored Treatment Strategies. Precision medicine is commonly defined as a methodology that customizes treatments according to the distinct needs of individual patients. This customization is rooted in considerations such as their genetic composition, biomarkers, phenotypic traits, or psychosocial characteristics, which set apart one patient from another despite sharing highlighted clinical symptoms (König et al., 2017), (Jameson & Longo, 2015).

Widely embraced definitions of precision medicine span several dimensions. Alternatively, precision medicine is perceived as an all-encompassing, data-driven paradigm that underscores the significance of integrating diverse types of data, incorporating clinical information and other pertinent data sources. This data integration enables the categorization of patients into distinct subgroups, with the expectation that these subgroups share a common foundation of disease susceptibility and presentation. Ultimately, this data-driven approach aims to facilitate more accurate and personalized therapeutic solutions. These dual aspects of precision medicine collectively underscore its commitment to tailoring medical care to individual patients while leveraging advanced data analysis and integration techniques for enhanced diagnosis and treatment (Gupta & Kumar, 2023).

While a precise definition of PM may remain elusive, it is widely acknowledged that the PM approach transcends the traditional symptom-based method in clinical decision-making, emphasizing the consideration of individual variabilities as presented in Figure 5. As a result, this groundbreaking concept embraces a diverse range of individual data, including biomarkers, lifestyle factors, environmental

influences, and genetic information, thereby underlining the comprehensive and personalized nature of PM(König et al., 2017), (Sanchez-Pinto et al., 2023),(Pelter & Druz, 2024).

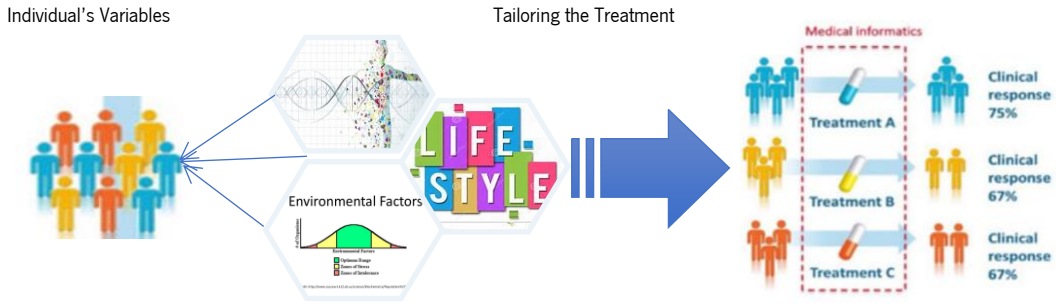


Figure 5: Precision Medicine
(NM, Santos 2021)

Hence, the conclusion can be drawn that precision medicine is undergoing rapid evolution, driven by key factors such as the abundance of big data, advancements in genomics and AI, and ultimately, the field of data analytics.

2.2.2 Catalyst of Change Is Revolutionizing The Clinical Decision-Making

The shift from conventional clinical decision-making to a data-driven model is propelled by an array of influential factors. Illustrated in Figure 6, these encompass technological advancements, the availability of diverse data, and considerations related to patient expectations, government initiatives, and educational levels.

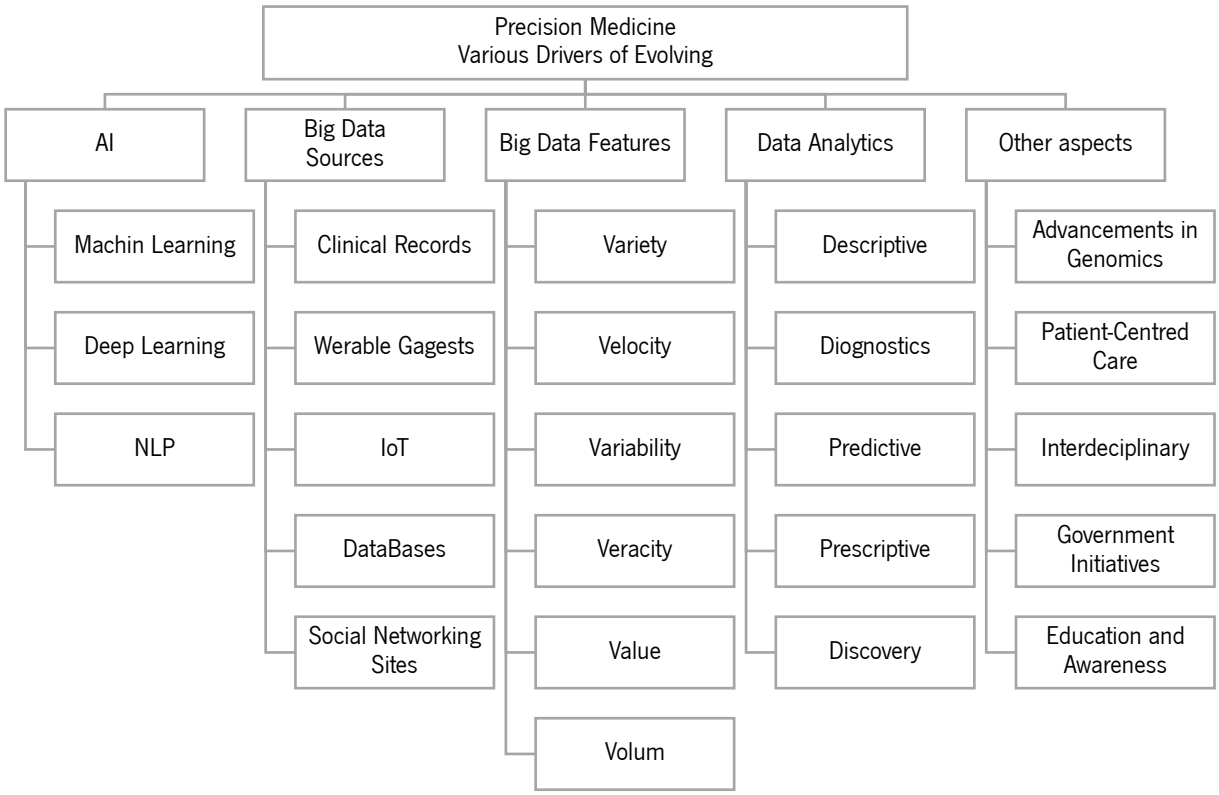


Figure 6: Various Drivers of evolving

a. Emerging Big Data: Transformative Advances and Paradigm Shifts in the Clinical Decision-making

The term 'Big Data' emerged prominently in 1997 within the realm of data visualization, originally addressing challenges in biomedicine and healthcare. Its inception sought to revolutionize data analysis in these fields. Over the years, big data analytics has catalyzed transformative advancements in clinical and molecular knowledge discovery. Initially inspired by big science and business intelligence, a pivotal moment in 2009 showcased the potential of big data when Google searches accurately predicted flu-like illnesses, rivaling established monitoring networks. This marked a turning point, recognizing big data as a viable solution for modern healthcare challenges(N. S. Gupta & Kumar, 2023).

In the last two decades, medical organizations and researchers have shifted towards extracting data from digitized clinical records, departing from traditional collection methods. Enabled by the features of big data, including Volume, Velocity, Variety, Veracity, and Variability (Vs) (Leff & Yang, 2015), and information load (Delen, 2020) data analytics now handle vast and intricate patient-related information. This includes signs and symptoms, clinical records, patient behavior, imaging data, medications, and insurance claims. This paradigm shift has yielded benefits such as real-time care, cost reduction, enhanced data accuracy, and improved efficiency.

In the current digital revolution, our data landscape has expanded significantly, as depicted in Figure 7. Traditionally, data primarily originated from business transactions with relatively low to moderate levels

of volume, variety, and velocity, such as imaging data, clinical notes, genomics, research data, insurance, and pharmacy records. However, with the surge in human-generated transactions via web applications and social media (e.g., IoT, Twitter, Facebook), this data complexity has evolved to moderate to high levels. Furthermore, machine-generated data from the Internet of Things (IoT), sensors, automated transactions, and smartphone applications, which are increasingly prevalent in monitoring patient data, present significant opportunities for advanced analytics (Mosavi & Santos, 2022b),(Delen, 2020).

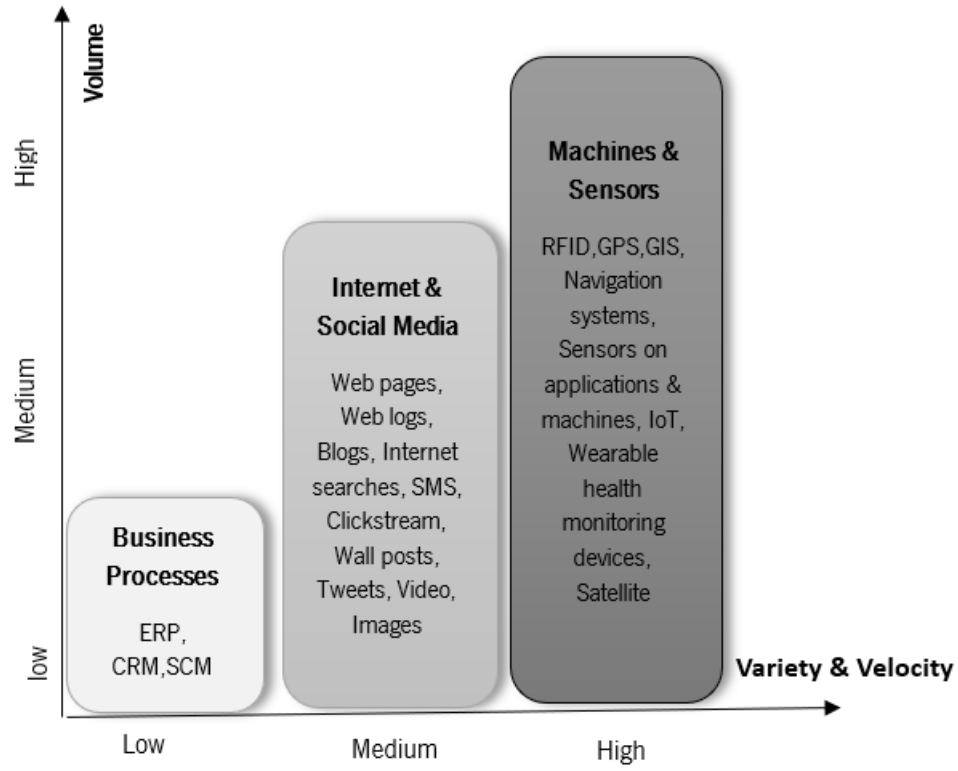


Figure 7: Sources of Big Data
Adapted from (Delen, 2020)

Table 1 outlines the key characteristics of the six Vs (Volume, Variety, Velocity, Variability, Veracity, and Value) in the context of healthcare, as elucidated by Wang (2017). The processing and analysis of the ever-expanding data with these attributes necessitate the use of supercomputers and advanced analytics to facilitate timely clinical decision-making (Mesko, 2017c).

Volume: The size of healthcare data is substantial, driven by factors such as patient treatment plans and varying patient conditions. Variety: Reflecting the diversity of data, healthcare information is non-uniform, originating from various sources like email, documents, XML, OLAP, and others. Moreover, Velocity: The speed at which data is generated and processed poses a challenge for timely decision-making in

healthcare. In addition, Variability: social media serves as a prime example, highlighting the inconsistency in data flow and presenting a challenge in managing such variability.

Veracity: Given the multitude of resources generating patient data, the accuracy, quality, and reliability of the data are subject to debate (Lasorsa et al., 2016). Finally, Value: Undoubtedly, the value of such data in longitudinal studies is significant and holds considerable influence (Delen, 2020; Wang, 2017).

Table 1: Features of Big Data in healthcare (Wang, 2017)

Characteristics	Description
Volume	High-throughput technologies, continuous Monitoring of Vital Signs
Variety	Heterogeneous and unstructured data sources, different in frequencies and taxonomies
Velocity	Increasing data generation rate by the Health infrastructure, high-speed processing fo fast Clinical Decision Support.
Variability	Seasonal Health effects and disease evolution, non-deterministic models of illness and health.
Veracity	Data coming from uncontrolled environments, data quality is unreliable.
Value	Clinically Relevant Data, Longitudinal Studies.

b. Emerging AI: Revolutionizing clinical decision-making

In this groundbreaking approach, the power of AI is harnessed to analyze medical data, transforming decision-making processes and ensuring the delivery of accurate and personalized treatments. The seamless integration of health data into platforms like Electronic Medical Records (EMR) and Electronic Health Records (EHR) opens up exciting possibilities for analytics to drive precision in care, as demonstrated by Heart et al. in 2017. This integrated AI-driven approach not only holds the potential to elevate the quality of care but also exerts a positive influence on population health and the overall cost of healthcare, as elucidated by Chase 2018 (Chase et al., 2018).

Precision medicine extends beyond genomics to include lifestyle, environmental, and clinical data. The integration of these diverse data sources creates a comprehensive understanding of an individual's health, unraveling the intricate factors that shape it. The era of Emerging AI in healthcare is marking transformative advances and paradigm shifts, placing unparalleled precision at the forefront of clinical decision-making.

c. Emerging Analytics: Transformative Advances and Paradigm Shifts in the clinical decision making.

While analytics has been in use since the late 1960s, primarily involving computers in decision-making tasks, its significance gained prominence in 2010, particularly with the surge in available Big Data.

Analytics leverages a spectrum of data, encompassing both structured and unstructured formats, alongside information and knowledge, to address business problems of varying complexity (Delen, 2020). Data Analytics has emerged as a transformative force in healthcare, pivotal in facilitating optimal clinical decision-making (Cirillo & Valencia, 2019; Sousa et al., 2019). Within the realm of analytics, distinct methodologies serve specific functions in harnessing the potential of clinical data (Pencina & Peterson, 2016; Vegter, 2018). As shown in Figure 8, Descriptive and diagnostic analytics uncover connections within data, shedding light on patterns and insights. Predictive analytics takes a step further by estimating future events, while prescriptive analytics assumes the role of a decision-making guide, recommending optimal courses of action. As illustrated in Figure 8, the spectrum ranges from informative (descriptive/diagnostic) to decision-focused (prescriptive).

The outcomes of descriptive and diagnostic analytics manifest as reports, dashboards, and visualizations, addressing questions related to "what happened" and "why it happened." However, these outcomes necessitate interpretation and analysis by domain experts. For instance, laboratory results provide crucial insights into a patient's health status concerning specific tests, and descriptive analytics empowers physicians to observe and analyze these results in the context of a patient's overall health. In healthcare, descriptive and diagnostic analytics play a pivotal role in disease detection and early intervention. Furthermore, predictive analytics becomes invaluable in identifying patients at risk, offering critical support for care-based practices (Alharthi, 2018).

elevation and prediction, the realm of prescriptive analytics tackles the question of “what we should do”, achieving optimal and effective treatment demands the application of mature and sophisticated analytics, capable of selecting the most efficient courses of action from a multitude of options. To attain such precision, prescriptive analytics leverages sensitivity analysis and what-if scenarios to evaluate all potential choices, ultimately optimizing outcomes (Mosavi & Santos, 2020, 2021a, 2021b),(Menezes et al., 2019; Sivarajah et al., 2017).

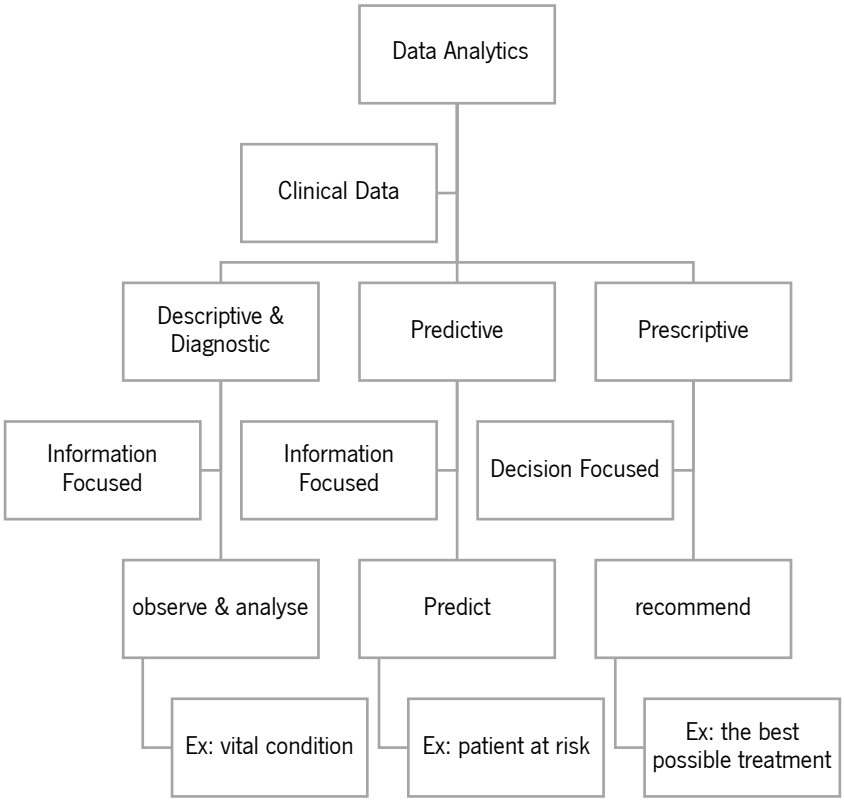


Figure 8: Healthcare Analytics

To underscore the implications of each analytics approach on clinical decision-making, we turn to the succinct and practical model of rational decision-making known as "Intelligence-Design-Choice" by Herbert Simon.

Simon introduced the concept of bounded rationality to address the reality that human decision-making often falls short of achieving truly optimal outcomes. He highlighted that human cognition is inherently limited, making it impossible for individuals to consider and analyze every conceivable alternative and its consequences during problem-solving. Consequently, decisions may be influenced by individual preferences rather than being strictly rational. In response to these limitations, Simon devised a systematic and logical framework, initially consisting of "Intelligence-Design-Choice" to guide rational decision-making. Later, he added an "Implementation" phase as a fourth step, incorporating monitoring and feedback from each preceding phase to validate, verify, and refine the decision-making process (Delen, 2020), (Woodside et al., 2016).

The "Intelligence" phase encompasses activities such as information gathering and scanning. This stage also involves the critical task of problem identification and definition. Moving to the "Design" phase, the focus shifts to conceptualizing the problem, defining assumptions, generating alternative solutions, and

constructing models. Finally, the “Choice” phase pertains to goal attainment, involving the evaluation and selection of the best possible course of action (Simon, 1959; Woodside et al., 2016).

Considering the role of analytics in rational decision-making, it becomes evident that each type of analytics corresponds to specific phases within this framework. Descriptive and diagnostic analytics are instrumental in data collection and analysis, problem definition, and knowledge extraction, aligning with the “Intelligence” phase. Predictive analytics, which involves formulating and generating alternatives, is well-suited to the “Design” phase, given its capacity to provide insights and options (Ramesh Sharda, Dursun Delen, 2014). Additionally, because predictive analytics offers informative capabilities, it also aligns with the “Intelligence” phase. Lastly, prescriptive analytics, with its focus on providing actionable recommendations and achieving optimal outcomes, naturally corresponds to the “Choice” phase. In summary, each analytics approach serves distinct functions within the decision-making process, contributing to a more informed and rational approach to clinical decision-making(Mosavi & Santos, 2020; Palanisamy & Thirunavukarasu, 2019).

d. Other influential factors references

A considerable number of individuals encounter the inefficacy of medications in their daily treatment, emphasizing the imperative need for a more precise approach(Schork, 2015). The rising aging population, coupled with the diverse responses of patients to treatments, imposes global economic (K. B. Johnson et al., 2021; McDonald et al., 2000).

As expectations for enhanced care and service quality rise, there is a simultaneous decline in the economic capacity to deliver high-quality services, posing a formidable challenge for healthcare providers. In addition, rapid progress in genomics, particularly with techniques like whole genome sequencing, has analyzed individual genetic information more accessible and affordable. This progress enhances our understanding of the genetic basis of various diseases and opens avenues for tailored treatments. Pharmaceutical companies are increasingly developing targeted therapies designed for individuals or specific patient subgroups based on their genetic profiles. This approach has shown promise in cancer treatment and beyond.

Another point of view is patient-centered care, precision medicine places a strong emphasis on patient-centric care, integrating individual preferences, values, and goals into the treatment decision-making process. Its application extends across medical specialties, including oncology, cardiology, psychology, and rare diseases, with potential applications in numerous other areas of medicine(Kosorok & Laber, 2019a; Latimer et al., 2017).

Furthermore, governments and healthcare organizations globally are investing in precision medicine research and infrastructure to foster its development and adoption. The growing emphasis on educating healthcare providers, researchers, and the public about the potential benefits of precision medicine and its integration into clinical practice is evident. As our understanding of genetics, data analysis, and medical technology expands, the concept of precision medicine is poised to evolve further. This evolution holds the promise of offering more targeted and effective treatments for a wide range of diseases, marking a significant shift toward personalized and data-driven healthcare, a trend expected to persist in the years to come(Milewa, 2009).

2.2.3 IDSS4PM; Gaps and Limitations

The evolution of intelligent decision support systems for precision medicine is a dynamic and rapidly advancing field. Despite extensive research at the application level, numerous challenges and scientific gaps persist in the development of data-driven clinical decision support systems. This section delves into the key obstacles and gaps encountered in this progressive domain which has been summarised In Figure 9.

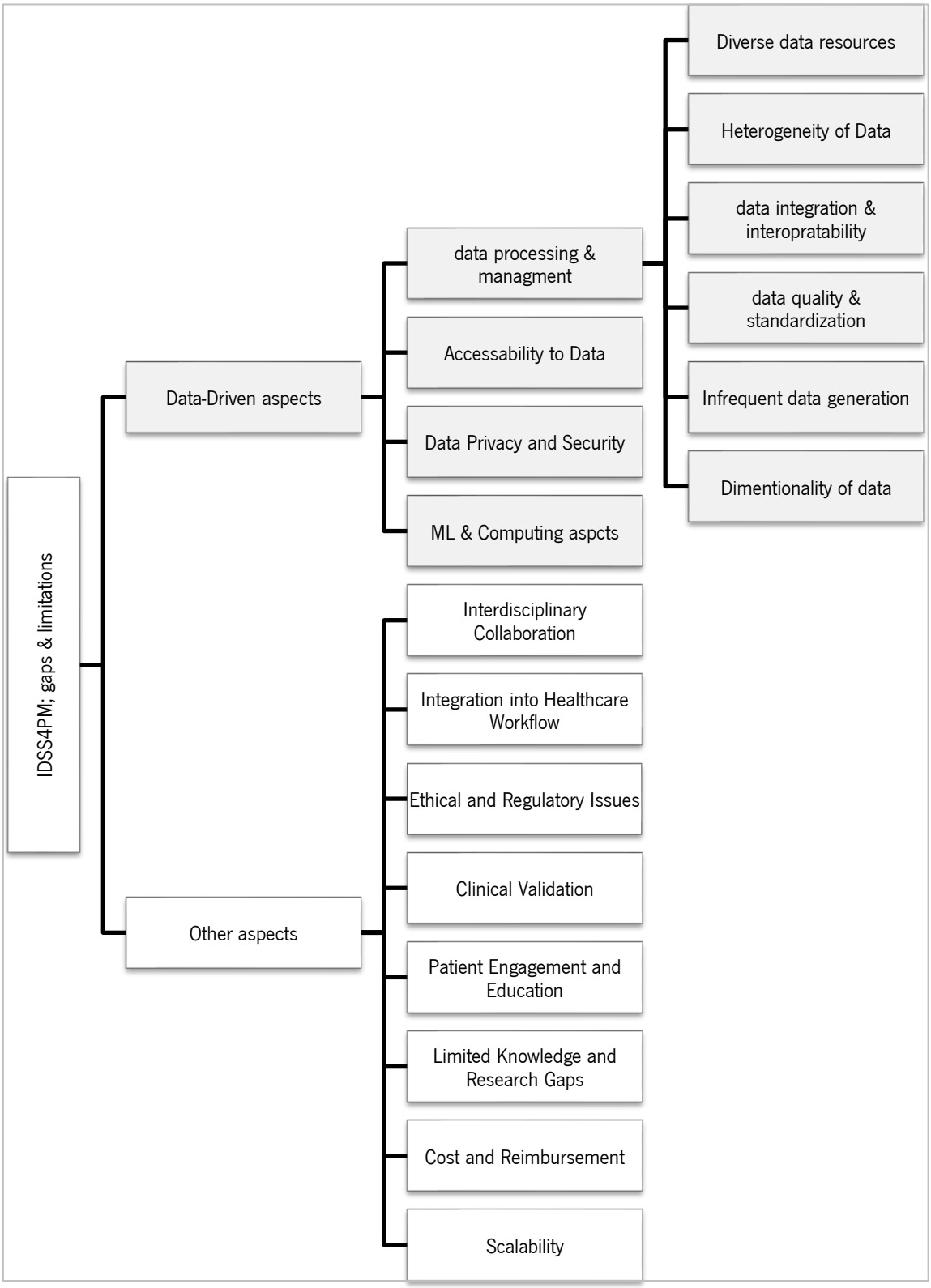


Figure 9:IDSS4PM; gaps and limitations

a. Data processing, and management:

Patient data in precision medicine is highly heterogeneous, including genetic, clinical, lifestyle, and imaging data(Ahmed et al., 2020).

In the realm of digitized healthcare platforms, a prevalent scenario involves the continuous monitoring and storage of various variables of patients, resulting in a vast amount of collected data. The integration of data is crucial for studying its correlation with other information generated during a patient's treatment journey and other clinical aspects. Ensuring that these data sources can work together seamlessly is essential for precision medicine but remains a technical challenge. Moreover, the presence of outliers and abnormal data introduces bias, necessitating the filtration of such data from modeling and study scenarios (Chang et al., 2019; Fenning et al., 2019). Despite the promising strides in Big Data analysis in clinical research, marked by a growing number of peer-reviewed articles, addressing these challenges remains limited. A concerted effort is needed to validate the knowledge extracted from clinical Big Data and seamlessly implement it in clinical practice(Carra et al., 2020).

Numerous studies have contributed fragmented solutions to address these challenges. Table 2 highlights a few major efforts that present promising frameworks. For instance, the "attention scores" technique for feature importance in time series clinical data is a complex yet applicable method suitable for nonlinear data(N. Johnson et al., 2021). Other research endeavors leverage summaries of patient time series data from the ICU, focusing on the early prediction of in-hospital mortality. This approach involves static observations and physiological data, including labs and vital signs, aggregated based on hourly circumstances, thereby addressing data extraction and integration within the context of data aggregation (N. Johnson et al., 2021).

Another significant limitation faced by clinical data for processing pertains to storage and computing, particularly concerning high-frequency data such as physiological indicators. The "Electron" framework has emerged as a solution designed to store and analyze longitudinal physiologic monitoring data (McPadden et al., 2019). Moreover, the Topological Data Analysis (TDA) approach proves effective for large-scale datasets, utilizing algebraic topology to analyze Big Data by reducing dimensionality, especially for geometric representations to extract patterns and gain insights into them. To cope with the velocity of data, the introduction of "anytime algorithms" to learn from data streaming has proven useful for time series data, contingent upon the computation capacity they can handle. Additionally, addressing heterogeneous data (data variety), the Generalized Nonnegative Matrix Tri-Factorization (GNMTF) framework is an efficient solution for data integration, although its competency complexity increases with

the number of data types to be integrated(Gligorijević et al., 2016), (Hulsen et al., 2019), (N. Johnson et al., 2021).

As discussed above, to overcome these limitations, researchers and practitioners have delved into diverse solutions, where those existing approaches have provided solutions to manage specific challenges such as feature importance, feature selection, dimensionality reduction, and addressing issues related to data processing tasks.

Table 2: Techniques to deal with limitations associated with data processing

Technique	Focused area	Limitations
attention scores	feature importance	feature selection, time series data, nonlinear features
post-hoc explanation techniques	Modeling and features relations	Interpretation- black box
summaries of patient time series data	Data extraction- time-series clinical data	vital signs and lab data
Filtering outliers	Data cleaning	Time series data- vital signs
Electron	Data extraction	Physiologic Signal Monitoring and Analysis
Topological Data Analysis (TDAs)	Dimensionality Reduction	
anytime algorithms	Velocity	learn from streaming data
GNMTF	Variety	Complex when data types increase

The other approaches include the implementation of standardized data formats, centralization of data, utilization of ontology and semantic interoperability, adoption of indexing techniques, application of Natural Language Processing (NLP), and exploration of blockchain technology. In this regard, it's important to note that the effectiveness of these solutions may vary depending on the specific context and the nature of the clinical data in terms of the characteristics of the data sources and the objectives of the integration. Additionally, each way to solve those challenges carries limitations and considerations.

i. Standardized data format: Implementing standardized data formats and coding systems can enhance interoperability and facilitate the integration of clinical data from different sources. The use of standardized vocabularies, such as SNOMED or LOINC, helps in achieving consistency. However, achieving universal adoption of standardized formats can be challenging. Furthermore, variations in the implementation and interpretation of standards may persist, leading to interoperability issues(Martínez-García & Hernández-Lemus, 2022).

ii. Ontology and Semantic Interoperability: Developing ontologies that define relationships between different clinical concepts can improve semantic interoperability. This approach enables a more

comprehensive understanding of the data, facilitating the integration of events across diverse sources (Kraus et al., 2018).

iii. Master Patient Index (MPI): Implementing an MPI helps in uniquely identifying and linking patient records across various systems. This is crucial for integrating clinical events related to the same individual, even if the data comes from different sources. The limitation associated with this method is Building and maintaining a reliable MPI requires accurate patient identification, which can be difficult in cases of variations in patient data or discrepancies. Privacy concerns and data security are also critical issues(Lintz, 2020).

iv. Data Integration Platforms: Although, utilizing data integration platforms and tools can streamline the process of combining data from various sources. Complex data integration processes may require skilled personnel and resources. Handling heterogeneous data formats and ensuring data quality during the integration process still has remained a challenge. These platforms often include features for data cleansing, transformation, and loading (ETL) to ensure data quality (Martínez-García & Hernández-Lemus, 2022).

v. Data Warehousing: Creating a centralized data warehouse that consolidates information from disparate sources can provide a unified platform for analysis. Data warehouses are designed to handle diverse data structures and can offer a consolidated view of clinical events. This solution carries major challenges and limitations in terms of in handling real-time data updates and ensuring data quality over time. Because developing and maintaining a comprehensive data warehouse can be resource-intensive (Bae et al., 2015; Diouf et al., 2018; Kherdekar & Metkewar, 2016; Khnaisser et al., 2015; Nacional, 2003; Nakache, 2003; Yang et al., 2024)

vi. Blockchain Technology: Some studies explore the use of blockchain to create a secure and transparent environment for sharing and integrating clinical data. Blockchain can enhance data integrity and traceability. However, While blockchain enhances security and transparency, its scalability for handling large volumes of clinical data is a concern. Regulatory and legal challenges, as well as potential resistance to adopting blockchain in healthcare, are also limitations(Al-Nbhany et al., 2024; Hasan et al., 2022; Karafiloski & Mishev, 2017; Velmovitsky et al., 2021)

vii. Machine Learning aspects: Applying machine learning algorithms and NLP techniques can aid in the identification and extraction of relevant information from unstructured clinical text. This can contribute to a more comprehensive view of patient records

The effectiveness of such techniques is dependent on the quality of input data which is a significant limitation. NLP models may struggle with understanding context and may require manual validation. Generalizability across different healthcare settings and languages can be a challenge(Bzdok & Meyer-Lindenberg, 2018; Naik et al., 2023; Quazi, 2022; Republic Of Korea AI Strategy, 2019; Wilkinson et al., 2020).

In addition, overfitting and underfitting are common challenges in ML. Overfitting occurs when a model learns to fit noise in training data, leading to poor performance on new data. Conversely, underfitting arises when a model is too simplistic and fails to capture patterns in the data. These issues can be addressed by adjusting training data or optimizing model parameters. Successful model development requires large datasets, informatics expertise, domain knowledge, and proper validation. Collaboration between medical and informatics experts is crucial to effectively utilize big data in healthcare (Tran et al., 2021).

b. Data Quality and Standardization:

ensuring data quality verification is a pivotal step in the data processing journey. The quality of data is significantly influenced by key factors, including the clinical team's assessment of the patient's condition, potential misinterpretations of original documents, and errors during data entry(Brown, 2016). Additionally, medical waveforms, such as electrocardiograms and electroencephalograms, commonly employed in physiological examinations, may introduce random noise and gaps (Khadanga et al., 2019). Addressing missing values is another crucial aspect to be handled during data cleaning and preparation(Adiba et al., 2021). Thus, Lack of standardized protocols for data collection and quality control across different healthcare systems and institutions.

c. Data Privacy and Security:

Precision medicine relies on large amounts of sensitive patient data. Protecting patient privacy and ensuring data security while making this information accessible for research and treatment is a significant challenge. Concerns about data breaches, unauthorized access, and patient consent may limit the availability of comprehensive datasets for decision support(Sadat Mosavi & Filipe Santos, 2021).

d. Ethical and Regulatory Issues:

The ethical use of patient data, informed consent, and the potential for discrimination based on genetic information are all important concerns. Regulatory frameworks must evolve to address these ethical issues. Therefore, ensuring that decision support systems are ethically designed and implemented can be challenging, and a lack of ethical guidelines may hinder progress (Maceachern & Forkert, 2021).

e. Interdisciplinary Collaboration:

Precision medicine requires collaboration between healthcare professionals, data scientists, bioinformaticians, and information technology experts. Thus, Gaps in interdisciplinary communication and collaboration may impede the development and implementation of effective decision-support systems (Tran et al., 2021).

f. Scalability:

Precision medicine solutions need to be scalable to handle large and diverse datasets. Therefore, Lack of scalable infrastructure and resources may limit the ability to analyze and process data efficiently.

g. Longitudinal Data Challenges:

Precision medicine often requires longitudinal data to track changes in health over time.
The limited availability of long-term, comprehensive datasets may hinder the development of accurate predictive models.

h. Clinical Validation:

Transitioning from research findings to clinically validated and actionable insights can be slow and challenging. Furthermore, the maturity of existing clinical practices in this domain requires careful consideration. The establishment of new policies, regulations, and collaborative pipelines between stakeholders is essential to expedite the evolution of PM. Successful and valid projects on a larger scale have a positive impact not only on overall performance quality but also indirectly contribute to enhancing best practices, particularly in problem-solving aspects associated with technical areas such as data processing (Blasimme, Fadda, Schneider, & Vayena, 2018). Robust clinical trials and validation studies are required to prove the effectiveness of precision medicine approaches.

i. Integration into Healthcare Workflow:

Integrating precision medicine into routine clinical practice can be complex. Healthcare providers may need training, and electronic health record systems must accommodate new data sources and decision-support tools (Wagner, 2019) (Ahmed et al., 2020).

In addition to the data-driven aspects, other aspects impacted the successful development of such a framework such as:

a. Access to High-Quality Genomic and Molecular Data:

Many patient populations, especially in underserved areas, may lack access to the necessary genetic testing and data. Disparities in access can limit the application of PM.

b. Cost and Reimbursement:

Precision medicine often involves expensive genetic tests and targeted therapies. Ensuring that these are affordable and that reimbursement models support precision medicine can be difficult(Wagner, 2019).

c. Limited Knowledge and Research Gaps:

There are still many gaps in our understanding of the genetic basis of many diseases. Research is ongoing to identify new biomarkers and therapeutic targets.

d. Rare and Undiagnosed Diseases:

Precision medicine is particularly challenging for rare and undiagnosed diseases, where limited data and treatment options exist.

e. Patient Engagement and Education:

Patients need to be educated about precision medicine and engaged in the decision-making process. Informed consent and understanding the implications of genetic testing and treatments are critical.

Addressing these challenges and scientific gaps requires a collaborative effort involving researchers, healthcare providers, policymakers, and patients. Ongoing research and innovation are essential to advancing precision medicine and realizing its full potential in improving patient outcomes and healthcare delivery (Duffy, 2016), (X. Liu et al., 2019),(Cazzola et al., 2017),(Sadat Mosavi & Filipe Santos, 2021).

3. METHODOLOGY AND RESEARCH STRATEGY

This chapter elucidates the research methodology employed, encompassing the methods utilized for crafting the artifact, evaluating the outcomes, and effectively communicating with audiences.

3.1 Design Science Research for Information Systems

Given our objectives, we have selected Design Science Research Methodology for Information Systems (DSR-IS) as the research methodology. DSRM has been a cornerstone paradigm in the field of Information System (IS) research since its inception. As outlined by Shirley Gregor and Alan R. Hevner (2013), "Design science creates and evaluates IT artifacts intended to solve identified organizational problems."

Figure 10 shows the rigorous process involved in creating the artifact to address observed problems. DSR entails six sequential steps, including problem identification and motivation, defining the objectives for a solution, design and development, demonstration, evaluation, and communication, as articulated by Ken Peffers (Peffers et al., 2007).

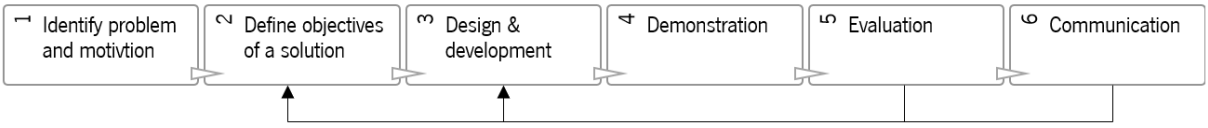


Figure 10: Design science research methodology

1. Identify Problem & Motivate

To create the artifact, it's imperative to identify the problem or research opportunity. As emphasized by Salvatore T. March and Veda C. Storey (2008), problem awareness is crucial for finding relevant solutions(Hevner et al., 2004). Following this principle, our study addressed significant problems that led to the development of our artifact.

The first year of the PhD program provided a valuable opportunity to delve deep into the literature on "Decision Support Systems" and analyze the work of other researchers in the field. Additionally, engaging with reports and formal publications from practitioners offered insights into the gaps between academia and industry demands, as well as potential avenues for future research. Moreover, participation in seminars and relevant subjects as part of the PhD program laid a strong theoretical foundation, essential for creating and designing the artifact.

By leveraging one year's worth of activities to identify business problems and research opportunities in the realm of "Decision Support Systems," we identified "Intelligent Decision Support System for Precision Medicine" as the focal point of our PhD research project. This innovative artifact aimed to address identified problems and demonstrate contributions effectively.

The first step of DSRM involved identifying and summarizing a research opportunity that challenges the conventional approach to clinical decision-making. For decades, evidence-based medicine has been the predominant approach, treating all patients similarly regardless of individual variabilities and disease conditions. This approach has resulted in over-treatment, costly, and complex medication protocols. Additionally, the availability of integrated data platforms in the healthcare domain, coupled with the increasing role of business analytics, AI techniques, and machine learning algorithms for disease diagnosis and prediction, has challenged traditional medical decision-making. The rise in consumerism and the growing aging population, leading to an increase in chronic diseases, further underscores the need for a paradigm shift in medical decision-making towards personalized care. This cultural shift, as outlined by Stephen J. Fenning, Gregor Smith, and Catherine Calderwood (2019), requires considering not only biological factors but also environmental factors, lifestyle, and patient preferences(Fenning et al., 2019).

2. Define the Objectives of a Solution

Defining objectives requires a comprehensive understanding of the problem and the existing situation. Building upon the problem identification phase, our new approach of "patient-like-me" challenges the outdated "one-size-fits-all" approach. In "Precision Medicine," analytical techniques utilize patient health records to support medical decision-making, aiming to tailor treatments for optimal outcomes. This PhD project aimed to propose a data-driven framework within the concept of "Decision Support Systems" to address the new paradigm of "Precision Medicine."

3. Design & Development

This phase involves creating the artifact by identifying and designing the expected architecture and functionality, followed by constructing the actual artifact. Relevant theoretical knowledge is essential for this phase. Following the task schedule and plan, activities such as choosing relevant cases, data collection, preparation, and selecting appropriate methods, are undertaken to design and create the desired functionality via a laboratory experiment.

4. Demonstration

The fourth step involves demonstrating the applicability of the artifact in the real world and proving its efficacy in solving identified problems. This phase requires knowledge of utilizing the artifact to address the problem. A case study method, a proof of concept certifies the validity, reliability, and usefulness of the artifact. This phase falls outside the scope of our current endeavor, as our primary focus has been on developing and presenting a prototype framework. However, it's worth noting that the result of analytical insight serves as a deliverable during the demonstration phase.

6. Evaluation

In the evaluation phase, the effectiveness of the artifact is assessed, comparing objectives with observed results. Utilizing evaluation methods, metric techniques, and indicators, the degree to which the artifact supports problem-solving is demonstrated. In our study, we employed pertinent metrics to evaluate the performance of our laboratory experiments. Furthermore, we elucidated how the final results align with our objectives and research question.

7. Communication

The final step involves communicating the problem, its solution, and the novelty of the solution to researchers and relevant audiences. Publication of the project's results and sharing the outcomes with academia are planned for communication purposes (Peffers et al., 2007).

The outcomes of each phase of this project have been extensively disseminated across academic platforms and indexed by reputable libraries such as Scopus, ScienceDirect, and Google Scholar. Additionally, our work has been featured in renowned journals including Nature. Furthermore, we have actively engaged with audiences through various conferences such as IEEE, where the findings and insights from our research have been shared and discussed.

3.2 CRISP-DM

For our current study, we employed the CRISP-DM 1.0 approach, an acronym for CRoss-Industry Standard Process for Data Mining. While Data Mining (DM) constitutes one of the phases in the Knowledge Discovery from Database (KDD) process, CRISP-DM serves as a practical guide for implementing DM in real-world systems. This methodology assists in translating business problems or application requirements into structured data mining projects, ensuring the effectiveness of outcomes by extracting knowledge from raw data.

Originally introduced in the late 1990s for Knowledge Discovery from Database (KDD), CRISP-DM was developed by a consortium comprising Daimler-Chrysler, SPSS, and NCR. As illustrated in Figure 11, CRISP-DM 1.0 encompasses six phases, each comprising specific tasks and outcomes:

The process begins with Business/Application Understanding, where the objectives and goals for data mining are identified, and the potential contribution of data mining to achieving these objectives is determined. Following this, the Data Understanding phase involves exploring available data, assessing its quality and suitability, and identifying any data issues or challenges. Subsequently, in Data Preparation, relevant data is selected and preprocessed, including cleansing, transformation, integration, and feature engineering. The Modeling phase entails selecting appropriate techniques, building, and training models using prepared data, and evaluating and refining these models for enhanced performance. Evaluation assesses the model's effectiveness in meeting business objectives, validates its performance, and identifies areas for improvement and fine-tuning. Finally, Deployment involves integrating the model into business operations or decision-making processes, developing a deployment and monitoring plan, and documenting results while providing recommendations.

By adhering to the structured approach outlined by CRISP-DM 1.0, we systematically navigate through the data mining process, ensuring the successful utilization of data-driven insights for decision-making.

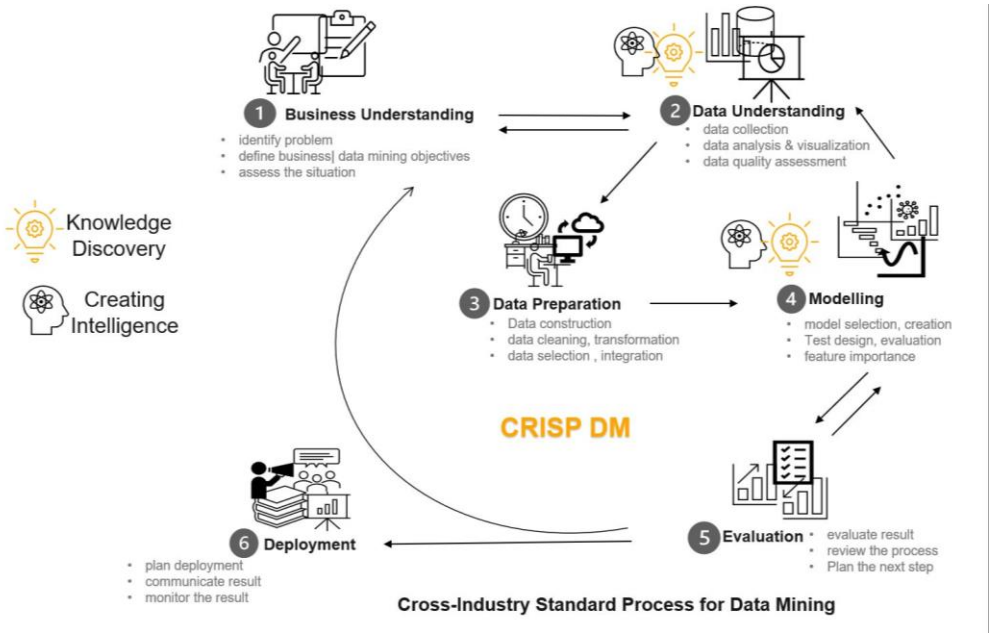


Figure 11: CRISP-DM 1.0

3.3 Mapping Methodologies

In Figure 12, we delineate the key performances of our two experiments across two methodologies. The blue numbering corresponds to CRISP-DM phases, while the purple numbering signifies steps in DSR. Contextual background study and objective identification are allocated to Phase One in CRISP-DM, whereas in DSR, they constitute distinct phases (1 and 2). Furthermore, activities such as data analysis, processing, integration, and modeling are intertwined with the design and development of the artifact. In CRISP-DM, data collection and analysis are associated with Phase 2, while data processing, clinical event interrelation, data transformations, and effective data preparation for modeling align with the data preparation phase. Both methodologies feature an evaluation phase that serves as a practical means to assess artifact performance and compare results with objectives. Additionally, analytical insights are defined in the demonstration phase of DSR, which corresponds to deployment in CRISP-DM. Finally, communication in Phase 6 encompasses result presentation, reporting, and dissemination through academic platforms.

In this study, we applied the CRISP-DM methodology to both projects. In phase one, which primarily involved a literature review and identifying scientific gaps, as well as analyzing the latest investigations and defining objectives, we laid the foundation. Moving to phase two, we conducted data analysis to comprehend and identify patterns, and outliers, and explore data relationships. During the data preparation phase, we utilized techniques such as data cleaning, construction, correlation analysis, and feature selection to effectively prepare the data for modeling.

Subsequently, in the modeling and evaluation phase, we engaged in predictive modeling and evaluation using relevant metrics to rank and assess the performance of algorithms. Across both projects, analytical insights and multidimensional filtering and visualization were integral components in the deployment phase. Furthermore, we communicated the obtained results and outcomes through academic platforms, which were assigned to the deployment phases.

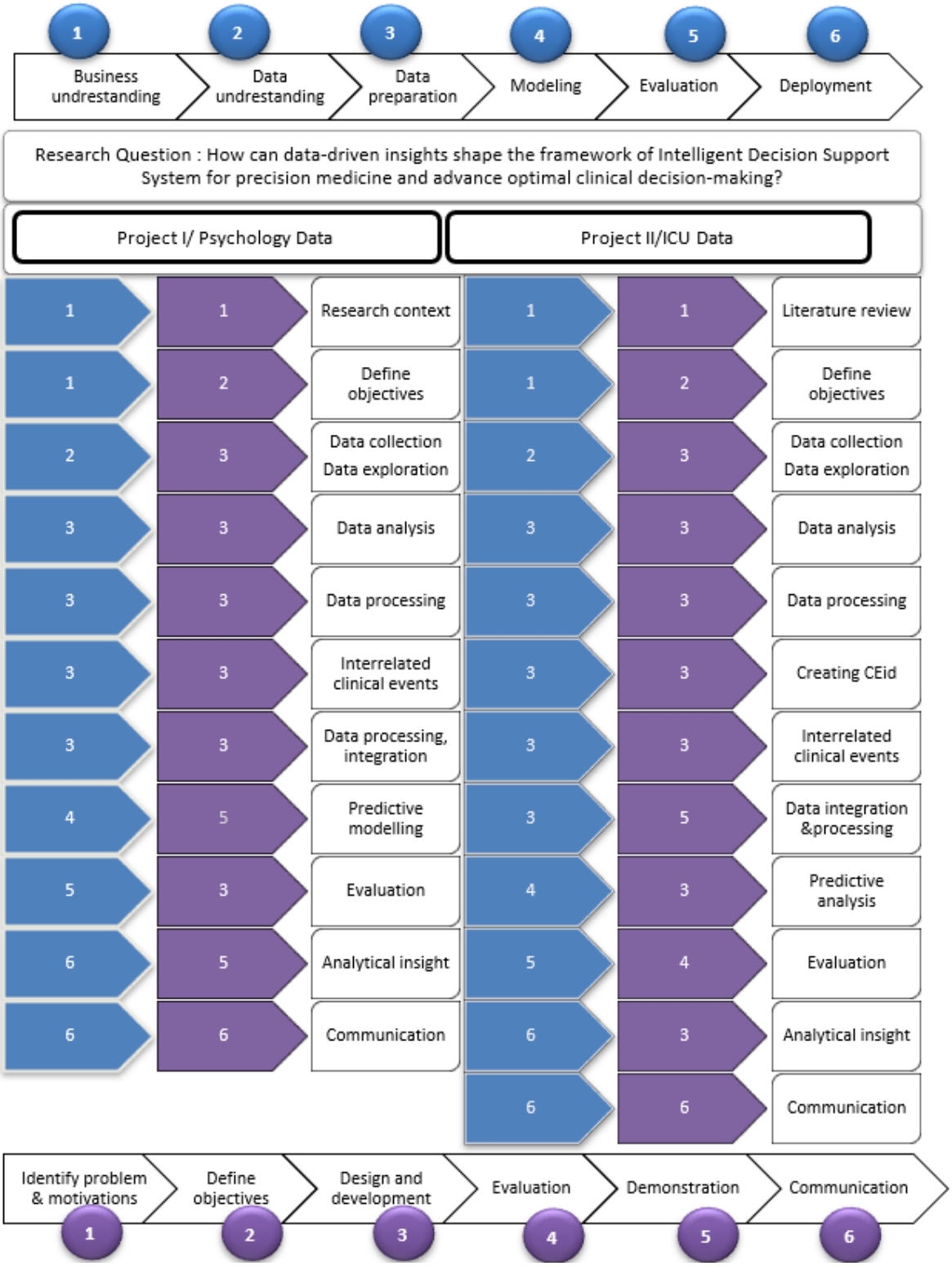


Figure 12: Mapping methodologies

3.4Literature Review Strategy

Summarized in Figure 13, in this study, we conducted an extensive search using key terms such as "Intelligent Decision Support," "Clinical Decision Making," "Precision Medicine," "AI applied in healthcare," and "Optimal Decision Making." We utilized various libraries including Scopus, ScienceDirect, and Google Scholar. The sources of papers were explored across two main areas: Information Systems and Technology, and Decision Support and Health Informatics. Additionally, we delved into interdisciplinary sources from the healthcare domain, including biotechnology, medicine, annual reviews, and statistics, spanning the period from 1997 to 2024. It's noteworthy that foundational concepts such as decision science and theoretical aspects have been extensively cited and date back further in time.

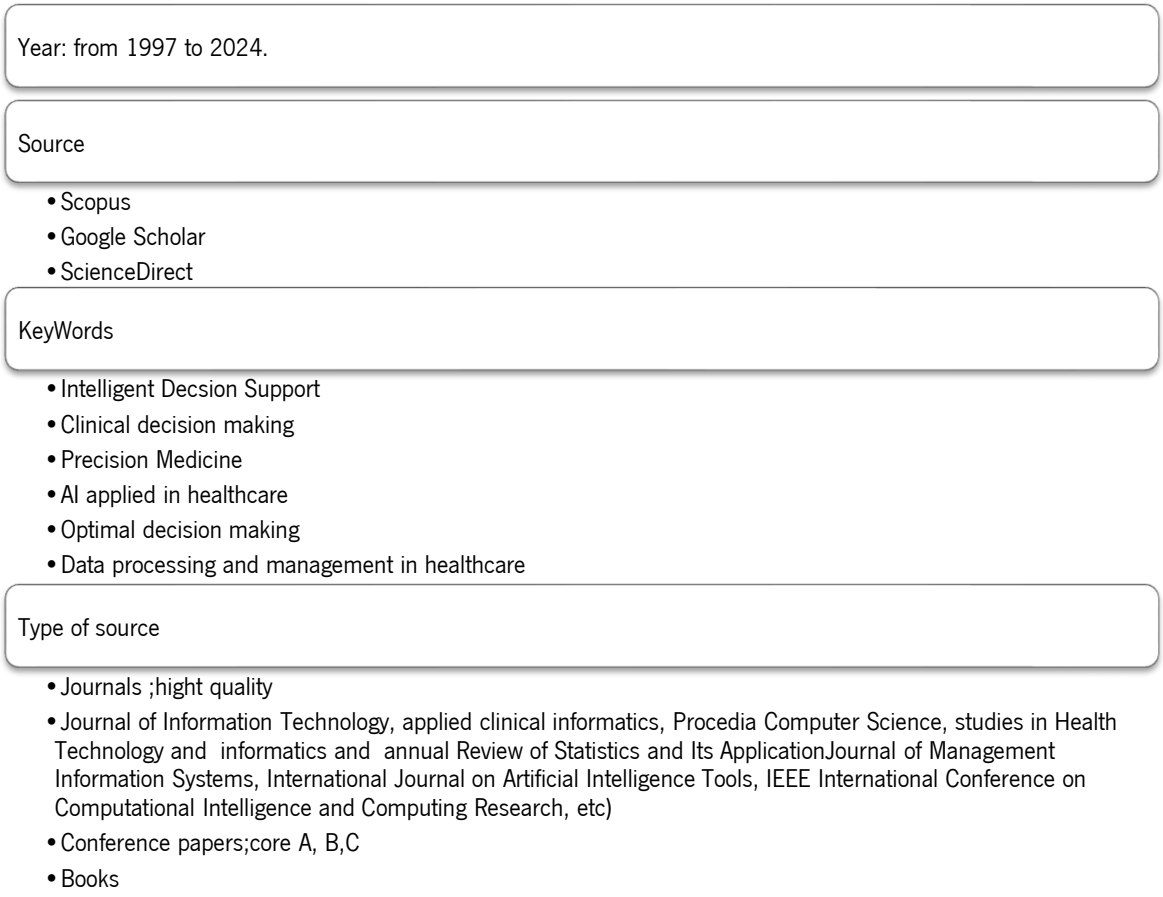


Figure 13: Literature review strategy

4. DATA-DRIVEN PSYCHOTHERAPY DECISION SUPPORT (DD_PT_DS)

By employing data mining techniques, the study seeks to unveil hidden patterns, relationships, and trends within the data, thereby advancing the comprehension of therapeutic alliance and potentially guiding therapeutic practices. Our investigation centered on two primary questions:

1. How can analytical insight be effectively applied to investigate and understand the therapy factors influencing therapeutic alliance?
2. How can diverse datasets encompassing physiological and psychological factors from both clients and therapists be utilized to accurately predict therapeutic alliance?

the pivotal components of the project's framework encompass a diverse range of physiological and psychological variables associated with both clients and therapists. The process initiates with knowledge discovery, manifested in the form of interconnected clinical events. In this phase, we conduct a comprehensive analysis, presenting insights that unveil patterns and relationships within an integrated platform.

Following knowledge discovery, predictive modeling takes center stage. Through the adoption of optimization techniques, we rank algorithms to pinpoint the most effective one. In this pivotal phase, we meticulously identify the best-performing elements, and consequently, the most significant predictors emerge. This insight contributes profoundly to the interpretation of results, influencing various aspects of tailoring treatment. By intricately considering patient data, clinical backgrounds, and the information about their therapists, we achieve a holistic approach to treatment customization.

4.1 Comprehensive Exploration of Dataset

The sample included 18 women and 8 men, aged between 18 and 52 years. The therapists involved in the study comprised 5 women and 1 man, with clinical experience spanning from 4 to 22 years.

The initial dataset comprised 24,525 rows and 45 variables, featuring 13,676 missing cells. Figure 14 provides a detailed description of each variable. Unique identifiers for clients and therapists are represented by the client's ID and therapist's ID, with six unique therapist IDs and twenty-two unique client IDs. Each client was exclusively assigned to one therapist for all therapy sessions, while therapists could attend to different clients.

The variable "Sex" was categorized as male (sex = 0) or female (sex = 1). The "Diagnosis" variable indicated whether the record was associated with anxiety (diagnostic = 0) or depression (diagnostic = 1).

The “Outcome” variable identified the result of each therapy process, categorized as poor (outcome = 0) or good (outcome = 1). Additionally, the “Termination” variable indicated whether the client experienced dropout (termination = 0) or completion (termination = 1).

The number of therapy sessions is denoted as “Session,” fragmented into epochs with a fixed period of one minute. The “Time” variable represents the period in seconds, approximately equal to sixty epochs/session numbers. The value of the “Condition” variable is constant as “1,” indicating under-treatment status, and “0,” representing the baseline before the session.

Biological variables for both clients and therapists included Heart Rate (mean, mean_baseline, SD_baseline, standardized) and EDA (mean, mean_baseline, SD_baseline, standardized). The values of EDA and HR reported in the context of _baseline, SD baseline, and standardized are repetitive for different epochs within the same session at baseline. However, the HR (mean) and EDA (mean) values differ in each epoch, referring to the in-session period. The WAI-total score which will be discussed as “WAI” in technical machine learning represents the value of TA for the client and therapist (client’s WAI, and therapist’s WAI, respectively).

The therapeutic alliance was assessed using the Portuguese version of the Working Alliance Inventory-Short Revised (WAI-SR). This inventory evaluates the quality of the therapeutic alliance across three dimensions: (1) agreement on tasks, (2) agreement on goals, and (3) development of a bond. Each dimension contributes to the total score of the WAI-SR. The WAI-SR (client) comprises 12 items rated on a 5-point Likert scale (1-5), with higher scores indicating a stronger alliance (total scores range from 12 to 60). Similarly, the therapist's WAI-SR consists of 10 items on the same Likert scale, with total scores ranging from 10 to 50.

In this study, special care was taken to prioritize privacy and ethical concerns. To address these issues, the data underwent rigorous de-identification and coding processes to ensure complete anonymity and prevent any association with real patients.

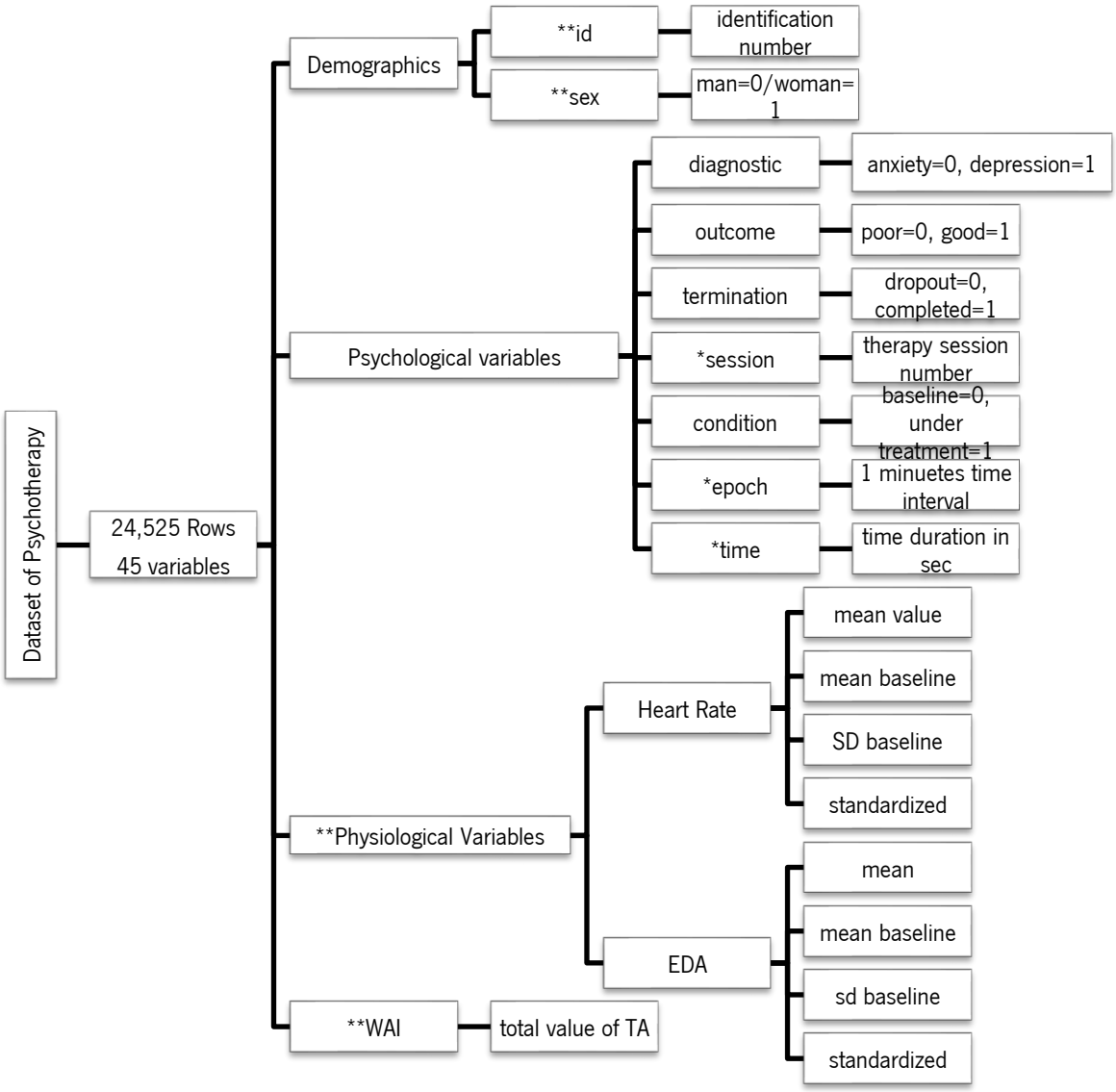


Figure 14: Data exploration (Psychotherapy)

4.2 Analytical Insight Through Interrelated Clinical Events

In this phase, to comprehend the intricate relationship within the data and extract meaningful patterns, we conducted a comprehensive analysis. We systematically observed and analyzed three significant relationships: Therapeutic Alliance (TA) categorized by session, gender, and individual IDs (specific therapist and client). Additionally, we scrutinized the distribution of the number of therapy sessions attended by clients. As a concluding step, we delved into the exploration of potential linear relationships between Heart Rate (HR) and Electrodermal Activity (EDA) with Therapeutic Alliance (TA).

4.2.1 Therapeutic Alliance by Session-Client

In Figure 15, we illustrate the correlation between the number of therapy sessions and Therapeutic Alliance (TA) for clients who completed the therapy (termination = 1). Each data point on the scatter chart represents the client ID (legend), session number (x-axis), and the corresponding TA value for that specific session (y-axis).

Examining the scatter chart reveals variations in TA values among clients in initial sessions, with some displaying high TA values, while others, after more sessions, exhibit lower TA values. For instance, client ID 20 had a TA value of 44 in session 2, which decreased to 38 by session 18. Notably, client ID 23 consistently demonstrated the maximum TA value (60) across all therapy sessions, while client ID 13 exhibited an upward trend in WAI with an increasing number of sessions.

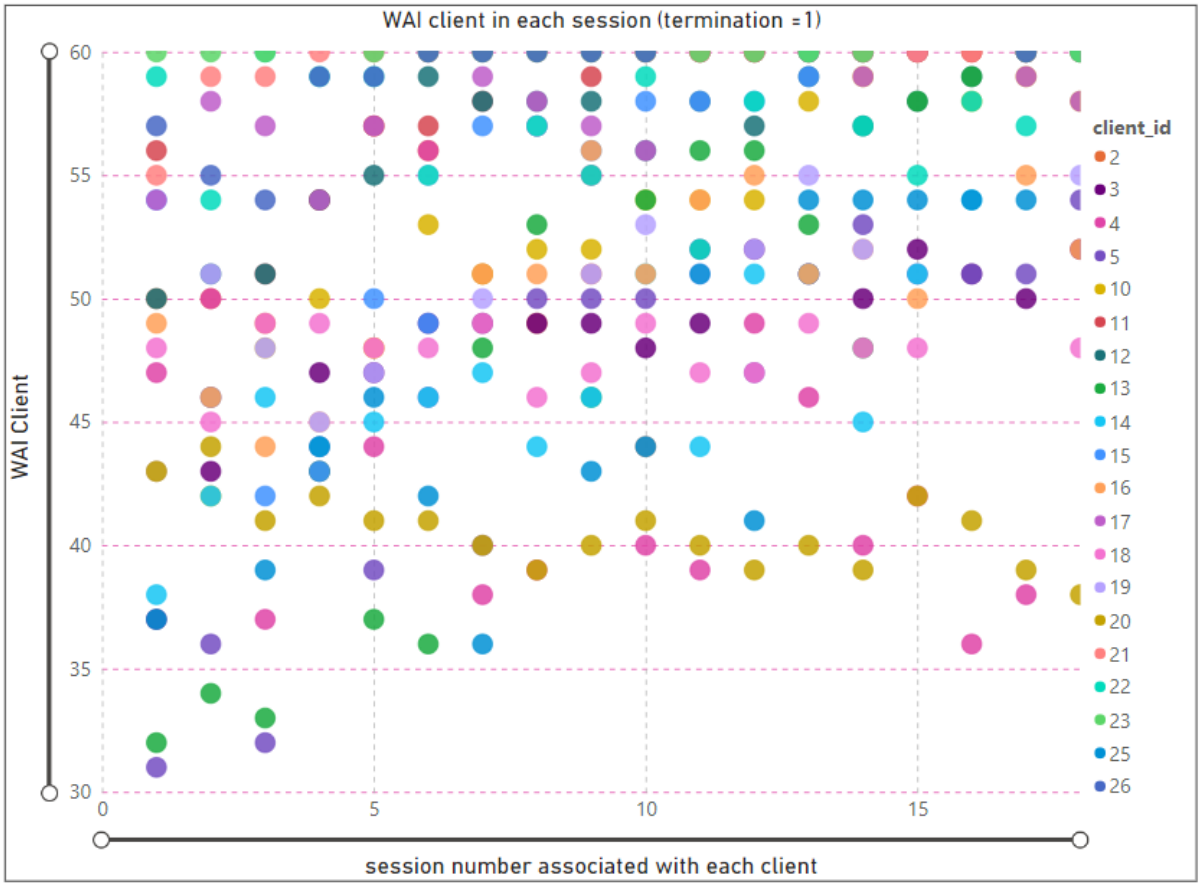


Figure 15: Relationship between session and WAI score (TA) for clients; termination=1

Figure 16 presents a parallel analysis focusing on clients who terminated therapy prematurely (termination = 0). As depicted in the scatter chart, only two clients opted to discontinue therapy – one with 8 sessions (ID 7) and the other with 11 sessions (ID 9).

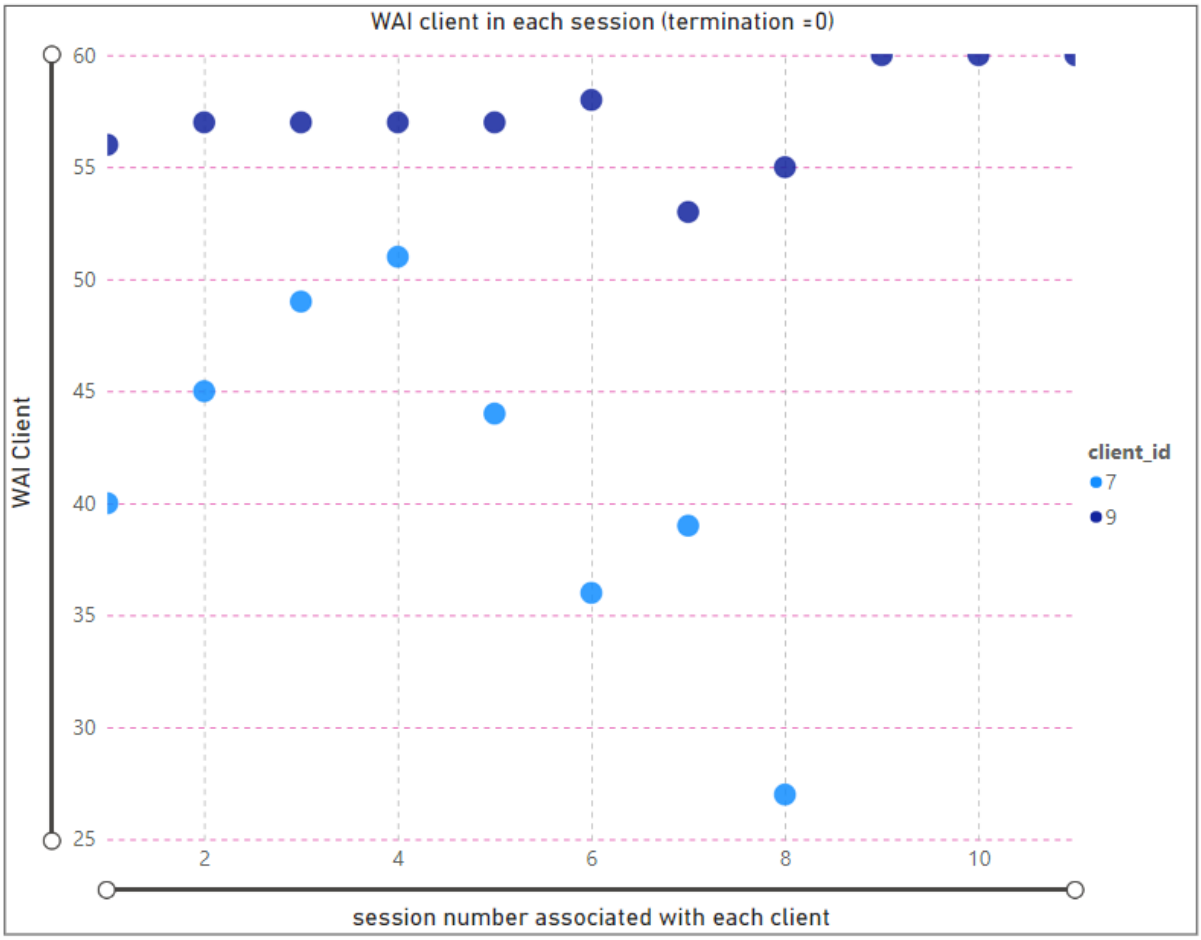


Figure 16: Relationship between session and WAI score (TA) for clients; termination=0

4.2.2 Therapeutic Alliance by Session-Therapist

To explore the correlation between the therapist's Therapeutic Alliance (TA) and the number of therapy sessions, we factored in the client ID, recognizing that each therapist may have multiple clients. Figure 17 illustrates a scatter chart featuring the therapist's ID, the count of clients, and the respective session numbers.

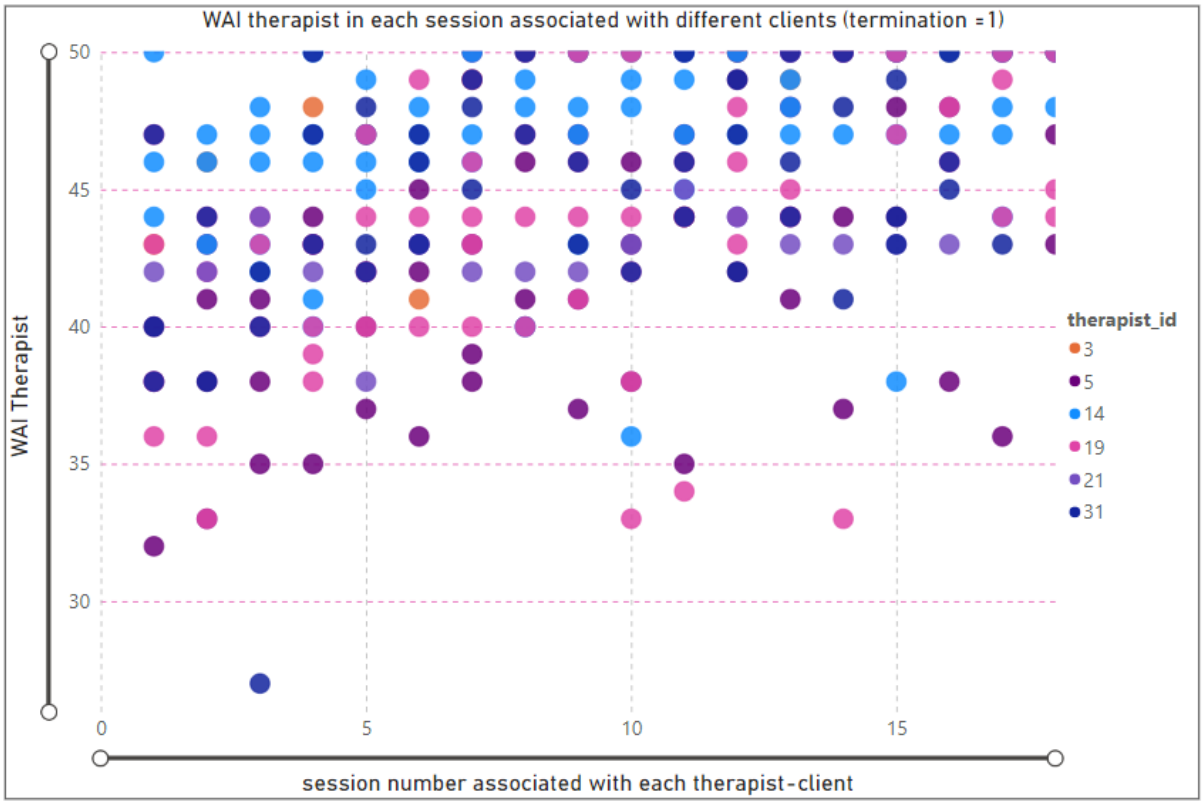


Figure 17: Relationship between session and WAI score (TA) for therapists; termination=1

As depicted in Figure 18, two therapists (ID = 14, 31) have conducted sessions with clients who eventually dropped out.

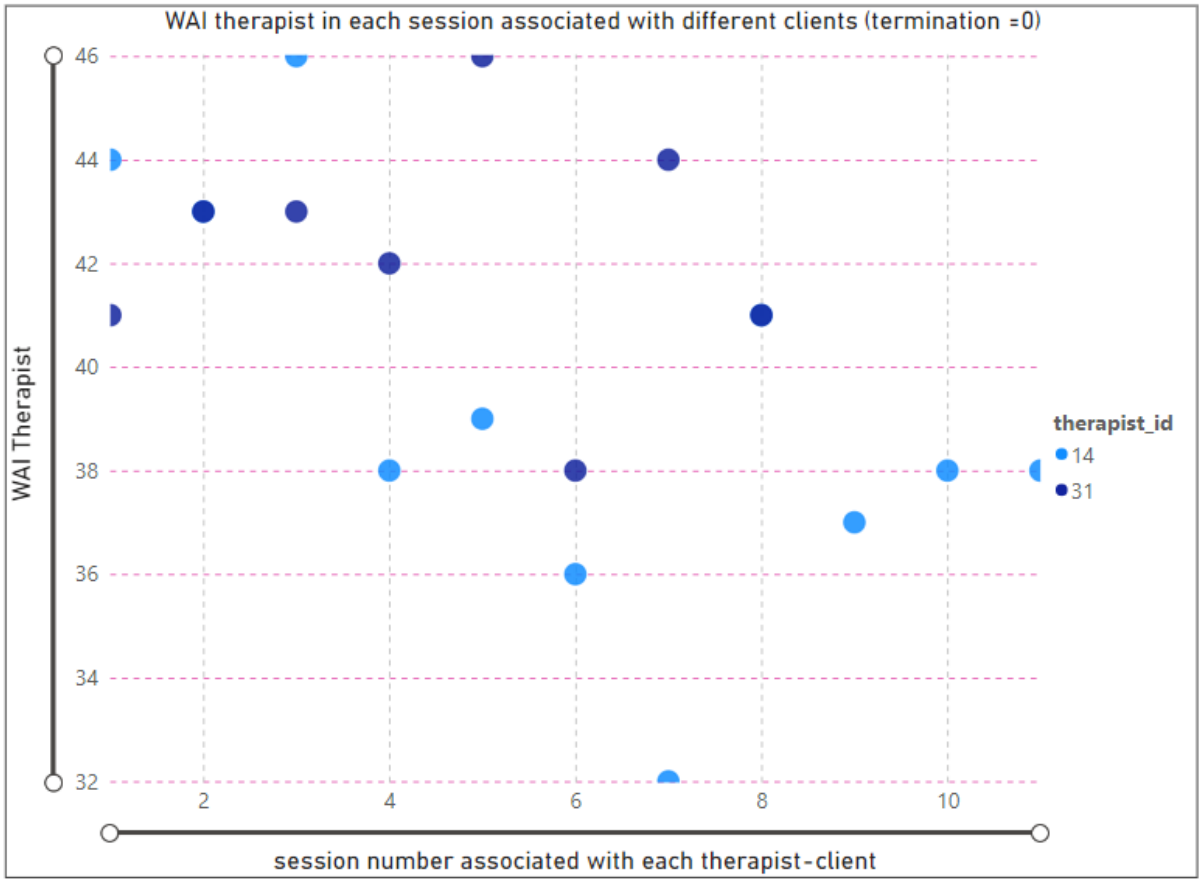


Figure 18: Relationship between session and WAI score (TA) for therapists; termination=0

4.2.3 Therapeutic Alliance by Sex-Client

In Figure 19, we present the average value and standard deviation of Therapeutic Alliance (TA) categorized by the client's sex ("client_sex" equals zero for male clients and one for female clients). Through this analysis, we observed that out of the twenty-two clients, seven were male (client ID = 3, 5, 12, 19, 20, 22, 23), and the remaining fifteen were female (client ID = 2, 4, 7, 9, 10, 11, 13, 14, 15, 16, 17, 18, 21, 25, 26). In Figure 6, the mean TA value for female clients was 51.74, while for male clients, it was 51.60. Additionally, the standard deviation of TA for female clients equaled 7.38, and for male clients, it was 7.10.

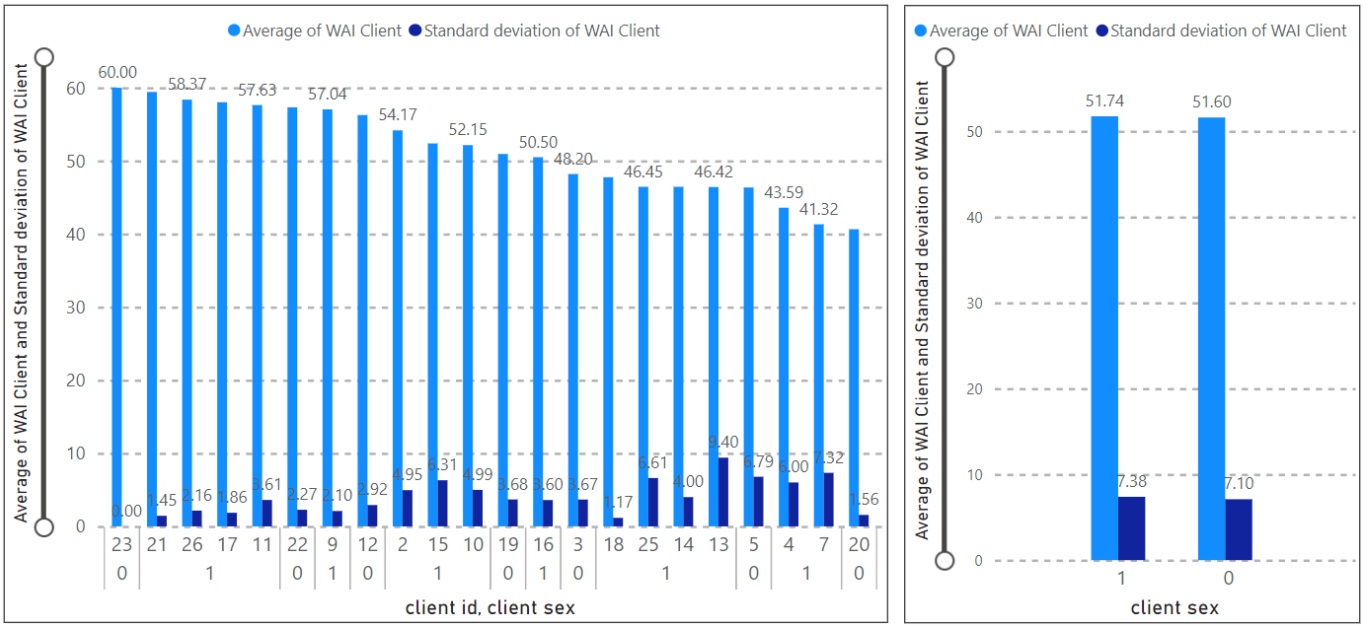


Figure 19: Average of WAI score (TA) by sex – client

4.2.4 Therapeutic Alliance by Sex-Therapist

In the TA by sex analysis (Figure 20), it's notable that only therapist ID=19 is male (therapist sex = 0), and the average TA value associated with this therapist is 43.70. Contrarily, the average TA value for other female therapists (therapist id = 3, 5, 14, 21, 31) stands at 45.07. Hence, female therapists exhibit a higher TA value compared to their male counterparts. Referring to Figure 6, the standard deviation of TA for therapist ID equal to 19 is 4.33, whereas for female therapists, it is 4.17.

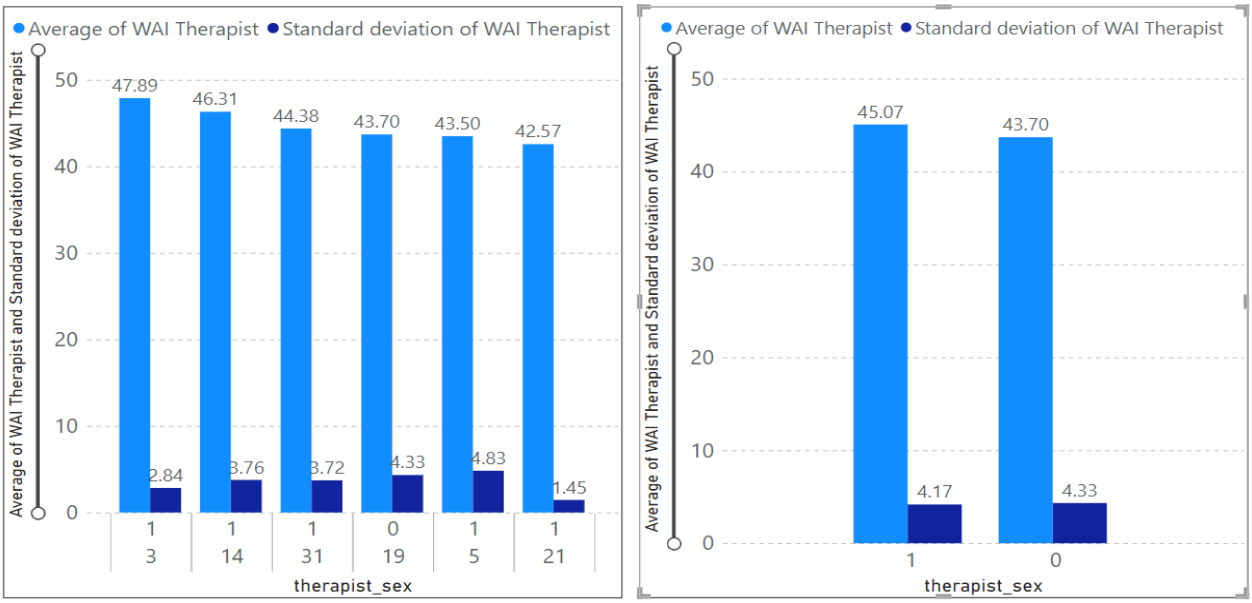


Figure 20: Average WAI score (TA) by sex therapist

4.2.5 Therapeutic Alliance for Specific Client

In this analysis (Figure 21), we've showcased the average scores recorded by client IDs. Figure 9 highlights that the lowest average value of Therapeutic Alliance (TA) was 40.64, attributed to client ID number 20, while the highest score, reaching 60.00, was achieved by the client with ID number 23.

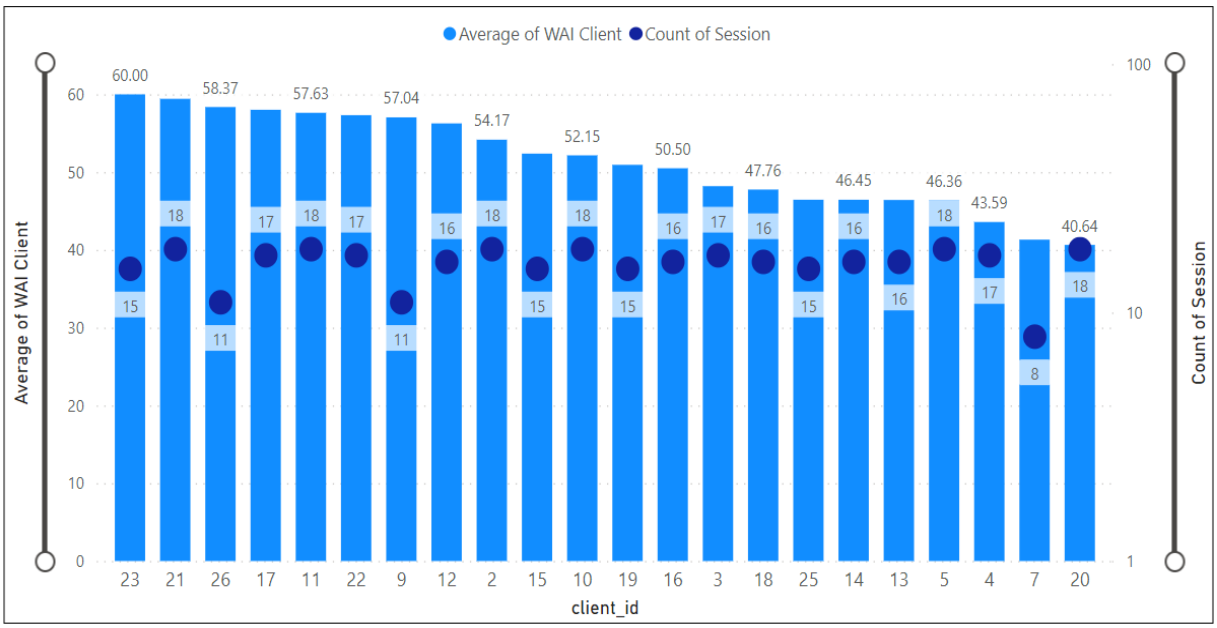


Figure 21: Average client's WAI score (TA) by client ID

4.2.6 Therapeutic Alliance for Specific Therapists

In a parallel manner, therapist ID number 3 attained the highest average value of the Working Alliance Inventory (WAI) score, reaching 47.89, while the therapist with ID number 21 recorded the minimum average at 42.57. Given that each therapist oversees multiple clients, Figure 22 details the association of therapist IDs with specific clients. To elaborate, therapist ID 14, paired with client ID 21, achieved an average Therapeutic Alliance (TA) value of 48.55, whereas therapist ID 5, aligned with client ID 20, reported the lowest average TA value at 37.56.

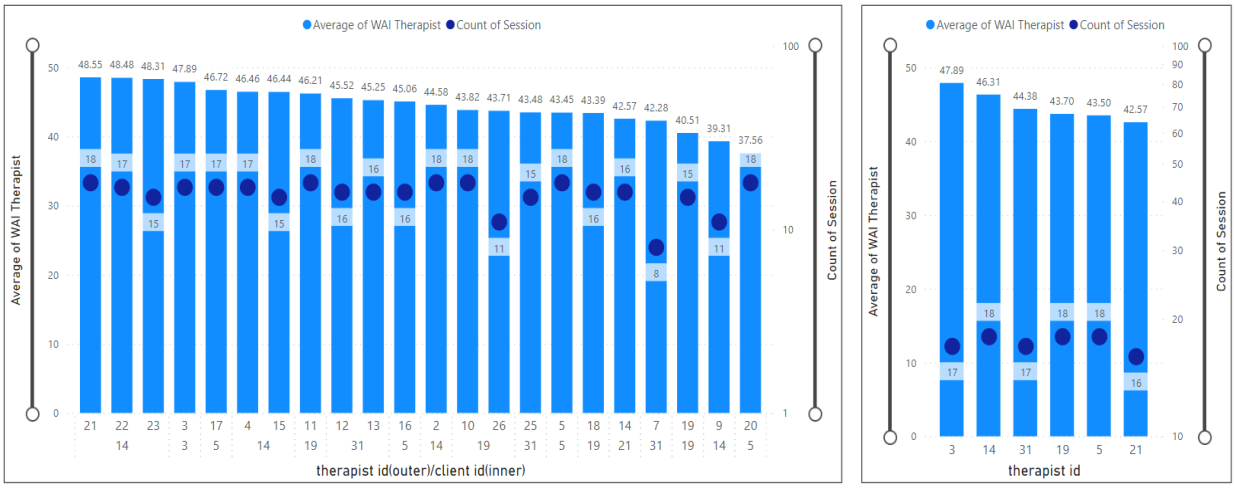


Figure 22: Average therapist's WAI score (TA) by therapist ID

4.2.7 Therapy Session for Client

The bar chart depicted in Figure 23 provides a comprehensive analysis of the total therapy sessions attended by clients. On the x-axis, twenty-two client IDs are represented, while the y-axis illustrates the total therapy sessions for each client. Additionally, the legend (termination) indicates whether the client dropped out.

Among the clients, those with IDs 11, 21, 2, 5, 10, and 20 exhibited the maximum number of therapy sessions, each totaling 18. Conversely, clients with ID number 26 attended eleven therapy sessions. Notably, clients ID 9 and 7 were categorized as dropout clients (termination = 0) and attended 11 and 8 sessions, respectively.

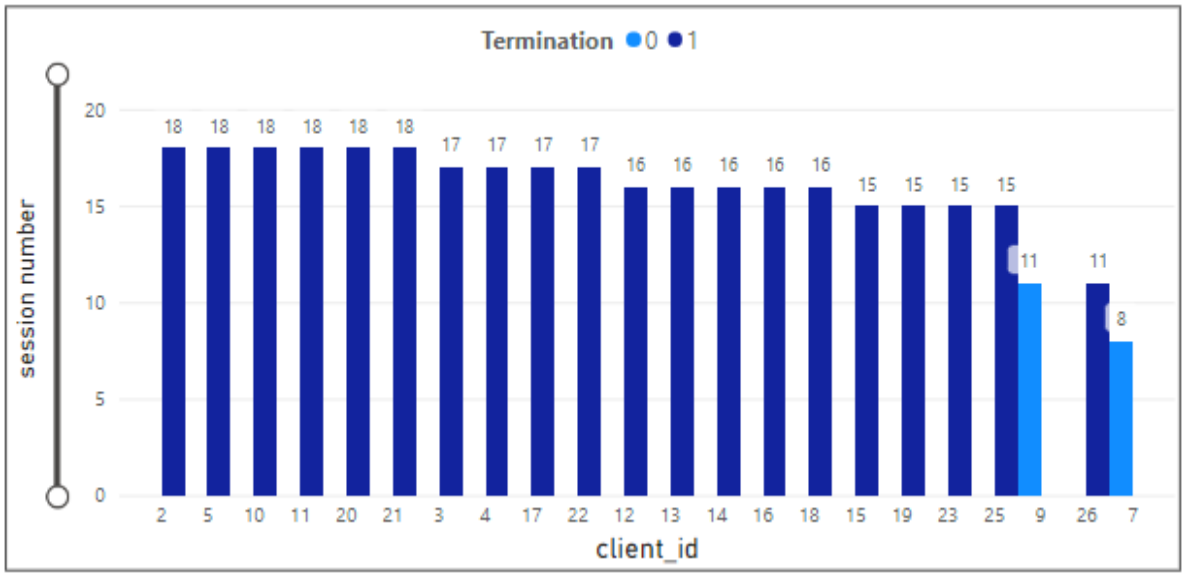


Figure 23: Therapy sessions attended by clients

4.2.8 Therapy Session for Therapist

Figure 24 illustrates a parallel analysis for therapists. On the x-axis, therapist IDs are paired with the respective client IDs. The most extensive number of therapy sessions (18) was linked with therapist-client pairs 14-2/21, 5-5/20, and 19-10/11.

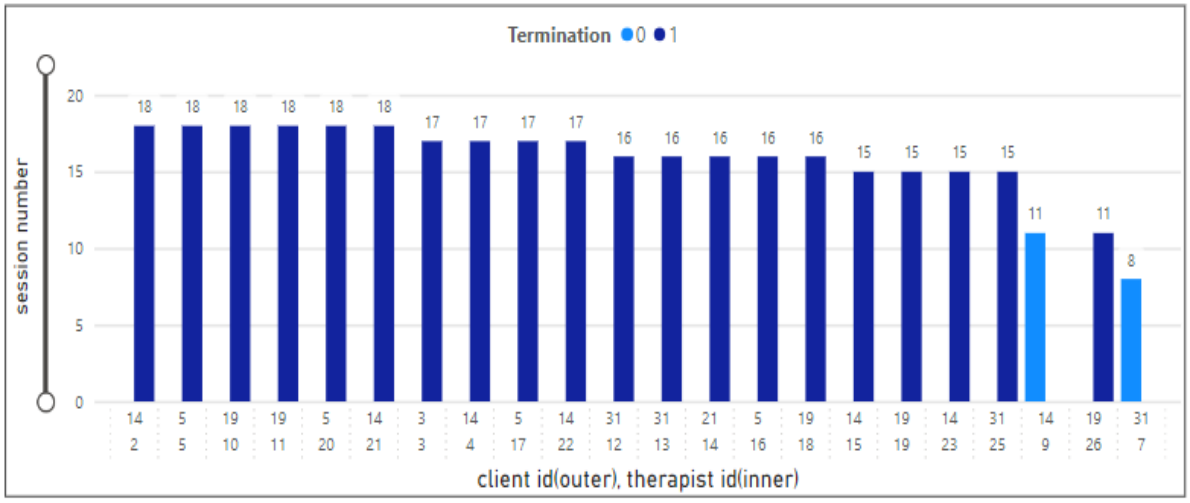


Figure 24: Therapy sessions attended by therapists

4.2.9 Investigating the Relation of Heart Rate on Therapeutic Alliance with Clients

In Figure 25, we delve into the correlation between the Therapeutic Alliance (TA) value (x-axis) and the average Heart Rate (HR) value (y-axis) for each client (termination equals one). The legend vividly identifies

each client ID with colorful representation, while session numbers are thoughtfully displayed as data labels. Additionally, the size of each data point, represented as a triangle, corresponds to the session number – initial sessions are subtly presented with smaller triangles for clearer interpretation.

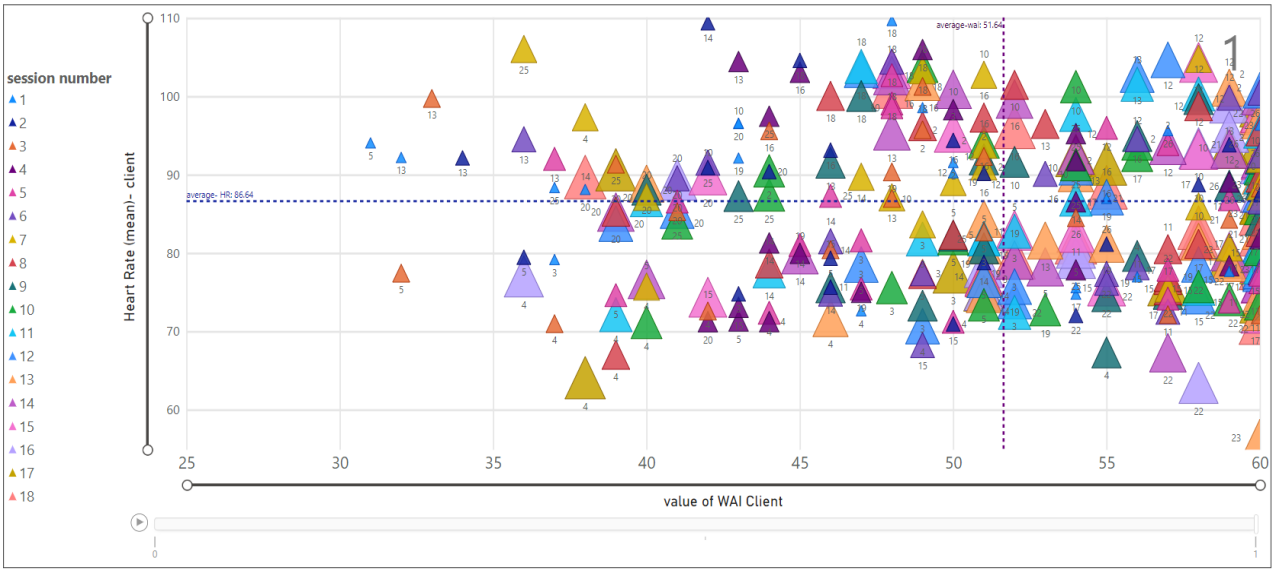


Figure 25: Relationship between HR and WAI score (TA) for clients; termination=1

Similarly, in Figure 26, we explore the connection between Heart Rate (HR) mean and Therapeutic Alliance (TA) for clients with termination status equal to zero. The chart highlights instances where clients exhibited HR values above the average (86.64). Specifically, for client ID = 9, the TA value surpasses the average (51.64), while for client ID = 7, it falls below 51.64.

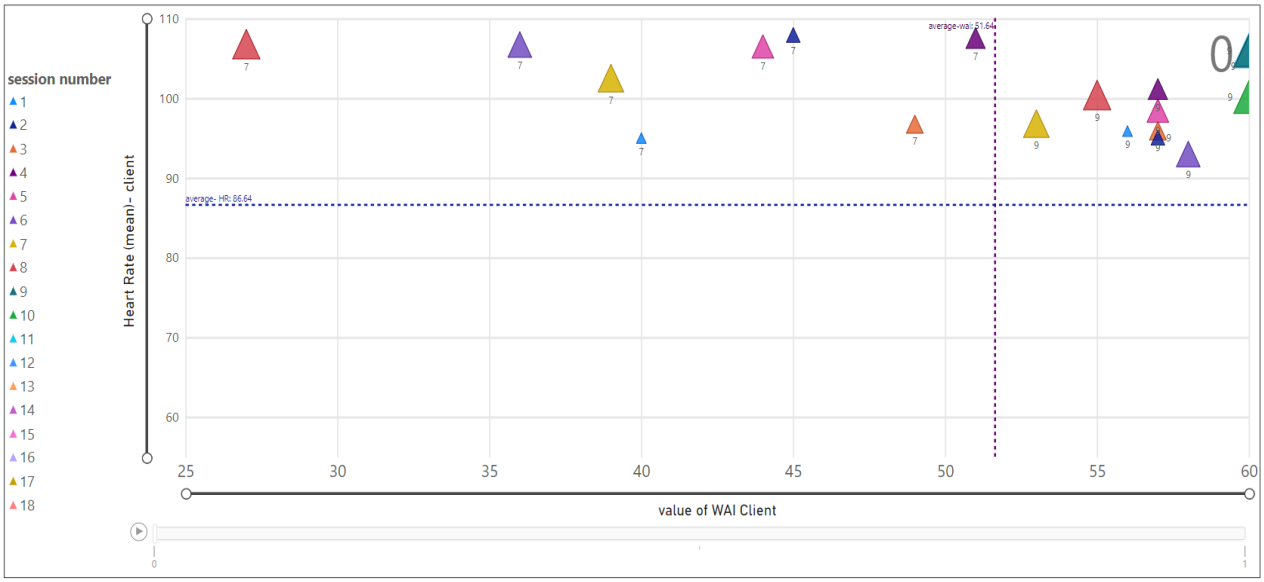


Figure 26: Relationship between HR and WAI score (TA) for clients; termination=0

4.2.10 Investigating the Relation of Heart Rate of Therapeutic Alliance - Therapist

In Figure 27, we examine the dynamic relationship between the Therapeutic Alliance (TA) value (x-axis) and the average Heart Rate (HR) value (y-axis) for each therapist ID, considering the diverse clients each therapist oversees. The highlighted data points spotlight the average HR (99.64) and TA (27) in session 3 for a therapist with ID = 31, associated with a client with ID = 13. Additionally, a vertical line represents the average TA value for therapists (44.82), while a horizontal line indicates the average HR value (84.28). This visual analysis provides insights into the interplay between therapist-specific dynamics and the associated TA and HR values.

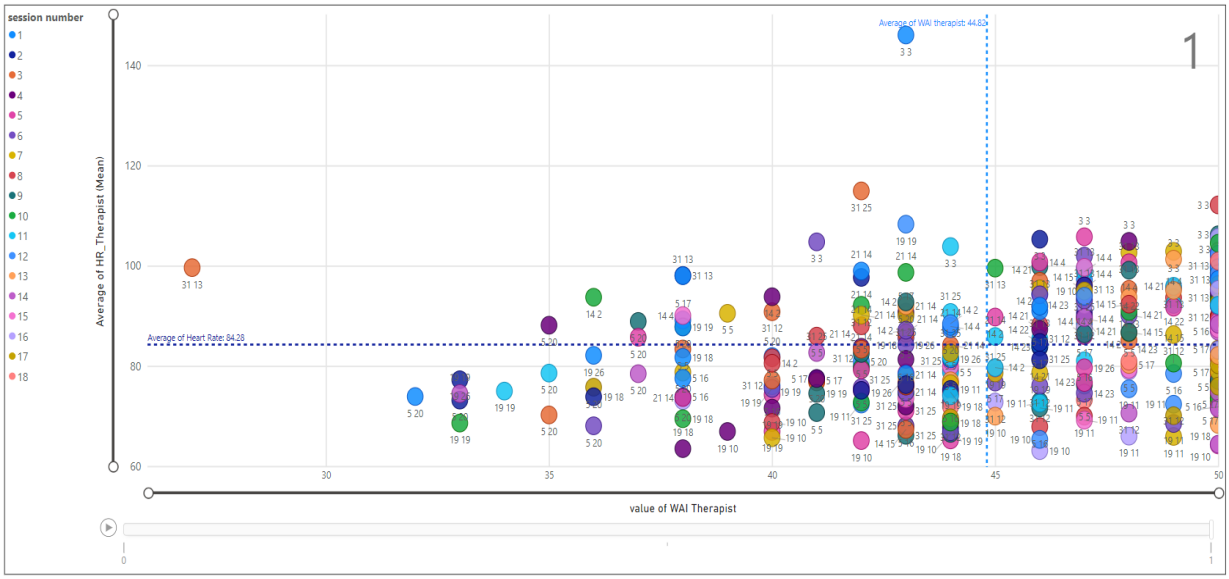


Figure 27: Relationship between HR and WAI score (TA) for therapists; termination=1

The identical analysis is presented in Figure 28, this time focusing on cases where the termination status is zero. Two distinct scenarios emerge: Therapist ID = 14 is paired with client ID = 9, and Therapist ID = 31 is associated with client ID = 7. The majority of data points cluster within a Therapeutic Alliance (TA) value lower than the average (44.82). Specifically, the data points linked to Therapist ID = 14 and client ID = 9 in session 1 reveal an average Heart Rate (HR) value of 98.04 and a TA value of 44. This visual exploration sheds light on nuances in therapist-client dynamics when termination is absent.

Figure 28: Relationship between HR and WAI score (TA) for therapists; termination=0

4.2.11 Investigating the Relation of EDA Of the Therapeutic Alliance - Client.

In Figure 29, we explored the correlation between Therapeutic Alliance (TA) and the average Electrodermal Activity (EDA) for clients with a termination status equal to one. The x-axis denotes the TA value, while the y-axis represents the average EDA for each client across therapy sessions. The legend introduces session numbers through colorful triangles, with the smallest triangle denoting session one and the largest corresponding to session eighteen. Additionally, a vertical average line for TA at 51.64 and a horizontal average line for EDA at 5.16 provide reference points.

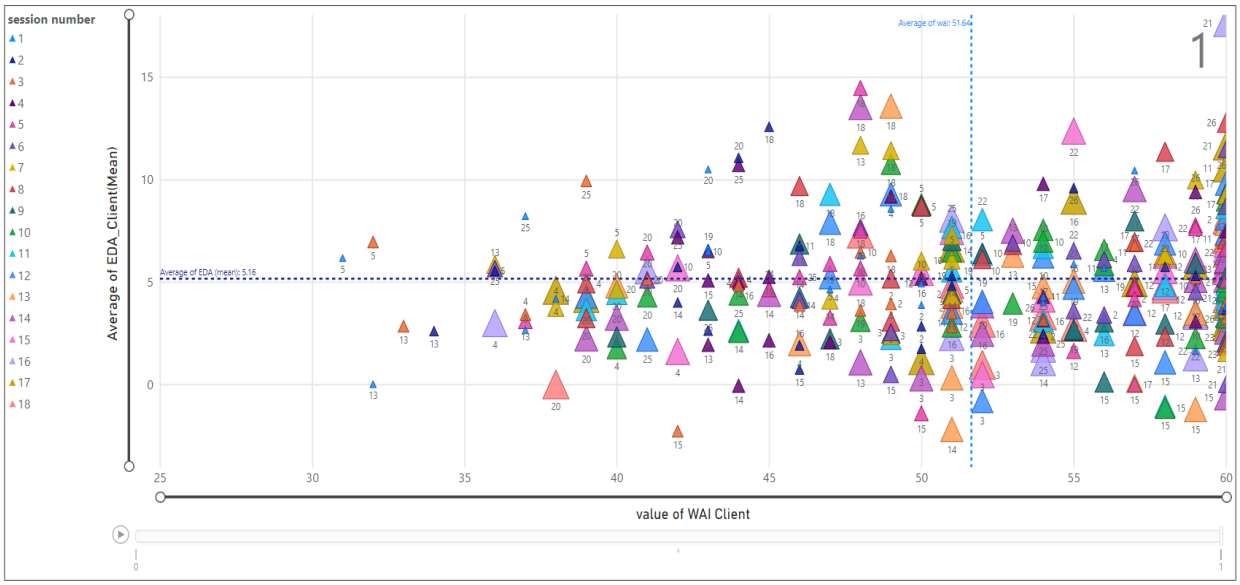


Figure 29: Relationship between EDA and WAI score (TA) for clients; termination=1

In Figure 30, we delved into the connection between Therapeutic Alliance (TA) and the average Electrodermal Activity (EDA) for clients with termination status set to one. The x-axis signifies the TA value, while the y-axis illustrates the average EDA for each client across therapy sessions. Session numbers are elegantly introduced through colorful triangles in the legend, where the smallest triangle signifies session one, and the largest corresponds to session eighteen. To offer a frame of reference, a vertical average line for TA at 51.64 and a horizontal average line for EDA at 5.16 are incorporated, enriching the visual interpretation of the correlation dynamics.

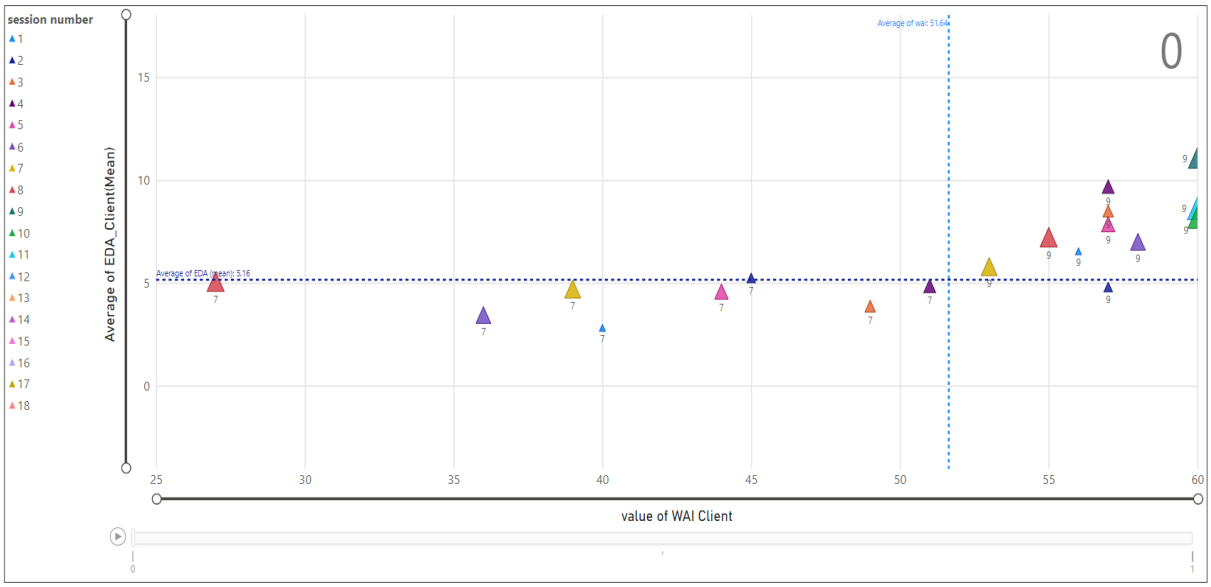


Figure 30: Relationship between EDA and WAI score (TA) for clients; termination=0

4.2.12 Investigating The Relation of EDA of the Therapeutic Alliance - Therapist

Exploring the association between Electrodermal Activity (EDA) and Therapeutic Alliance (TA) for therapists is visualized in Figure 31. The highlighted data point spotlights therapist ID = 21, assigned to client ID = 14 in session 16, with an average EDA value of 13.65 and a corresponding TA value of 43. This focused illustration provides insight into the interplay between EDA and TA for therapists across specific therapy sessions.

.

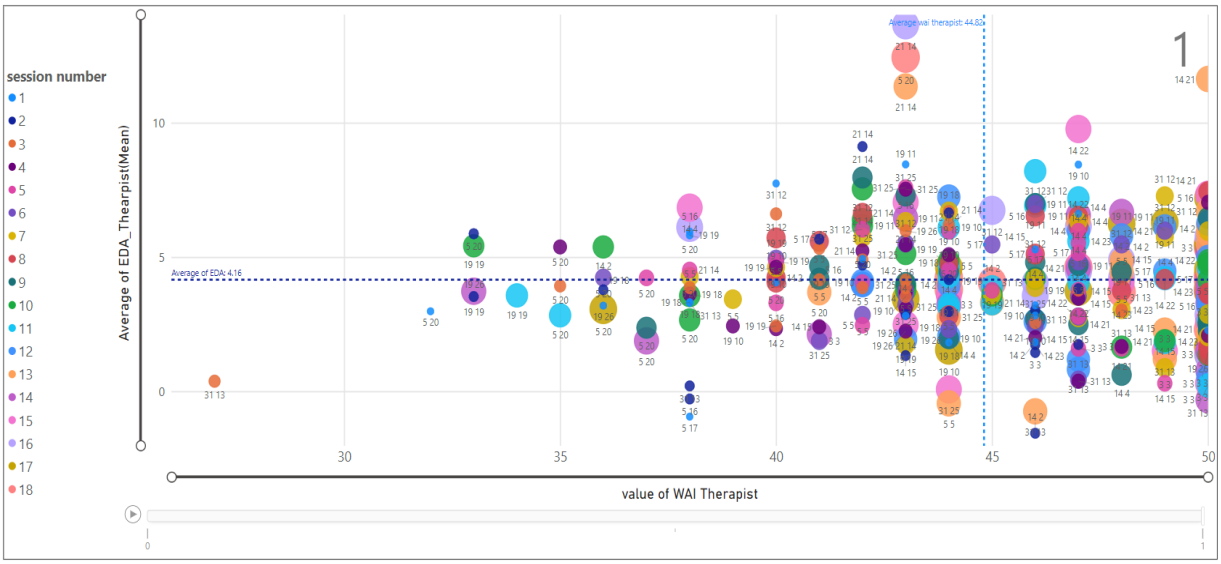


Figure 31: Relationship between EDA and WAI score (TA) for therapists; termination=1

In Figure 32, the scatter chart replicates the analysis, this time with a termination status equal to 0. A specific data point is highlighted, corresponding to session 10 for therapist ID = 14 and client ID = 9. In this instance, the therapist's average Electrodermal Activity (EDA) is 5.44, while the Therapeutic Alliance (TA) value stands at 38. This focused example provides a glimpse into the dynamics of EDA and TA in sessions where termination did not occur.

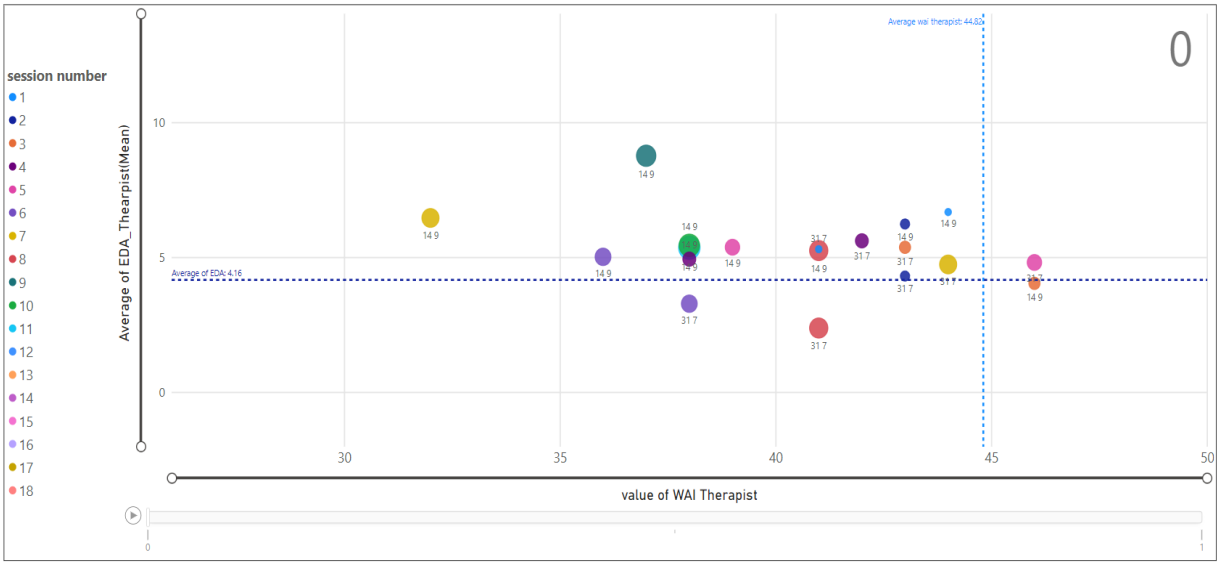


Figure 32: Relationship between EDA and WAI score (TA) for therapists; termination=0

4.2.13 Results and Discussion

Performing analytical insights contributes to tailoring treatment and therapy plans for each client by understanding the intricate links between indicators and the relationship between physiological and psychological factors to TA. Such analysis and data observation are vital for monitoring client and therapist

status, enabling early detection, and enhancing patient engagement, ultimately leading to improved outcomes. Additionally, conducting statistical analysis provides a better understanding of data quality and boundaries, while visualization techniques offer insights into variable relationships and pattern identification, aligning with study objectives. Data processing, including grouping and aggregation at the session level, along with integration based on therapist-client session linkage, supports this investigation by preparing effective data for modeling. This facilitates and advances predictive analysis for the next phase. Figure 33 provides a summary of the contribution of this phase (Analytical Insight) to Intelligent Decision Support Systems for Precision Medicine (IDSS4PM) from a data-driven perspective. It also outlines its application in precision medicine, bearing in mind that the project is an exploratory investigation.

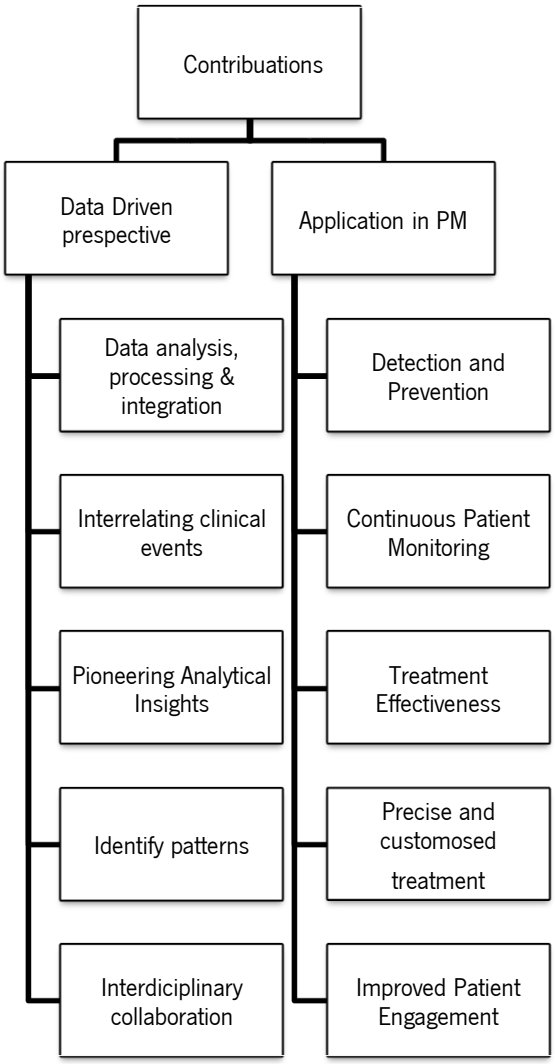


Figure 33:Contributions IDSS4PM; Analytical Insights

4.3 Predictive Modeling for Therapeutic Alliance

4.3.1 Comprehensive Data Preparation For Modelling

This phase encompasses a series of tasks such as feature selection, feature construction, data cleaning, and formatting.

In the data preparation for modeling, we systematically eliminated records with missing values. Employing Pearson correlation analysis (denoted as "Pearson's r"), we quantified the linear relationship between variables to discern correlation coefficients among features. The results of the correlation analysis are presented in Figure 34. This step is pivotal for selecting influential predictors and excluding features with a negative impact on modeling results. Recognizing that a high correlation between features is redundant and doesn't enhance model accuracy, we excluded those with less influence on predicting the target variable (WAI-SR score). The constant feature "condition" with a value of 1 was also removed. The variable "ID," deemed non-contributory and potentially causing overfitting, was likewise excluded. Additionally, noting the substantial skewness (Y1 = 20.54) in Client_HR (Standardized), we omitted this feature from modeling.

To enhance model efficiency, we adjusted variable types to the most suitable ones, categorizing some variables and converting others to float type. Figure 34 illustrates the correlation matrix. By considering Pearson correlation coefficients with a range of 0.1 <= x <= -0.1, we identified effective indicators for predicting the WAI value. Furthermore, a correlation level of 0.4 <= x <= -0.4 was employed to examine relationships among variables. This comprehensive approach ensures a refined dataset, optimized for subsequent modeling endeavours.

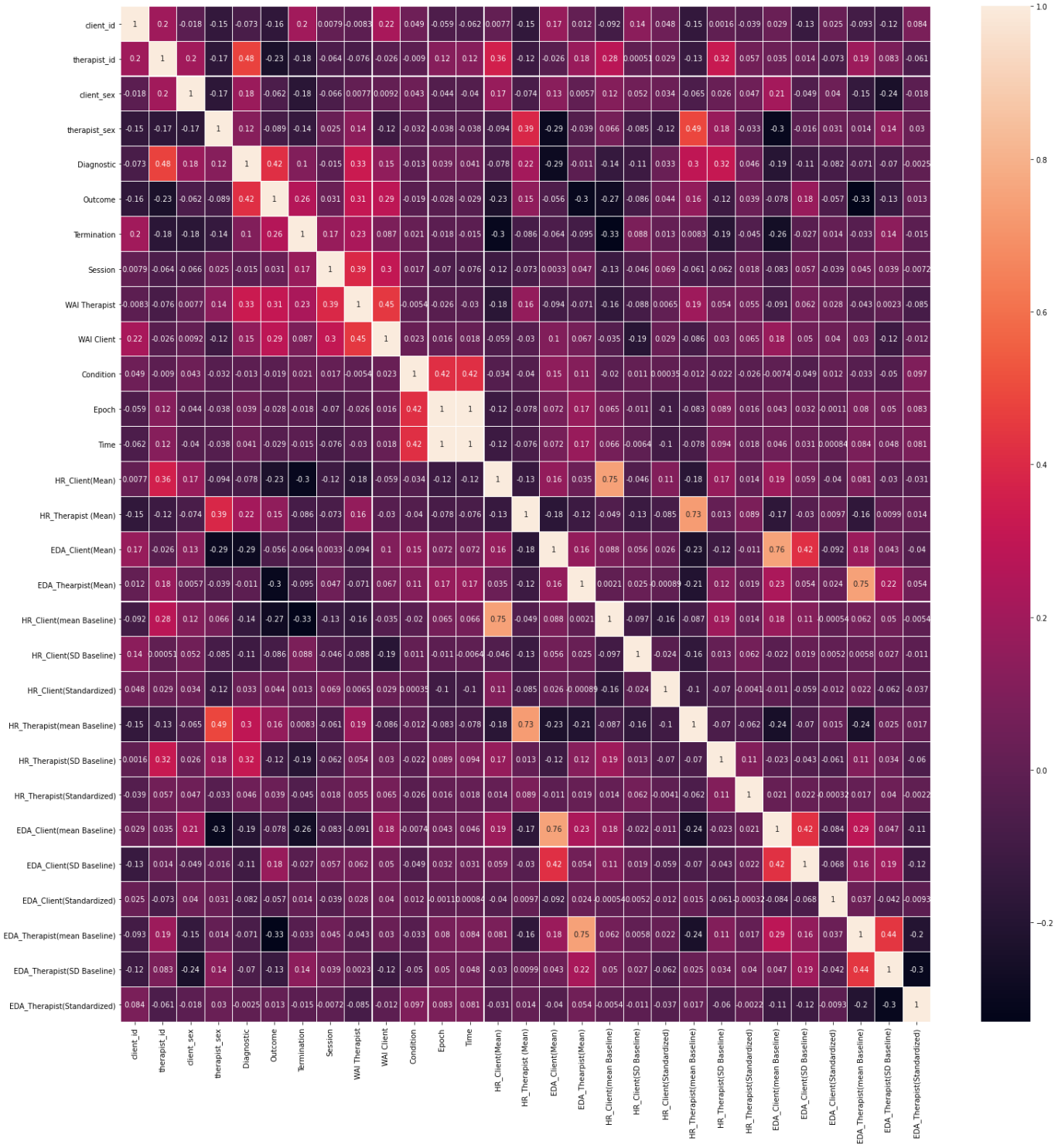


Figure 34: Pearson Correlation Coefficients Matrix

Table 3 highlights that "WAI_Therapist" stands out as the most influential indicator with a positive impact on predicting the target, boasting a correlation coefficient score of 0.44. Additionally, variables such as "Session," "Outcome," "HR-Client(SD_Baseline)," "EDA_Client(mean_Baseline)," "Diagnostic,"

"EDA_Therapist(SD_Baseline)," "Therapist_sex," and "EDA_Client(mean)" were identified as other effective predictors for "WAI-Client," presented as correlated variables to "WAI_Client."

Moreover, "WAI_Client" emerged as the most potent variable in predicting the value of "WAI_Therapist," holding a correlation of 0.44. Several other indicators, including "Session," "Diagnostic," "Outcome," "Termination," "HR_Therapist(mean_Baseline)," "HR_Client(mean)," "HR_Client(mean_Baseline)," and "HR_Therapist(mean)," were acknowledged as influential factors.

Interestingly, EDA did not exhibit a significant association with "WAI_Therapist," while Heart Rate emerged as a crucial variable. As per Table 3, the therapist's HR positively influenced WAI, while the client's HR showed an inverse relation to "WAI_Therapist." Furthermore, "Diagnostic" and "Termination" exhibited a stronger association with "WAI_Therapist" than with "WAI_Client."

Table 3: Pearson Correlation Between TA (WAI Scores) And Predictors

#	Predictors for WAI_Client	score	Predictors for WAI_Therapist	score
1	WAI therapist	0.45	WAI-Client	0.45
2	Session	0.30	Session	0.39
3	Outcome	0.28	Diagnostic	0.33
4	HR-Client(SD_Baseline)	(-) 0.18	Outcome	031
5	EDA_Client(mean_Baseline)	0.17	Termination	0.23
6	Diagnostic	0.15	HR_Therapist(mean_Baseline)	0.19
7	EDA_Therapist(SD_Baseline)	(-) 0.11	HR_Client(mean)	(-) 0.18
8	Therapist_sex	(-) 0.12	HR_Client(mean_Baseline)	(-) 0.16
9	EDA_Client(mean)	0.10	HR_Therapist(mean)	0.16

In the process of selecting the final predictors, we meticulously examined the correlation coefficients among the predictors. The details of the included and excluded variables for predicting "WAI_Client" are presented in Table 4. Notably, due to the robust correlation observed between "session" and "WAI_Therapist" (0.39), and between "Diagnostic" and "Outcome" (0.42), these variables were excluded from the modeling process. Additionally, considering the strong correlation (0.75) between "EDA_Client(mean)" and "EDA_Client(mean_Baseline)," the former was also excluded.

As highlighted in the data understanding phase, where the values of all EDA and HR mean_Baseline and SD_Baseline were repeated for each epoch within the same session number, we aggregated the data based on the session number to predict "WAI_Client." This step ensures a refined selection of predictors for an effective and accurate modeling approach.

Table 4: Pearson Correlation Among" WAI_Client" Predictors

#	Predictors for WAI_Client	Correlation Coefficient Score	Included	Description
1	WA therapist	0.44	YES	
2	Session	0.30	NO	high correlation with "WAI-Therapist" (0.39)
3	Outcome	0.28	YES	
4	HR-Client(SD_Baseline)	(-) 0.18	YES	repeated value in each epoch for each session number
5	EDA_Client(mean_Baseline)	0.17	YES	repeated value in each epoch for each session number
6	Diagnostic	0.15	NO	high correlation with "Outcome" (0.42)
7	EDA_Therapist(SD_Baseline)	(-) 0.11	YES	repeated value in each epoch for each session number
8	Therapist_sex	(-) 0.12	YES	
9	EDA_Client(mean)	0.10	NO	high correlation with "EDA_Client(mean_Baseline)" (0.75)

Upon scrutinizing the predictors of "WAI_Therapist" as outlined in Table 5, it became apparent that "Outcome" exhibited a robust correlation with "Diagnostic" (0.42) and was consequently excluded from the modeling process. Moreover, in the decision between "HR_Therapist(mean_Baseline)" and "HR_Therapist(mean)," the latter was omitted due to its lesser influence on "WAI_Therapist" compared to the mean_baseline value.

Regarding "HR_Client(mean_Baseline)," its noteworthy correlation score of 0.75 with "HR_Client(mean)" prompted a careful evaluation. In this context, we opted for the baseline value, considering its stronger correlation and impact on "WAI_Therapist." It's crucial to highlight that "HR_Therapist(mean_Baseline)" comprised repeated values for each epoch within the same session, necessitating the aggregation of data at the session level for accuracy.

Lastly, the variable "Therapist_Sex," exhibiting a Pearson correlation score of 0.49 with "HR_Therapist(mean_Baseline)," was excluded from the modeling process. This decision ensures a streamlined and effective predictor selection process for robust modeling outcomes.

Table 5: Pearson Correlation Among" WAI_Therapist" Predictors

#	Predictors for WAI_Therapist	Correlation Coefficient Score	Included	Description
1	WAI-Client	0.44	YES	
2	Session	0.39	NO	Repeated in each epoch.
3	Diagnostic	0.33	YES	
4	Outcome	031	NO	High correlation with "Diagnostic" (0.42)

5	Termination	0.23	YES	
6	HR_Therapist(mean_Baseline)	0.19	YES	Repeated value in each epoch for each session number
7	HR_Client(mean)	(-) 0.18	NO	High correlation with “HR_Client(mean_Baseline)” (0.75)
8	HR_Client(mean_Baseline)	(-) 0.16	YES	Choose to aggregate repeated value in each epoch for each session number
9	HR_Therapist(mean)	0.16	NO	High correlation with “HR_Therapist(mean_Baseline)” (0.73)
10	Therapist_Sex	0.13	NO	High correlation with “HR_Therapist(mean_Baseline)” (0.49)

Leveraging the results of correlation analysis, we strategically picked the most impactful variables for predicting the target which are listed in Table 6. To predict "WAI_Client," the top six highly correlated predictors were chosen, including "WAI_Therapist," "Outcome," "Therapist_sex," "HR-Client(SD_Baseline)," "EDA_Client(mean_Baseline)," and "EDA_Therapist(SD_Baseline)." Similarly, for predicting "WAI_Therapist," we selected key predictors such as "WAI_Client," "Diagnostic", "Termination","HR_Therapist(mean_Baseline)," and "HR Client(mean_baseline)."

Table 6: List of selected predictors

Target		Predictors
WAI_Client	1	WAI_Therapist
	2	Outcome
	3	Therapist_sex
	4	HR-Client(SD_Baseline
	5	EDA_Client(mean_Baseline)
	6	EDA_Therapist(SD_Baseline)
WAI_Therapist	1	WAI_Client
	2	Diagnostic
	3	Termination
	4	HR_Therapist(mean_Baseline
	5	HR Client(mean_baseline)

4.3.2 Modelling And Evaluation

In this phase, we tailored our approach based on the problem type and target variable, employing various machine learning (ML) techniques.

In addressing the research question on predicting therapeutic alliance using physiological and psychological factors from both clients and therapists, we utilized a range of popular machine learning methods including Artificial Neural Network (ANN), Decision Tree (DT), Random Forest (RF), Linear Regression (LR), and Support Vector Regression (SVR). This approach enabled us to explore the relationships between independent and dependent variables.

Evaluation techniques involved Nested cross-validation (CV=5) and GridSearchCV for model ranking and hyper-parameter tuning. Nested cross-validation, incorporating both inner and outer loops, facilitated unbiased performance estimates on unseen data while considering hyper-parameter tuning and model evaluation, mitigating overfitting risks. Although computationally demanding, we opted for CV=5 folds for both inner and outer folds to balance reliability and computational efficiency.

Performance evaluation metrics included Mean Squared Error (MSE) and Coefficient of Determination (R2). A lower MSE indicates better model performance, while a higher R2 demonstrates improved goodness of fit. These metrics collectively gauged the model's overall performance, optimization through hyper-parameter tuning, and predictive accuracy in capturing the relationship between features and the target value.

According to Table 7, in our comprehensive evaluation of various algorithms, Linear Regression (LR) and Random Forest (RF) emerged as the most adept techniques for predicting Therapeutic Alliance (WAI) for both clients and therapists. The LR utilizes a coefficient implementation, relying on weighted sums for predictions, where the coefficients serve as a crude feature importance score. Similarly, RF calculates the feature importance, offering insights into the relative significance of each predictor.

Table 7: Ranking algorithms

Technique	Client'sWAI		Therapist's WAI	
	MSE	R2	MSE	R2
Artificial Neural Network (ANN)	6.42	0.46	8.51	0.14
Random Forest (RF)	5.49	0.59	9.66	0.16
Decision Tree (DT)	8.04	0.33	16.48	(-) 0.23
K-Nearest Neighbors (KNN)	7.59	0.42	9.88	0.20
Support Vector Regression (SVR)	5.79	0.54	9.15	0.17
Linear Regression (LR)	6.06	0.53	7.90	0.30

4.3.3 Significance of Predictors

Table 8 shows the predictive outcomes of Logistic Regression (LR) and Random Forest (RF) models for the therapist and client Working Alliance Inventory (WAI). Notably, for therapist WAI, "Diagnostic" (score = 3.62) and "Termination" (score = 2.93) ranked as the most impactful variables with a positive influence. Conversely, "HR_Client(Mean_Baseline)" exhibited less influence, while "HR_therapist(mean_baseline)" and "HR_client(mean_baseline)" played significant roles. Specifically, a lower HR(mean_baseline) in therapists correlated with higher therapist WAI, with HR_therapist(mean_baseline) showing importance as a physiological indicator (coefficient score of (-) 0.15).

For client WAI, the therapist's WAI (score = 0.40) and "Outcome" (score = 0.31) emerged as substantial predictors. In the realm of physiological indicators, EDA_client(mean baseline) (score = 0.13) stood out as a significant factor.

In summary, concerning physiological indicators, while Heart Rate(mean_baseline) proved crucial for predicting therapist WAI (with a negative impact), EDA(mean_baseline) emerged as a noteworthy indicator for predicting client WAI.

Table 8: Significance of predictor

top predictor for " client's WAI" by RF				top predictors for "therapist's WAI" by LR			
physiological	Score	NOT physiological	Score	physiological	Score	NOT physiological	Score
EDA_Client (mean Baseline)	0.13	therapist's WAI	0.41	HR_Therapist (mean_baseline)	(-)0.15	Diagnostic	3.62
—	–	Outcome	0.31	—–	—	Termination	2.93

4.3.4 Results and Discussion

To conduct predictive analysis, we studied correlations among features and between features and targets, selecting relevant features and grouping repeated physiological indicators per session to prepare effective data for modeling. In the modeling phase, we employed various regression algorithms and ranked their performance through parameter tuning to identify the most effective algorithm for predicting the target. During the evaluation phase, we utilized different metrics to assess performance. Finally, we identified the significance of predictors to discuss outcomes at the application level. While this study is exploratory in nature, we successfully achieved the defined objectives and provided answers to the research question. Importantly, this investigation was conducted through interdisciplinary collaboration, closely working with the Department of Psychology at the University of Minho to support and validate our study, gain domain knowledge, and facilitate data collection.

At application level, this study offers crucial insights into the application of data mining in psychotherapy. data mining and ML techniques are increasingly employed in psychotherapy to aid clinical decision-making, considering factors such as patients' characteristics and treatment history. These algorithms excel at capturing intricate patterns and relationships, leading to more accurate predictions of therapeutic alliance.

Despite the rarity of such technologies in the field of therapeutic alliance, we have demonstrated their potential to strengthen it. By training prediction models on complex relationships identified in the data,

we can predict the level or quality of therapeutic alliance based on various input variables and patterns. Furthermore, integrating significant amounts of biological and psychological data allows us to understand, monitor, and predict the outcomes of clients' psychotherapeutic processes.

Machine learning models are particularly adept at tailoring predictions to individual clients and therapists, providing personalized insights into the factors contributing to a strong therapeutic alliance. This personalized approach empowers clinicians to deliver customized interventions and enhance the effectiveness of the therapeutic process.

Moreover, machine learning algorithms effectively address the complexity and non-linear relationships inherent in psychotherapeutic data, further enhancing their utility in this domain.

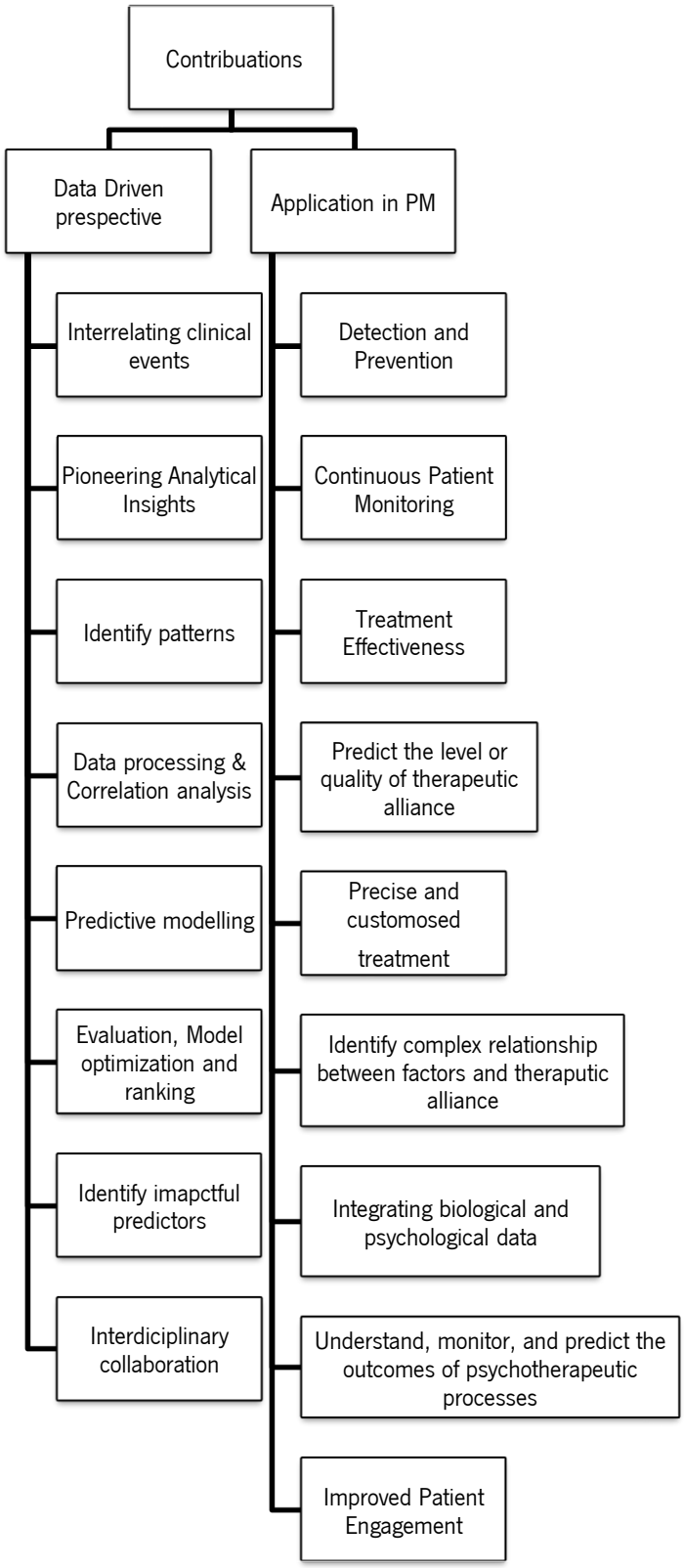


Figure 35: Contributions to IDSS4PM; Predictive Modelling

5. DATA-DRIVEN ICU MEDICINE DECISION SUPPORT (DD_ICUM_DS)

This chapter outlines the creation process of the framework, which comprises four key phases. Firstly, it involves the analysis of interrelated clinical events, applicable to each patient. Secondly, the identification of cluster references and analysis of patterns are undertaken. Thirdly, leveraging cluster references and mapping them onto clinical data to identify similar data points and study their behavior. Finally, predictive modeling and classification(clustering) is employed to place new data into the most suitable cluster, facilitating the enhancement of treatment plans for each cluster. Throughout the development of the artifact, we have effectively tackled significant challenges related to data aspects.

In each phase, we delve into the significant contributions to data-driven decision support systems, highlighting the relevance and impact of the artifact (IDSS4PM). We validate the architecture of this artifact by utilizing collected data from patients who have been in the Intensive Care Unit (ICU), ensuring its efficacy and applicability in real-world scenarios.

The IDSS4PM embodies several key characteristics presented in Figure 36 that significantly enhance its functionality. These include:

a. Advancing interoperability: The system seamlessly integrates with various data sources and platforms, ensuring smooth communication and data exchange.

b. Leveraging AI and machine learning for intelligence: By employing advanced algorithms and learning techniques, the IDSS4PM can analyze complex datasets and derive actionable insights.

c. Interdisciplinary approach: The system incorporates knowledge from diverse fields such as medicine, data science, and technology, fostering collaboration and innovation.

d. Evidence-based decision-making: Based on rigorous analysis of empirical data and established best practices, the IDSS4PM enables informed and evidence-driven decision-making processes.

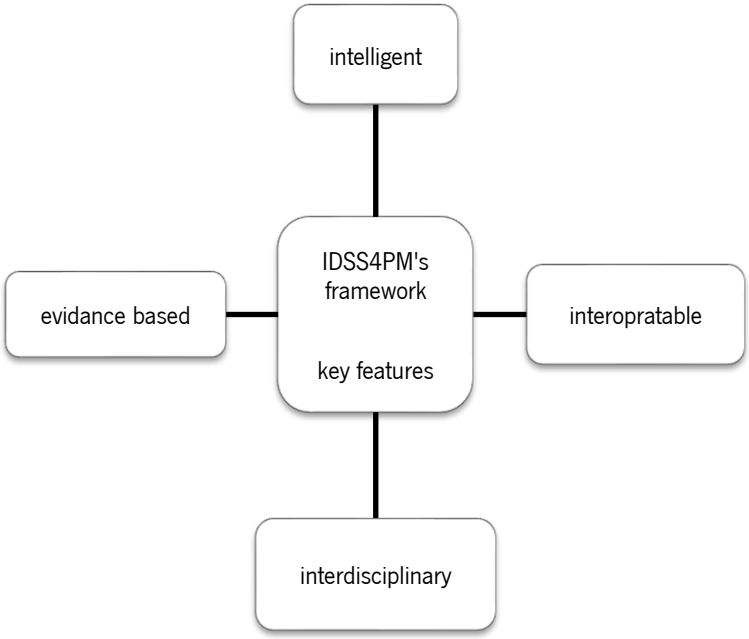


Figure 36: Key characteristics of IDSS4PM ‘s framework

5.1 Comprehensive Exploration of Datasets

According to Figure 37, the dataset encompasses 10 distinct categories, each providing valuable insights into various facets of patient health and medical interactions. Specifically, the data pertains to the clinical background of 70 patients who have been in the ICU.

1) “Vital Signs”: This dataset contains 43,9025 records and 108 biological variables, focusing on vital signs that play a crucial role in assessing a patient's overall condition.

2) “Laboratory Results”: Comprising 11,3320 records and 9 variables, this dataset provides information about various laboratory exams conducted, aiding in diagnosing and monitoring patients' health.

3) “Procedures”: With 911 records and 6 variables, the "Procedure" dataset sheds light on the medical actions recommended and prescribed by healthcare professionals.

4) “Sepsis (Gravity Score)”: Capturing data from 176 records and 6 variables, this category gauges the severity of patients' conditions, particularly in cases of sepsis.

5) “Glasgow Coma Scale”: Containing 861 records and 6 variables, this dataset evaluates patients' consciousness levels, a vital indicator of neurological well-being.

6) “Diagnosis”: Encompassing 124 records and 9 variables, the "Diagnosis" section focuses on recording signs, symptoms, and potential medical conditions.

7) “Medication Prescriptions”: This category with 35,422 records and, 39 variables provides data on medications prescribed by clinicians, helping track patient treatment plans.

8) Medical administration: with 993,496 records and 17 variables associated with drug administration.

9) SOAP: encompassing 2435 records and 8 variables, contains critical data related to the Subject, Object, Assessment, and Plan components following the SOAP framework

10) “Intervention Actions”: Capturing information about various interventions, this category showcases actions taken to manage patients' health.

11) ICU Admission and Discharge: "Admin-Discharge" houses data about patients' admissions and discharges from the ICU, facilitating comprehensive patient care management. This dataset consists of the date and time of admission and discharge.

12) Reference Dataset: Serving as a point of reference, this dataset includes episode and process numbers, wherein the episode number represents clinical events, while the process number signifies patient identity.

In Figure 37, datasets marked by: “|” (“vital sign”, “procedure” and “diagnosis”) have the time or date of the clinical event, and others with ||, include both (time and date). In addition, the symbol: “*” shows that data is associated with ICU (all tables except “med prescription). Moreover, in this table “R” shows the number of records, and “V” means the number of variables. In addition, two variables include distinct values whether “Process Number” (DP) or “Episode Number” (DE).

Collectively, these datasets offer a multidimensional view of patients' health journeys, medical interactions, and treatment outcomes. This rich and diverse collection of data allows for in-depth analyses, enabling healthcare professionals and researchers to gain valuable insights into patient well-being and medical practices. As such, this dataset holds great potential for advancing medical knowledge, enhancing patient care, and guiding evidence-based decision-making.

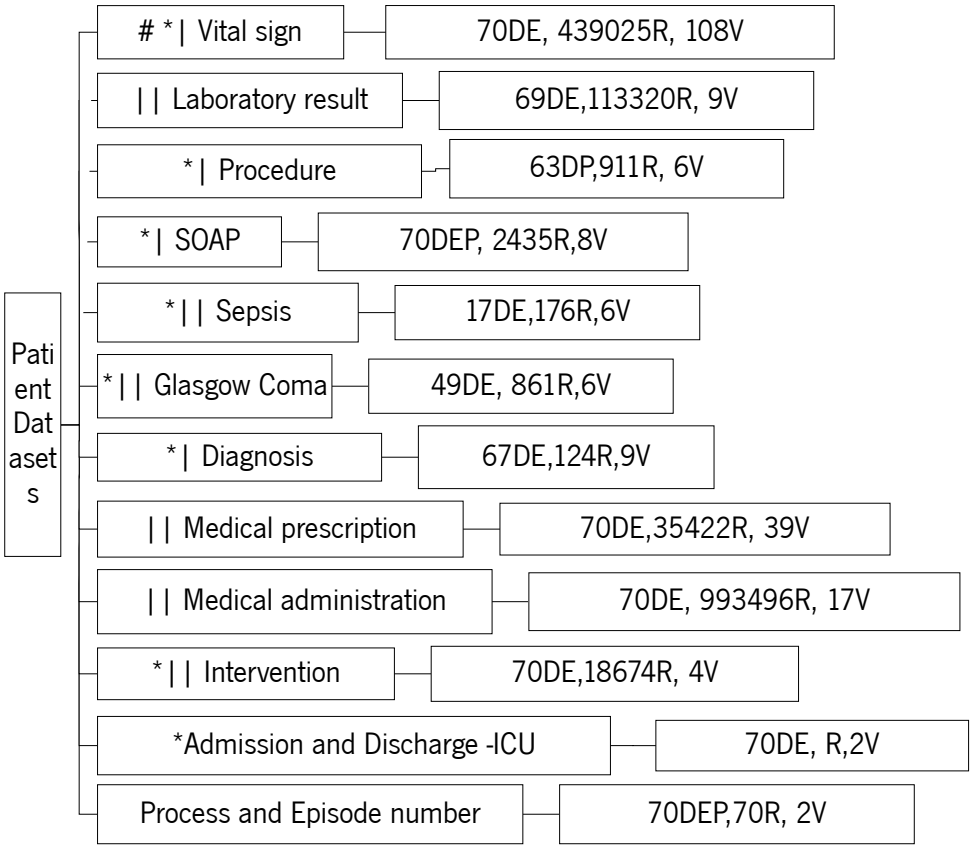


Figure 37: Data exploration

Like the previous Project (DD-PT-DS), special care was taken to prioritize privacy and ethical concerns. To address these issues, the data underwent rigorous de-identification and coding processes to ensure complete anonymity and prevent any association with real patients.

5.2 Interrelated Clinical Events With CEid

As depicted in Figure 38, our approach to conducting temporal analysis on interconnected clinical events involved several key steps, including "data understanding," "data preparation," "CEid," and "Interrelating clinical events." Notably, the introduction of the CEid (Clinical Event Identity) emerged as an innovative solution, streamlining the integration process and enhancing the correlation of clinical events for each patient concerning the key information.

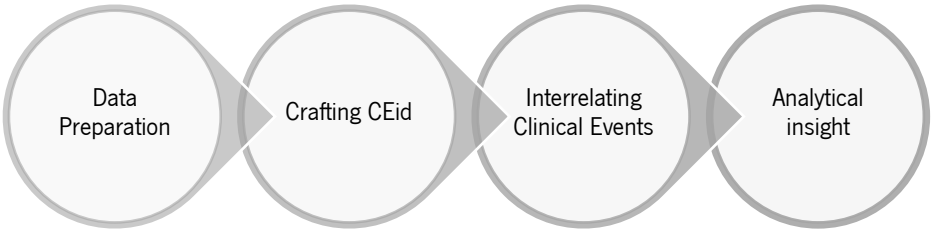


Figure 38: Steps to perform analytical insight

5.2.1 Data Processing and Preparation

This phase encompasses a series of pivotal data processing tasks, aimed at enhancing the quality and applicability of the collected data. These tasks revolve around critical aspects such as:

a. Feature Selection: Employing a meticulous approach to feature selection, we leveraged domain knowledge, statistical analyses, and data quality assessments. In instances like the vital sign dataset, variables with over 90% missing data were systematically excluded.
b. Feature Engineering: In several datasets, we introduced novel variables to enhance analytical insight and harness valuable information. A prime illustration involves deriving the length of hospital-ICU stay by utilizing admission and discharge dates, thus transforming raw data into meaningful metrics.
c. Extraction and Processing: The effectiveness of analytical performance was significantly amplified through rigorous data extraction. An illustrative example is the extraction of demographic details such as age and gender from diverse datasets, culminating in the creation of a consolidated dataset that streamlined subsequent data processing.
d. Conditional Columns: Our transformative approach extended to the creation of conditional columns, converting raw data into actionable information. An example is our utilization of laboratory references for each exam, enabling a systematic comparison with exam results to ascertain their normal or abnormal status.
e. Grouping and aggregating time-series data for vital indicators, thus addressing sporadic data registration issues encountered with biological sensors in the ICU. By adopting an hourly aggregation approach, we effectively mitigated the challenge of infrequent data updates.
f. Cleaning-Missing cells: In addressing gaps within the vital sign dataset, a meticulous strategy was employed: we judiciously populated missing cells by computing the average value from neighboring cells preceding and following the voids. Similarly, within another

dataset, a pragmatic approach was taken by eliminating missing cells.

As an integral part of our preparations, we meticulously fine-tuned all datasets to ensure they boasted fitting data types and pertinent features.

Table 9: Key Transformations

#	Key data processing	Description | example
1	Exploratory data analysis	Identify mean, max, min, missing cells, and data quality
2	Feature Selection	Using Knowledge of domain and missing cells
3	Feature engineering	Construct new variable such as length of stay
4	Data extraction	Demography table; Extract age and gender
5	Conditional feature	Create Laboratory result status
6	The correct type of data	Considering categorical, numerical, and also time series data
7	Group and aggregation	For handling infrequent data generation, particularly from biological sensors
8	Missing cells	Imputation method and elimination.

5.2.2 Encapsulating Diverse Clinical Events with CEid

This phase harnessed a formula to craft a distinct key, uniquely pointing to each clinical event for individual patients. Illustrated in Figure 39, this key exhibits a structured composition: the initial number designates the event's day, followed by an abbreviation denoting the data type. Additionally, the third and fourth components encompass the patient's process number and episode number, respectively. The sequential order of events during a specific period is encapsulated within the event sequence, an ascending value ranging from one to n, culminating in the count of parallel events transpiring on the same day.

Figure 39 serves as an illustrative example, underscoring an event linked to an episode number (20016701) attributed to a patient identified by process number 859785. This informative data point delineates the event's occurrence on the fifth day within the ICU context (vital sign data), constituting the eleventh clinical transaction.

Creating this new variable holds significance in establishing connections among clinical transactions through a unique timeline based on the day of the clinical event. This process aims to facilitate the development of an analytical dashboard, wherein this code serves as a pointer to the key information of a patient. Also, to facilitate the clustering analysis.

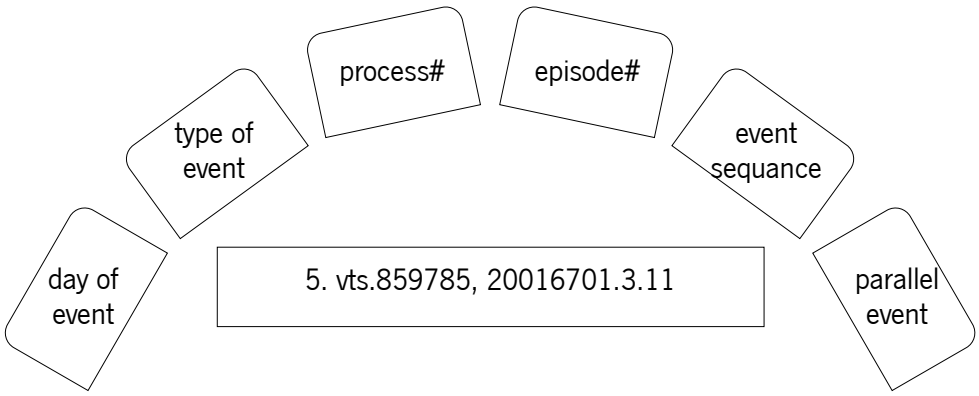


Figure 39: Structure of CEid

5.3 Analytical Insight Through Interrelated Clinical Events

Conducting analytical insight for the temporal analysis of interrelated clinical events stands as a potent data visualization tool in clinical decision-making. The paramount goal is to shed light on the chronological sequence and interconnections among a myriad of clinical events, providing invaluable insights into the temporal facets of patients' medical trajectories. This process enhances our understanding of the intricate web of events, contributing to a more nuanced comprehension of patients' healthcare experiences.

The innovative integration of CEid further enhances the interrelation of these clinical events, providing a foundation for the development of advanced analytical insight. This integration using the multidimensional model of tables, provides a holistic perspective on patients' medical trajectories. Moreover, introduces a range of filters, empowering decision-makers to meticulously observe and analyze the distinctive clinical journey of each patient over a timeline, especially during their hospital stay.

Figure 40 offers a visual representation of the access to analytical insights facilitated by the utilization of CEid, highlighting its capacity to provide a comprehensive view of patient's medical journeys. Through the utilization of various filters such as "gender," "process number," "episode number," "day of a clinical event," "age," and "type of clinical transaction," decision-makers can access and analyze patients' clinical trajectories. This allows for the identification of trends and patterns within the data.

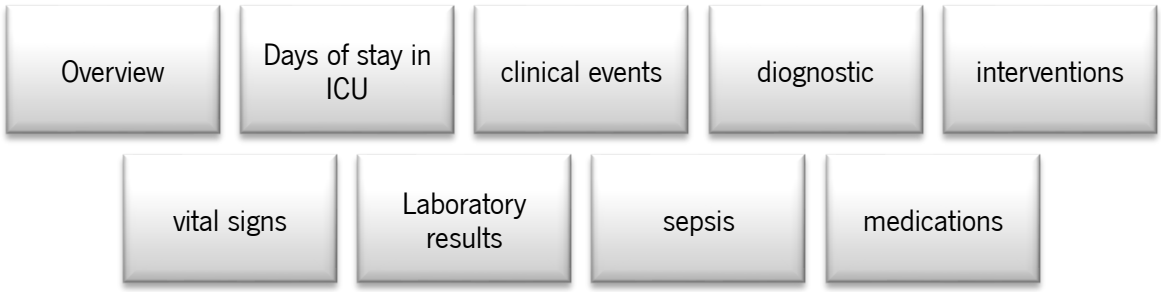


Figure 40: Access the Patient's analytical insight

5.3.1 Length of Stay in ICU

In Figure 41, the bar chart illustrates the length of stay in the ICU for each patient, with the process number serving as the identifier. The legend distinguishes between genders, indicating whether the patient is female or male. Additionally, the presentation includes pre-operation and post-operation distinctions in the scatter chart. Furthermore, a scatter chart showcases the relationship between the length of stay and age. This diverse set of information can be accessed and presented using various filters, providing insights into demographics alongside the length of stay for each patient.

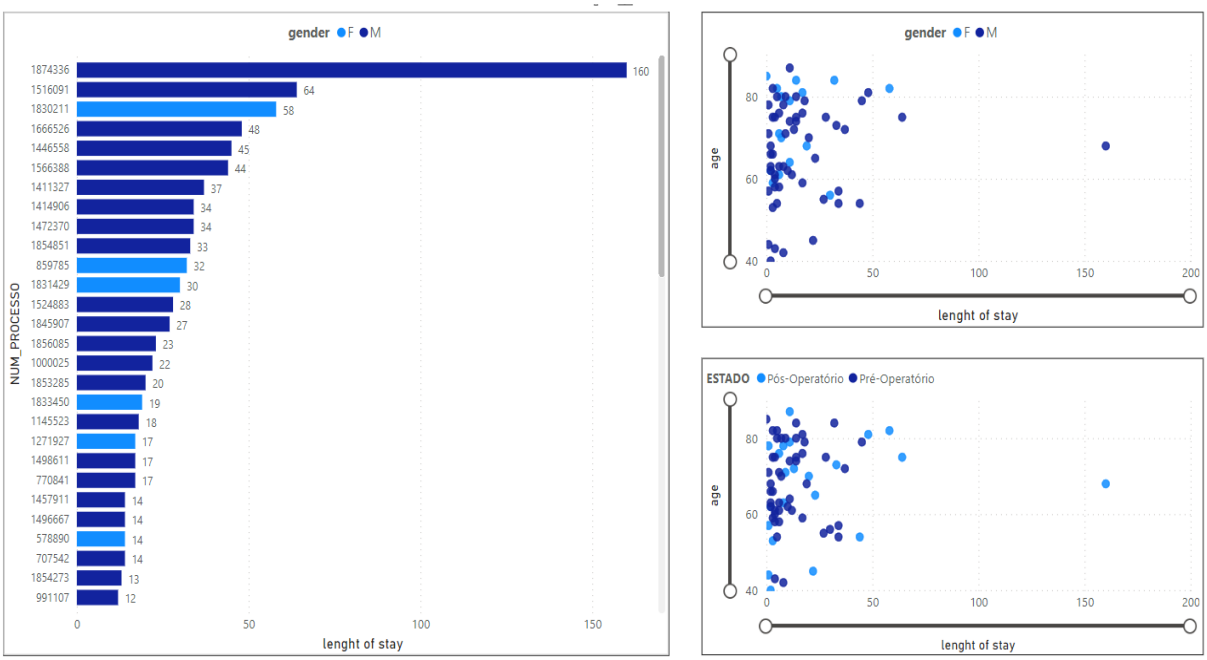


Figure 41: Length of stay in ICU

To present a comprehensive view of diverse clinical information on an integrated platform based on the "day of a clinical event," we focused on a specific patient with a process number equal to "1000025,"

as depicted in Figure 42. The total number of clinical transactions (TS) associated with this patient is 7798, with 5977 distinct transactions recorded. Notably, the patient is a 45-year-old male with the pathology code "N979." Their ICU status indicates "post-operation," with a total ICU stay of 22 days.

Figure 42: Analysing Clinical Data for Patients with process number:"1000025"

5.3.2 Total Number of Clinical Events

In Figure 43, the x-axis effectively chronicles the progression of days throughout the patient's tenure in the hospital, while the y-axis aptly quantifies the tally of clinical events, neatly categorized by transaction type.

This informative bar chart represents a filtered view, concentrating on three pivotal event types for patients sharing a common process number of "1000025": diagnostics (dig), interventions (int), and laboratory exams (lab). A discerning glance at this chart reveals that diagnostic events were notably concentrated on the first two days, with just a few transactions recorded. Furthermore, intervention actions persevered for a remarkable 27-day duration.

And 125 laboratory exams on day one were reduced to 24 exams on day two.

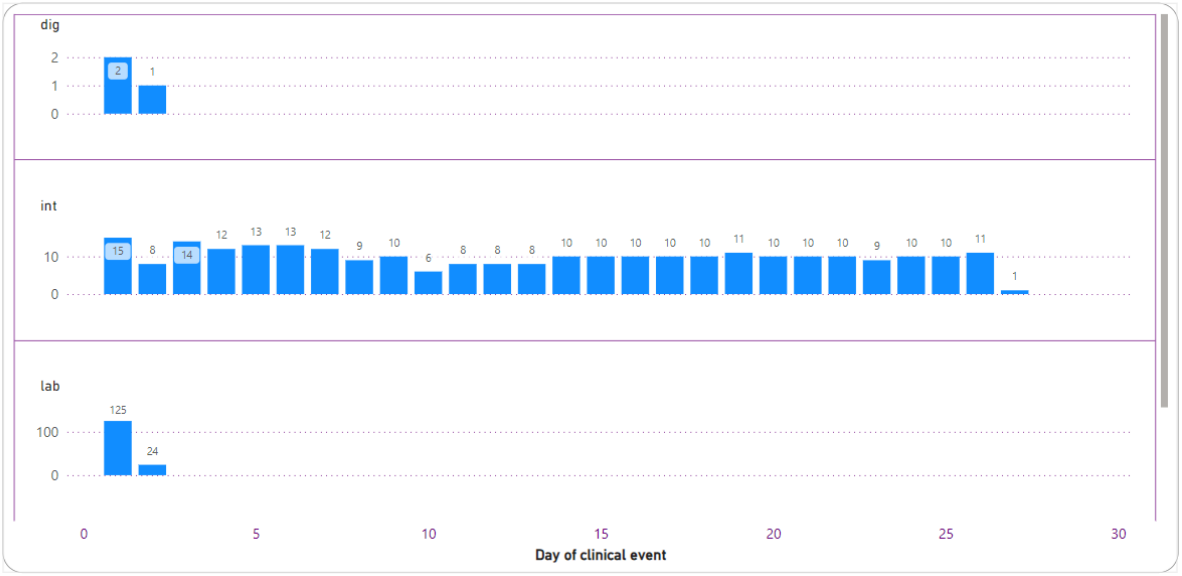


Figure 43: Temporal analysis of diverse clinical transactions

5.3.3 Diagnostics

Figure 44 illustrates the number and types of diagnoses for the selected patient. It reveals that COVID-19 was diagnosed on day one, followed by a diagnosis of SARS on day two. This observation and analysis provide valuable insights into the patient's health status over a distinct period, facilitated by the unique periodic platform of the day of the clinical event.

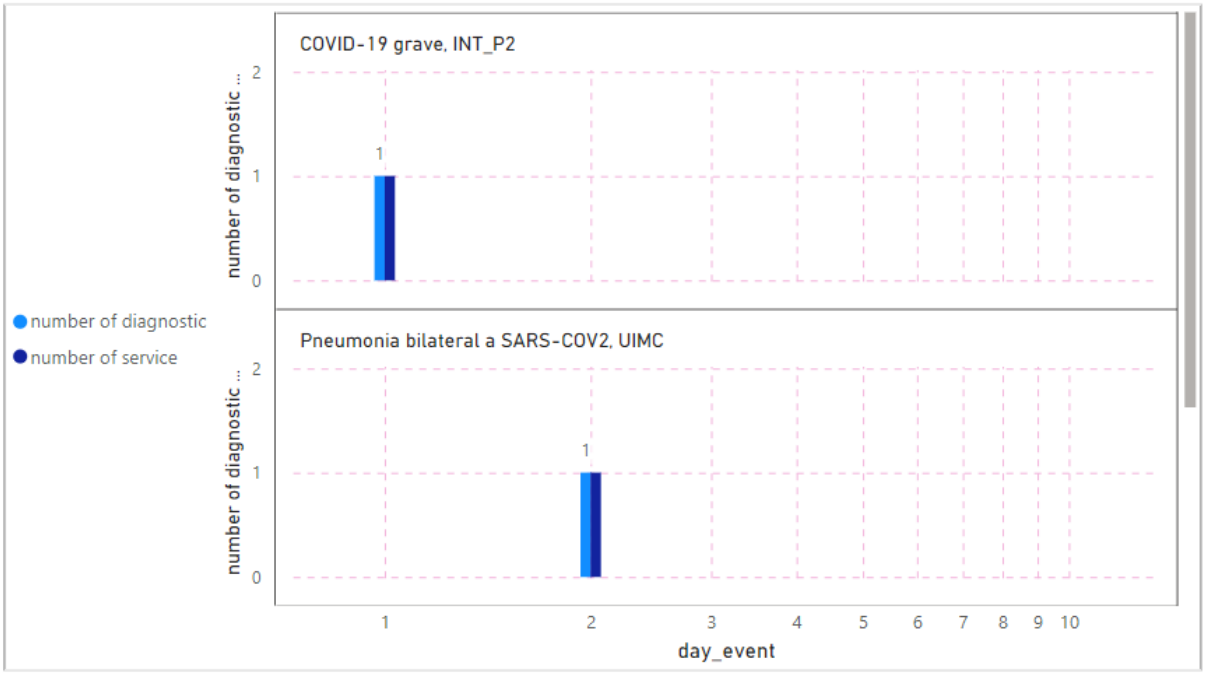


Figure 44: Temporal analysis of Diagnostic transactions

5.3.4 Intervention

Figure 45 displays two types of interventions repeated from day one to day 21 and day 30. The y-axis represents the total number of each specific type of intervention.

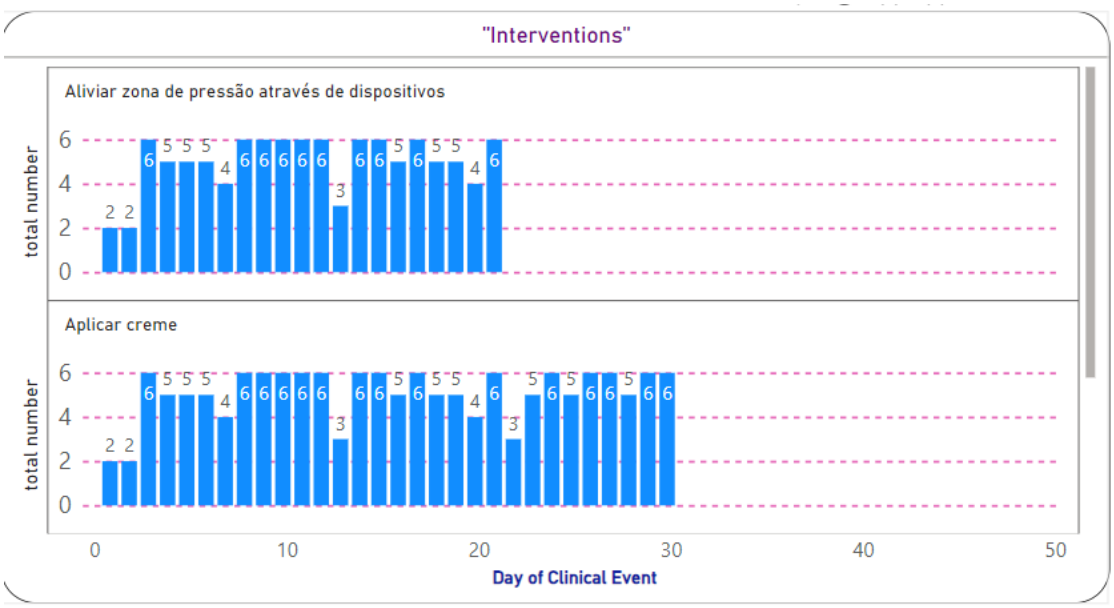


Figure 45: Temporal analysis of intervention transactions

5.3.5 Vital Signs

The vital sign dataset includes seven biological indicators registered by sensors in ICU, after hourly aggregation and transformation, all data points are placed in a timeline regardless of the time and date of data acquisition. Thus, there possible to monitor the average, minimum, and maximum value of each vital sign during the days of stay in the ICU. Figure 46 presents the fluctuation of the oxygen saturation average value (light blue), the maximum value (orange line), and the minimum value (blue line). The minimum value was observed on day three with a value of 83.9, and the maximum value of oxygen saturation was registered on days one and two with a value of 100. During four days of monitoring, fluctuations in the average value were observed to range between 100 and 97."

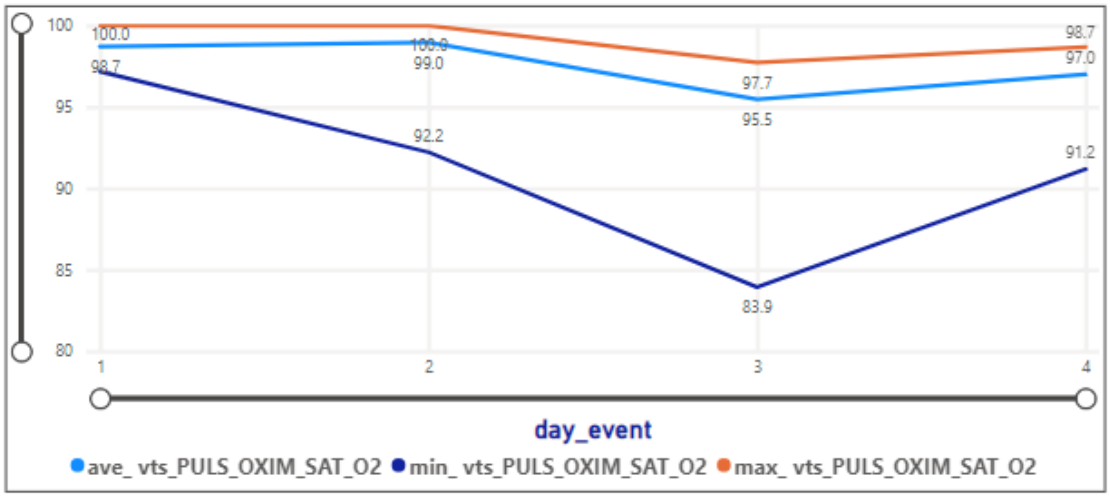


Figure 46: Temporal analysis of oxygen saturation

Figure 47 illustrates a line chart that provides a detailed view of the fluctuation in Heart Rate (HR) values over the course of the patient's stay in the ICU. The y-axis of the chart represents the HR values, showcasing the minimum, maximum, and average HR values recorded each day. This visualization offers valuable insights into the patient's HR trends during their ICU stay. On the second day of the patient's admission, the chart indicates a minimum HR value of 56.88 beats per minute (bpm). In contrast, the HR reached its highest point on the fourth day, recording a maximum value of 148.87 bpm.

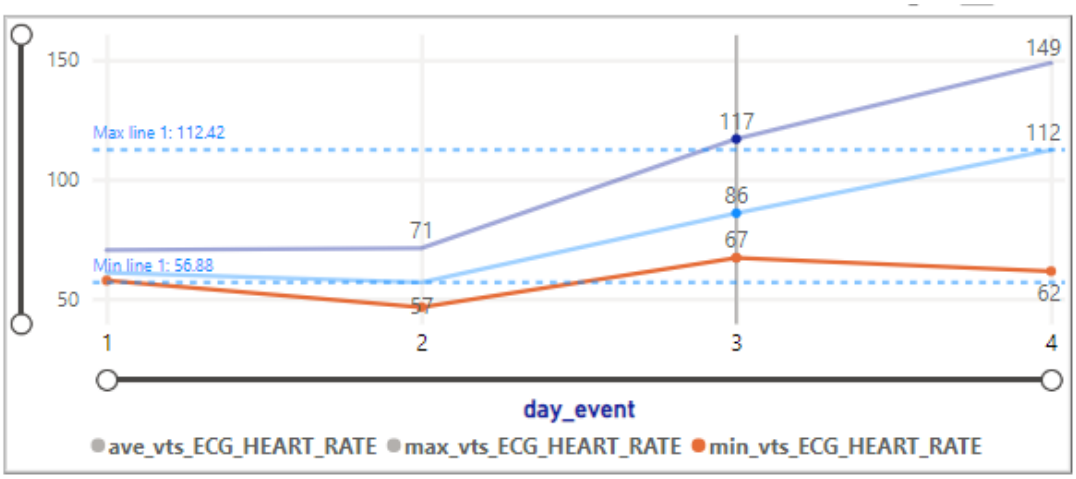


Figure 47: Temporal analysis of Heart Rate

5.3.6 Laboratory Result

Figure 48 shows the results of laboratory exams under the specific category. We transformed the minimum and maximum references associated with each test to define a conditional column. This feature construction identifies the condition of the result. Based on that, “Lr” means lower than minimum, “mr” is more than maximum, and “nr” means normal. According to the line chart, the result of “Neutrophiles” from 81 dropped to 6 on day one. It means that the value decreased from more than the maximum to a normal level. This method extracts significant information on the condition of a specific patient regarding the particular laboratory exam, and analyzing such information helps clinical decision-makers observe necessary laboratory results over days of stay in the hospital.

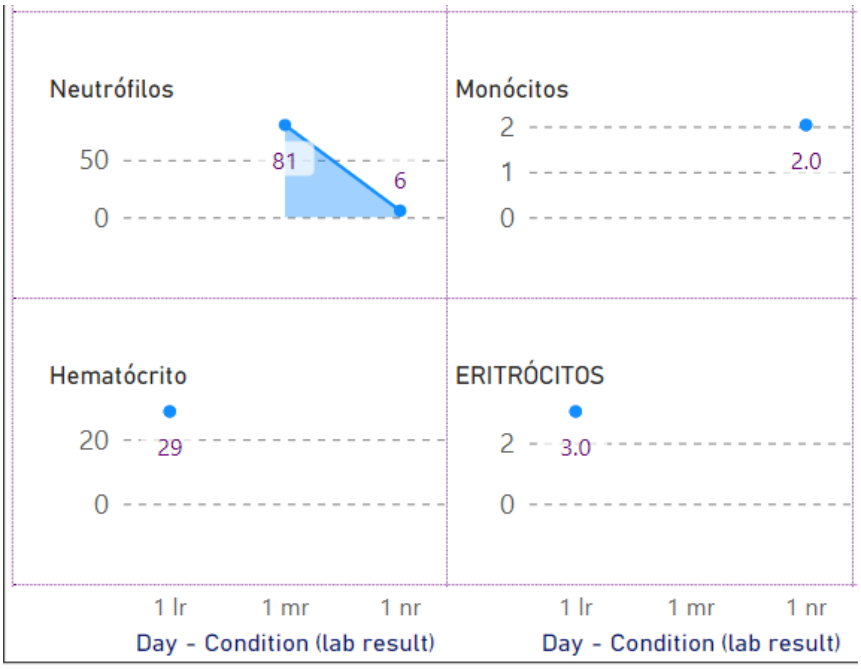


Figure 48: Temporal analysis of laboratory exams

5.3.7 Sepsis

Figure 49 presents a graphical representation illustrating the occurrence of four distinct sepsis events over four days. Each instance is denoted by a recorded value, which remained constant at 53 for the first, second, and third days. However, on the fourth day, a noticeable increase is observed, with the recorded value rising to 60. This chart provides a clear and concise visualization of the timeline of sepsis events, offering valuable insight into the progression and severity of these occurrences over the observed period.

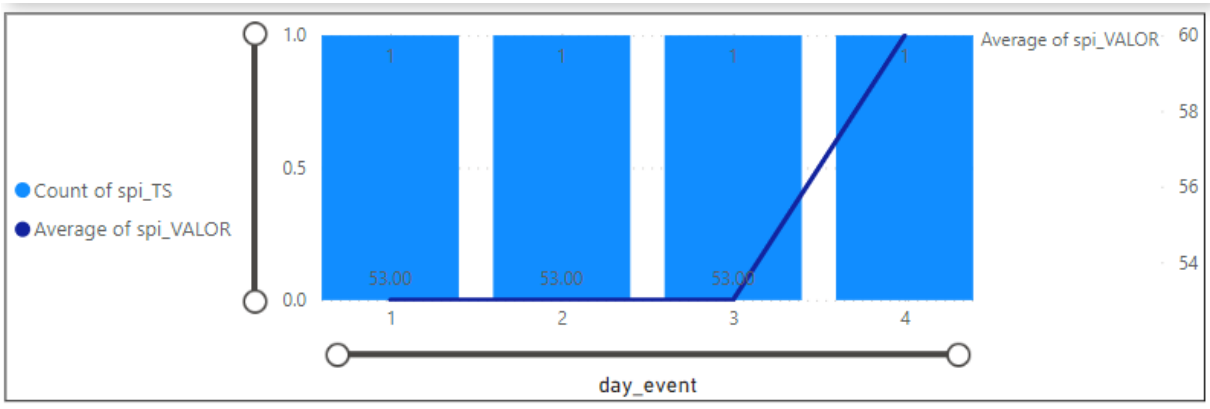


Figure 49: Temporal analysis of sepsis

5.3.8 Medication Prescription

Figure 50 provides insights into four different types of prescribed medications, complete with average dosages and their respective units of measurement. Notably, on day 26, a prescription for 500 milliliters (ml) of 'glucose 10%' was administered. 'Brometo de ipratropium' was consistently prescribed from day 14 to day 40 at a fixed dose of 500 milligrams (mg). 'Colloredo de sodio' was recommended for a duration of 40 days, with varying dosage levels. Lastly, 'hydrocortisone' was prescribed from day 13 to day 26, with a notable decrease in dosage from 200mg to 50mg during this period.

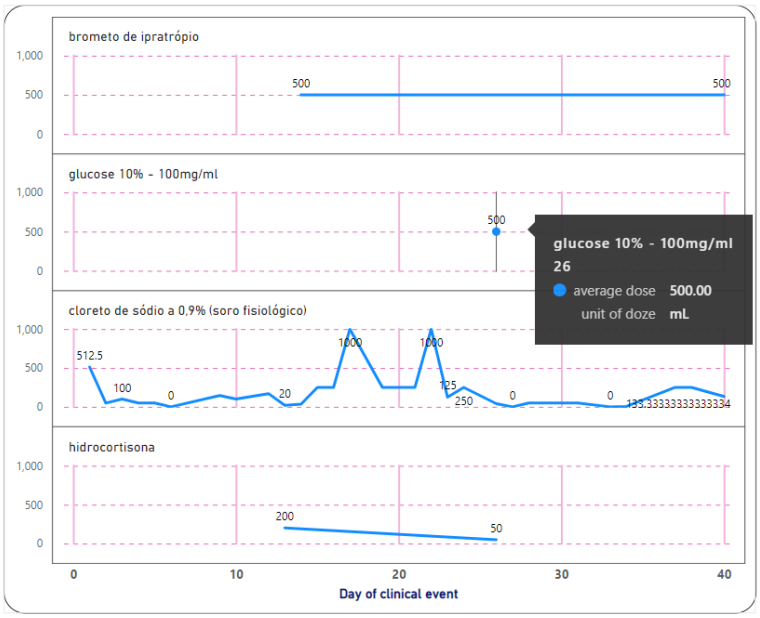


Figure 50: Temporal analysis of medications

5.3.9 Results and Discussion

Figure 51 illustrates the significant contributions derived from analytical insight. Introducing the CEid and establishing connections among clinical events for individual patients through an analytical dashboard constitutes a significant scientific contribution to data-driven clinical decision-making, particularly within the realm of precision medicine. This phase marks a substantial advancement in the evolution of intelligent decision support systems tailored for precision medicine. The impact of the analytical dashboard, facilitated by CEid, is comprehensively explored from both a data-driven perspective and practical application. Furthermore, we systematically address the encountered limitations during this phase, ensuring a comprehensive understanding of the challenges and opportunities in implementing this innovative approach.

a. The Transformative Impact of interrelating clinical events via CEid

The formulated key plays a crucial role in advancing the analytical insight and also performing the next phases of the experiment by solving key challenges and advancing the further steps towards the framework of an intelligent decision support system for precision medicine. From a data-driven perspective, there are key impacts such as:

i. Information Exchange:

The meticulously designed key within CEid functions as a standardized identifier for clinical events, fostering consistency and streamlining information exchange during data processing and integration.

In essence, CEid, serving as a common temporal reference, plays a pivotal role in facilitating smooth information exchange. It provides a standardized framework that enhances our ability to comprehend and interpret temporal relationships within clinical data. This not only promotes data consistency but also ensures a cohesive and interoperable approach to handling temporal aspects in diverse healthcare systems and platforms.

ii. A meaningful pointer to a data point

The implementation of this standardized key ensures the effortless extraction of individual data points, providing a consistent mechanism for the exchange and sharing of information across a variety of systems and platforms. Each component of this key holds significance, indicating whether the event is linked to a specific patient identity, episode number, and the day of that clinical transaction. Furthermore, it serves to discern whether there are parallel occurrences of similar clinical events or not.

iii. Standardization for interoperability

The formula introduces a standardized approach to creating keys for clinical events, ensuring uniformity and consistency in data representation. This standardization is crucial for interoperability and compatibility across various healthcare systems. The inclusion of the CEid variable adds a standardized temporal dimension to the dataset, enabling a consistent and uniform representation of time across different clinical events. This supports standardization in the analysis and interpretation of temporal relationships.

iv. Pioneering Analytical Insights through dynamic Dashboards

In particular, the CEid variable stands out as an asset, providing a dynamic and flexible means to analyze and interpret complex patient data. Beyond supporting informed decision-making, this approach has the potential to significantly enhance the quality of patient care by offering a nuanced understanding of individual medical journeys.

b. Predictive Modelling:

The structured key, with its detailed components such as the day of the event, data type, process number, and episode number, serves as a rich source for features in predictive modeling. This can contribute to more accurate predictions based on the temporal relationships between different events. For example, the event sequence, part of the crafted key, provides a chronological order of clinical events, enabling temporal clustering. This sequential information allows for the identification of patterns or clusters of events occurring closely in time.

Evidence-based Decision: Data-Driven Decision-Making: The integration of clinical events provides decision-makers with data-driven insights, allowing for evidence-based decisions that are more likely to result in successful treatment and patient satisfaction.

At the application level, this groundbreaking approach empowers medical professionals with a comprehensive perspective, enabling them to navigate and decipher intricate patient trajectories for more informed decision-making and improved patient care. The integration of diverse clinical events into a unified temporal framework, represented by the CEid variable, serves as a potent tool for healthcare practitioners and allows for versatile filtering options, facilitating a thorough review of various clinical events and their associated details. These are the key contributions at the application level:

a. Early Detection and Prevention: The ability to track and analyze clinical events over time allows decision-makers to identify trends and patterns that may indicate the onset of health issues. This early detection can lead to proactive interventions, reducing the severity and cost of treatment.

b. Continuous Patient Monitoring: The scrutiny of clinical data fosters ongoing patient monitoring. By observing data fluctuations over time, healthcare providers can identify deviations from a patient's baseline health and intervene early to mitigate complications.

c. Treatment Effectiveness: Clinical decision-makers can assess the effectiveness of treatment strategies by analyzing how patients respond to interventions over time. This information helps in adjusting treatment plans for better outcomes and resource allocation.

d. undrestanding of each patient's health journey**:** By unifying clinical events, decision-makers can gain a comprehensive understanding of each patient's health journey. This information

enables them to make more personalized treatment decisions, tailoring interventions to the individual needs of the patient.

e. Reduction of Unnecessary Costs: Understanding the progression of a patient's health condition and the impact of various clinical events allows decision-makers to avoid unnecessary and costly diagnostic tests and treatments that may not be effective.

f. Optimized Resource Allocation: Decision-makers can allocate healthcare resources more efficiently by focusing on patients who need immediate attention or personalized care. This improves resource utilization and patient outcomes.

g. Improved Patient Engagement: Patients who see their healthcare providers making data-informed decisions based on a unified view of their clinical events are more likely to be engaged in their care. This can lead to better adherence to treatment plans and improved health outcomes.

h. Preventive and Proactive Health Management: It's not just about treating diseases but also about identifying risk factors and taking proactive steps to prevent or delay the onset of diseases.

In summary, the unification of clinical events empowers clinical decision-makers to deliver more effective, precise, and cost-efficient care in the realm of precision medicine. It enhances the decision-making process by providing a holistic view of a patient's health history, enabling proactive care, and ultimately leading to improved patient well-being.

Overall, this analytical observation in healthcare aims to enhance decision-making by providing a visual representation of the temporal relationships among clinical events. It supports healthcare professionals in identifying patterns, making informed decisions, and improving patient care based on a comprehensive understanding of the temporal aspects of medical histories.

In this experimental phase, our primary aim was to discern patterns among data points displaying similar behavior, irrespective of patient assignments. Leveraging unsupervised learning, particularly clustering, we delved into the temporal dynamics of the data. This analysis holds substantial importance in tracking the progression of data, specifically health conditions, over a designated timeframe. Notably, we focused solely on numerical variables during this phase. Our secondary goal involved establishing cluster references by considering the minimum and maximum values of each numerical variable within each cluster. Subsequently, these cluster references were employed to map onto the original dataset,

encompassing both numerical and categorical variables, facilitating the assignment of clusters to the entire dataset.

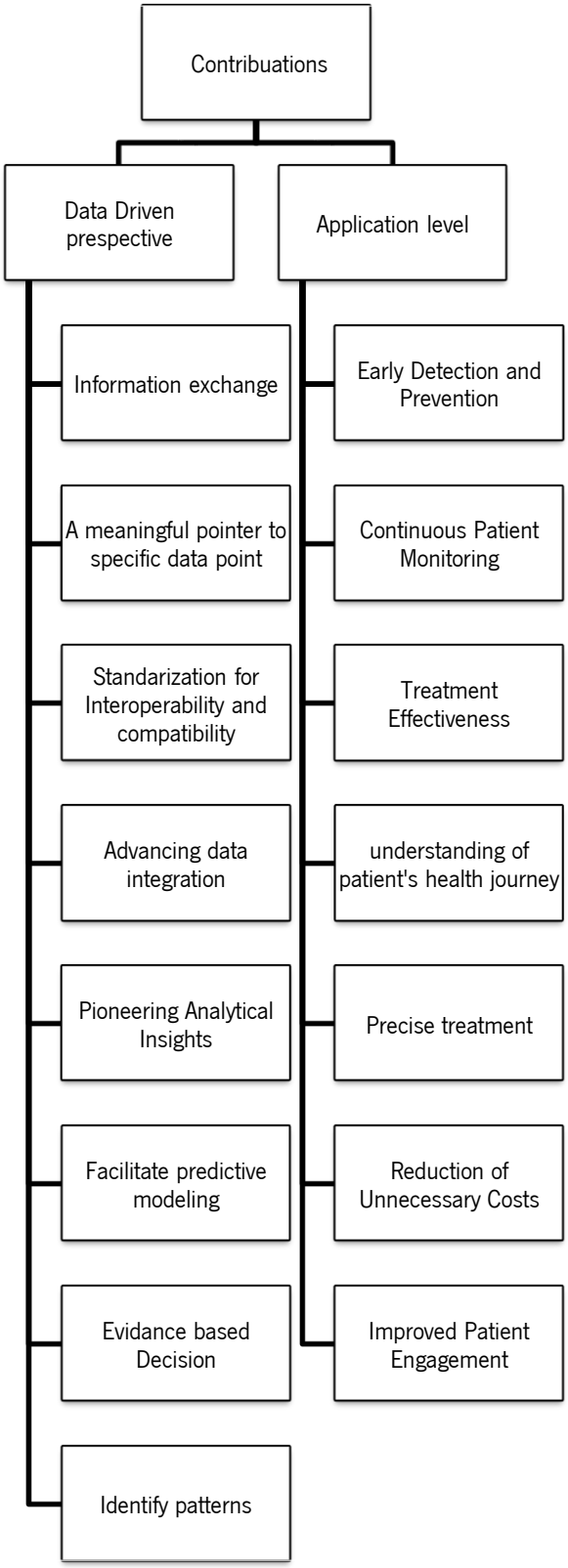


Figure 51: Contributions to IDSS4PM; Analytical Insights

5.4 Unveiling Patterns Throughout Temporal Clustering Analysis

In this experimental phase, our primary aim was to discern patterns among data points displaying similar behavior, irrespective of patient assignments. Leveraging unsupervised learning, particularly clustering, we delved into the temporal dynamics of the data. This analysis holds substantial importance in tracking the progression of data, specifically health conditions, over a designated timeframe. Notably, we focused solely on numerical variables during this phase. Our secondary goal involved establishing cluster references by considering the minimum and maximum values of each numerical variable within each cluster. Subsequently, these cluster references were employed to map onto the original dataset, encompassing both numerical and categorical variables, facilitating the assignment of clusters to the entire dataset.

As illustrated in Figure 52, our primary action involved consolidating all datasets into one cohesive unit. After this, we applied rigorous data processing techniques to render the data amenable to temporal clustering. The outcome proved instrumental in identifying patterns among similar data points.

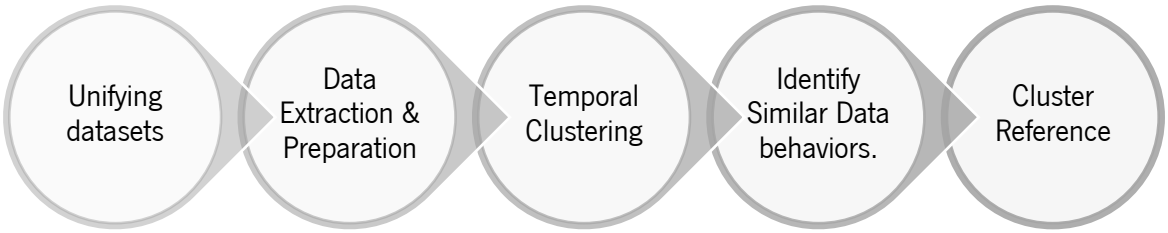


Figure 52: Towards temporal clustering analysis

5.4.1 Unifying Datasets For Clustering

To integrate the datasets, we meticulously followed the outlined procedures in Figure 53. This involved a meticulous examination of each dataset, with a specific focus on extracting the "day of clinical event" from the CEid dataset. For numerical data, we adopted a comprehensive approach, grouping the information based on "Process number," "day of clinical event," and computing the mean value. In handling categorical data, custom codes were generated for entities lacking inherent identity, such as interventions, procedures, diagnostics, and laboratory exams. Subsequently, rows were aggregated and grouped using the key variables "Process Number," "Day of Clinical Event," and "Code."

Further refinement involved pairwise merging of datasets, utilizing the common variable "Process Number" present in each dataset. To enrich the dataset, information about the "length of stay,"

admission/discharge timestamps, and date details were incorporated by merging with the admission-discharge dataset. The resulting consolidated dataset, comprising 1,705,174 rows, encompassing 29 variables, and associated with 70 patients (process numbers), was meticulously prepared for subsequent clustering analysis.

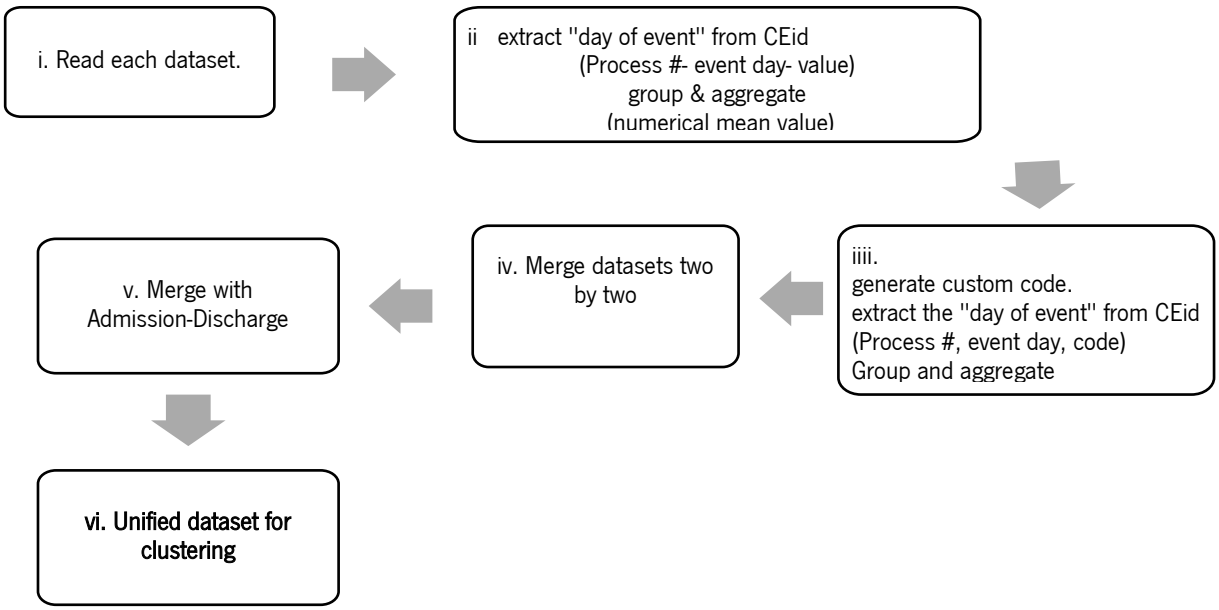


Figure 53: Pipeline to unify datasets

5.4.2 Clustering

In preparation for the k-means clustering approach, we focused exclusively on numerical variables, omitting the "Process Number" too. Our data preparation encompassed several crucial steps, including handling missing cells, discarding columns with more than 50% missing values (Glasgow Coma, Sepsis), and eliminating corresponding rows with missing data. This involved meticulous selection of numerical variables, rectifying variable types, and ensuring the dataset's integrity by addressing duplicates. Consequently, the refined dataset earmarked for clustering comprises 692 rows and 9 variables, each intricately linked to 70 patients across a span of 39 days of clinical events.

To choose the optimal number of clusters we performed the elbow method.

The elbow method is a heuristic approach employed to determine the optimal number of clusters (k) in a dataset. It involves running the k-means clustering algorithm on the dataset for a range of values of k (e.g., from 1 to a certain maximum). For each value of k, the Within-Cluster Sum of Squares (WCSS) or the sum of squared distances between data points and their assigned cluster centroid is calculated. The WCSS is then plotted against the number of clusters, and the "elbow" of the curve represents the point

where the rate of decrease in WCSS slows down significantly. The idea is to choose the value of k at the elbow, as it signifies a good balance between model complexity (number of clusters) and the fit to the data.

Figure 54 depicts the application of the elbow method to determine the optimal number of clusters. In addition to the elbow method, we explored a cluster range from n=2 to n=7, utilizing the average silhouette score for each cluster configuration. Through this analysis, we identified n=4 as the optimal number of clusters, ensuring a balance between cohesion within clusters and separation between clusters. The silhouette score is a valuable metric for assessing the quality of clustering, and our approach aimed to leverage it effectively in determining the most suitable cluster count for the given dataset.

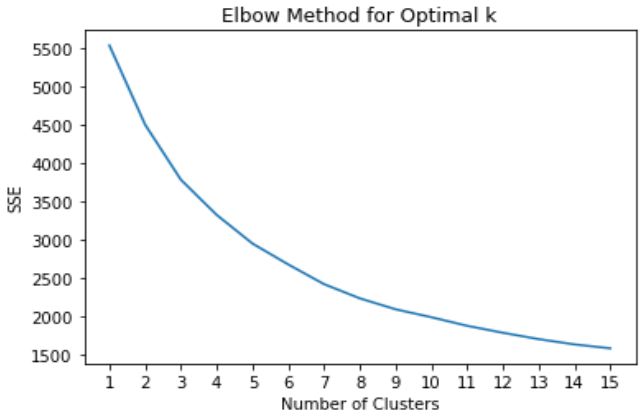


Figure 54: Elbow method for the optimal number of clusters

5.4.3 Identifying Patterns and Cluster References

To analyze the characteristics of each cluster we extracted the average and standard deviation (std) of each variable in each cluster. The presented Table 10 outlines key numerical indicators across four distinct clusters (CLUSTER 0, CLUSTER 1, CLUSTER 2, CLUSTER 3), each characterized by a Silhouette Score denoting the cohesion and separation of data points within the cluster. The Silhouette Score ranges from 0.12 to 0.36, providing an assessment of the clustering quality, with higher scores indicating better-defined clusters. Additionally, the table includes data distribution statistics, specifying the number of data points (rows) in each cluster.

a. The key observation of Cluster 0 with the Silhouette Score equal to 0.23 and Data Distribution equal to 271 rows shows:

- The average pulse rate of the arterial blood pressure is 100.66 with a standard deviation of 12.20.
- Diastolic arterial blood pressure is around 64.18 with a standard deviation of 9.6.
- Systolic arterial blood pressure is notably higher at 118.81 with a standard deviation of 15.46.

- Mean arterial blood pressure is 82.65 with a standard deviation of 11.14.

- Heart rate is 99.18 with a standard deviation of 17.49.

- The pulse oximetry oxygen saturation level is relatively high at 94.21 with a small standard deviation of 4.44.

- Body temperature is around 36.17 with a standard deviation of 1.27.

- The day of the clinical event is on average 10.60 with a relatively high standard deviation of 8.00.

b. Cluster 1 with a Silhouette Score equal to 0.12 and Data Distribution of 44 has a lower silhouette score, indicating less cohesion and separation.

- Variables exhibit higher standard deviations, suggesting greater variability within this cluster.

- Body temperature has a relatively high standard deviation of 2.72, indicating variability in this aspect.

Cluster 2 Silhouette has a higher silhouette score, suggesting better-defined clusters (Score: 0.26) and the Data Distribution equal to 285 rows. In This cluster

- Variables like systolic arterial blood pressure and mean arterial blood pressure show significant variability with standard deviations of 15.13 and 12.20, respectively.

- The pulse oximetry oxygen saturation level has a low standard deviation, indicating more consistency.

c. Furthermore, Cluster 3 has the highest silhouette score, indicating well-defined clusters (Silhouette Score: 0.36) and Data Distribution equal to 92 rows. The key Observations show:

- Systolic arterial blood pressure is notably high at 129.29 with a standard deviation of 15.13.

The day of the clinical event has the lowest average at 4.98, suggesting events are concentrated around this day.

In terms of any possible Trends and Patterns:

- Cluster 3 stands out with the highest silhouette score, indicating the most distinct cluster.

- Systolic arterial blood pressure is a key differentiator among clusters, with Cluster 3 having the highest values.

The day of the clinical event shows variation, with Cluster 2 having the highest average and Cluster 3 having the lowest.

Table 10: Characteristics of each cluster

Indicator – Numerical	CLUSTER 0	CLUSTER 1	CLUSTER 2	CLUSTER 3
(692 rows, 8 columns)	Silhouette Score: 0.23	Silhouette Score: 0.12	Silhouette Score 0.26	Silhouette Score: 0.36
	Data Distribution: 271	Data Distribution:	Data Distribution: 285	Data Distribution: 92

			44					
	Ave	std	Aver	std	Ave	std	Ave	std
The pulse rate of the arterial blood pressure	100.66	12.20	98.44	11.81	73.32	9.59	77.15	13.93
to diastolic arterial blood pressure	64.18	9.6	27.89	14.46	63.76	10.08	59.01	7.93
systolic arterial blood pressure	118.81	15.46	56.73	12.23	129.29	15.13	113.38	16.02
mean arterial blood pressure	82.65	11.14	43.09	7.26	86.90	12.20	77.21	9.62
Heart Rate	99.18	17.49	98.50	21.69	74.47	10.79	83.76	17.76
the pulse oximetry oxygen saturation level	94.21	4.44	93.11	6.2	95.52	2.73	94.27	3.93
Body Temperature	36.17	1.27	35.90	2.72	36.09	1.16	28.83	3.80
Day of clinical event	10.60	8.00	16.45	11.22	7.94	6.52	4.98	4.15

In this phase, we extracted the minimum and maximum values for each variable within every cluster. These extremal values serve as proximate references to be applied to the original dataset. By filtering the original dataset based on these boundaries, each data point is assigned to its respective cluster.

Table 11 presents the minimum and maximum values of each variable across all clusters, offering a comprehensive overview of the distribution and range within each cluster.

Based on that, the subsequent section of the table details the minimum (min) and maximum (max) values for various physiological parameters within each cluster. For instance, parameters such as pulse rate, arterial blood pressure, heart rate, oxygen saturation level, body temperature, and the day of the clinical event are captured. These min-max values offer a comprehensive view of the range and variability of each parameter within the respective clusters. For example, in CLUSTER 0, the pulse rate of arterial blood pressure ranges from 70.54 to 144.32, providing insights into the dispersion of this physiological metric within that specific cluster. This detailed breakdown facilitates a nuanced understanding of how clusters differ in terms of physiological characteristics. Also, to facilitate mapping and cluster assignments (next phase).

Table 11: Cluster references

Indicator – Numerical (692 rows, 8 columns)	**CLUSTER 0**		**CLUSTER 1**		**CLUSTER 2**		**CLUSTER 3**	
	Silhouette Score: 0.23		Silhouette Score: 0.12		Silhouette Score 0.26		Silhouette Score: 0.36	
	Data Distribution: 271		Data Distribution: 44		Data Distribution: 285		Data Distribution: 92	
	min	max	min	max	min	max	min	max
the pulse rate of the arterial blood pressure	70.54	144.32	70.34	123.46	42.69	101.43	55.27	124.39
diastolic arterial blood pressure	48.80	138.37	10.83	53.50	37.01	101.60	40.51	88.03
systolic arterial blood pressure	82.88	173.66	45.15	96.03	98.27	179.98	74.60	164.71

mean arterial blood pressure	51.20	143.86	31.65	61.22	67.61	165.66	58.62	108.86
heart rate	47.71	144.65	57.38	149.02	49.63	116.32	54.46	133.74
pulse oximetry oxygen saturation level	49.18	99.46	71.31	98.75	77.85	99.54	65.00	99.03
body temperature	30.65	39.00	23.53	38.67	29.64	37.91	16.74	32.63
day of the clinical event	1	37	1	39	1	29	1	19

5.4.4 Results and Discussion

This performance represents a significant stride in the continuous development of the IDSS4PM framework. Based on Figure 55, a crucial aspect of adopting a data-driven approach was the introduction of standardization, particularly through the creation of custom codes for categories lacking clinical codes. This transformative step not only enhanced data quality but also facilitated data integration and clustering techniques.

Furthermore, leveraging the CEid to extract the day of the event proved instrumental in grouping, aggregating, and implementing temporal clustering. By incorporating process numbers and aggregating based on the day of the clinical event, we successfully integrated datasets to apply clustering techniques. The extraction of homogeneous data points enabled a comprehensive study of the behavior of similar data, allowing for the identification of potential patterns. Additionally, extracting the minimum and maximum values of each variable in each cluster provided us with cluster references crucial for subsequent analysis.

This experimental phase marked a significant advancement in health profiling. The utilization of cluster references and their mapping onto the main dataset in the next phase will further enhance our understanding of the intricacies within the data, particularly in identifying anomalies or outliers beyond established boundaries.

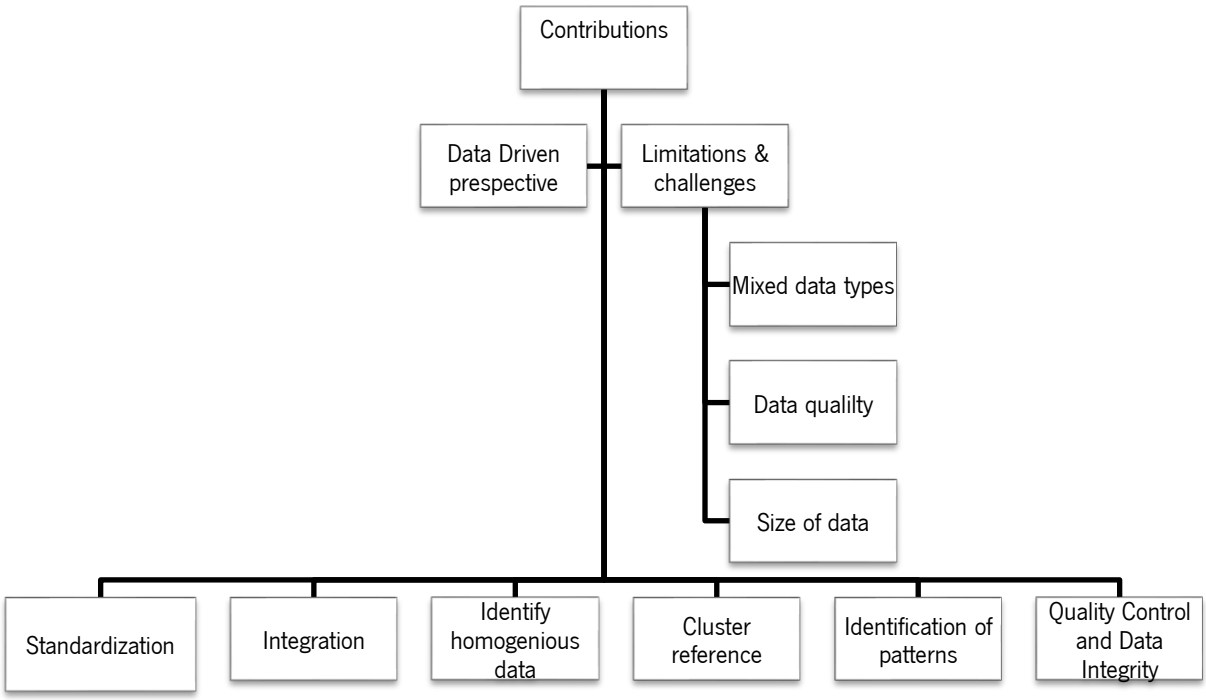


Figure 55: Contributions to IDSS4PM; Clustering

5.5 Cluster-Based Data Mapping and Assignment

This phase encompasses crucial steps aimed at grouping similar data points, both categorically and numerically. As illustrated in Figure 56, our initial approach involves leveraging cluster references for data extraction, followed by assigning clusters and data mapping. Subsequently, we meticulously analyze each cluster to gain insights and draw meaningful conclusions.

Figure 56: Cluster-based data mapping and cluster assignment

5.5.1 Leveraging Cluster References, Extraction and Mapping

In this pivotal phase, our focus was on translating the identified clusters into actionable insights within t he original dataset. Leveraging the minimum and maximum values established for each numerical varia

ble in every cluster, we meticulously mapped and assigned rows to their corresponding clusters. By emp loying these cluster references, we extracted pertinent data points from the original dataset and systema tically tagged each entry with its designated cluster number.

For the critical task of mapping clusters onto the original dataset—comprising both numerical and categorical data—we employed a unique approach. By determining cluster references through the minimum and maximum values of each variable within each cluster, we not only surmounted the challenge of clustering mixed data types but also successfully addressed the grouping of columns with 50% missing data, previously omitted during clustering.

This pioneering approach emerged as a cornerstone, seamlessly accommodating the diversity inherent in clinical data types—both categorical and numerical. The final step involved a thorough observation and analysis of each profile, offering a comprehensive understanding of the distinctive characteristics encapsulated within every cluster.

5.5.2 Identify Homogeneous Data

In the current phase of our study, we have successfully mapped and assigned clusters to a dataset comprising 1,631,242 records, encompassing 19 variables. This comprehensive analysis allows us to understand the distribution of data across different clusters, providing valuable insights into distinct health patterns.

Based on that the total size of mapped clusters is 1,631,242 rows, 19 variables. The individual Cluster Breakdown shows:

- Cluster 0: 598,266 rows
- Cluster 1: 110,042 rows
- Cluster 2: 593,192 rows
- Cluster 3: 325,289 rows

It's noteworthy that, within the total dataset, there are 4,453 unclustered data points. These unclustered data points represent instances that may not align with the identified patterns in the current clustering approach or could have missing values that need further investigation.

This cluster distribution summary provides a clear overview of the sizes and composition of each cluster, setting the stage for more detailed analyses and insights into the specific characteristics associated with each health pattern. Additionally, attention to unclustered data points allows for a more comprehensive understanding of the dataset and potential areas for refinement in future analyses.

5.5.3 Cluster Overview and Analysis

The bar chart, in Figure 57, shows the average value of each indicator based on cluster number. The key observation shows that:

Cluster 0

- Pulse Rate: High (111.04 beats per minute)
- Diastolic Arterial Blood Pressure: Moderate (64.57 mmHg)
- Mean Arterial Blood Pressure: Moderate (77.92 mmHg)
- Systolic Arterial Blood Pressure: High (108.65 mmHg)
- Pulse Oximetry Oxygen Saturation Level: Normal (96.04%)
- Body Temperature: Elevated (36.52°C)
- Heart Rate: High (107.05 beats per minute)
- Glasgow Coma: Low (12.27)
- Sepsis: Absent
- Day of Clinical Event: 1.06

Cluster 1:

- Pulse Rate: Moderate (90.43 beats per minute)
- Diastolic Arterial Blood Pressure: Low (35.65 mmHg)
- Mean Arterial Blood Pressure: Low (42.80 mmHg)
- Systolic Arterial Blood Pressure: Low (55.58 mmHg)
- Pulse Oximetry Oxygen Saturation Level: Normal (73.15%)
- Body Temperature: Low (27.02°C)
- Heart Rate: Moderate (77.68 beats per minute)
- Glasgow Coma: Moderate (5.33)
- Sepsis: Absent
- Day of Clinical Event: 1.09

Cluster 2:

- Pulse Rate: Low (74.88 beats per minute)
- Diastolic Arterial Blood Pressure: Moderate (64.98 mmHg)
- Mean Arterial Blood Pressure: Moderate (85.31 mmHg)
- Systolic Arterial Blood Pressure: High (128.07 mmHg)
- Pulse Oximetry Oxygen Saturation Level: Normal (95.41%)
- Body Temperature: Elevated (36.53°C)

- Heart Rate: Moderate (75.21 beats per minute)
- Glasgow Coma: High (14.52)
- Sepsis: Present (49.97% probability)
- Day of Clinical Event: 1.22

Cluster 3:

- Pulse Rate: Moderate (79.01 beats per minute)
- Diastolic Arterial Blood Pressure: Moderate (56.27 mmHg)
- Mean Arterial Blood Pressure: Moderate (70.16 mmHg)
- Systolic Arterial Blood Pressure: High (101.39 mmHg)
- Pulse Oximetry Oxygen Saturation Level: Normal (93.86%)
- Body Temperature: Elevated (30.01°C)
- Heart Rate: Moderate (80.84 beats per minute)
- Glasgow Coma: Moderate (14.11)
- Sepsis: Present (49.97% probability)
- Day of Clinical Event: 1.09

The analysis reveals pivotal observations (presented in Figure 57) indicating that each cluster exhibits unique patterns in health parameters, indicating potentially distinct health profiles. Cluster 0 has a high pulse rate, elevated body temperature, and high systolic arterial blood pressure, suggesting a more intense clinical event. Moreover, cluster 1 shows signs of lower blood pressure, lower heart rate, and lower body temperature, indicating a different health state compared to Cluster 0. Clusters 2 and 3 have relatively similar pulse rates but differ in other parameters, such as systolic arterial blood pressure and the presence of sepsis.

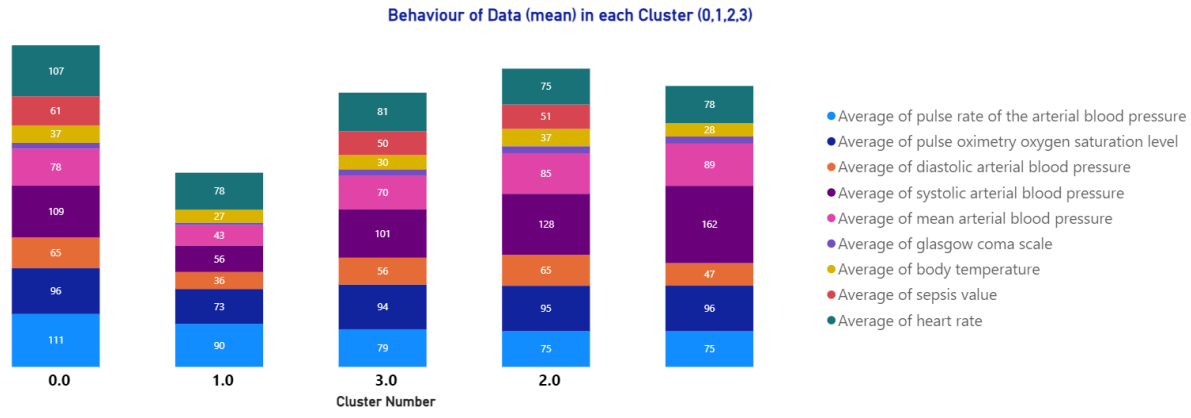


Figure 57: Behaviour of data in each cluster

5.5.4 Exploring Health Profiles

Analyzing the minimum, maximum, and mean variables in each cluster enhances the ability to understand the distribution of health parameters, identify characteristic patterns, and glean insights into the diverse health profiles represented within your dataset. These insights, in turn, contribute to more informed decision-making and the development of precise healthcare strategies.

Cluster 0:

- The day of the clinical event: Ranges from 1 to 37, with a mean of 1.06. Events are spread across the observed period.
- Glasgow Coma Scale: Varies between 3 and 15, with a mean of 12.27. Indicates a range of consciousness levels, potentially reflecting diverse patient conditions.
- the sepsis from 61 to 63, with a mean of 61.13. This variable shows minimal variability within the cluster.
- arterial blood pressure ranges from 70.54 to 136.56, with a mean of 111.04. Suggests a spectrum of blood pressure levels, potentially indicating varying severity of cardiovascular conditions.
- Systolic arterial blood pressure varies between 77.92 and 114.91, with a mean of 108.65. Indicates a range of systolic blood pressure levels.
- Mean arterial blood pressure ranges from 48.81 to 104.75, with a mean of 96.04. Indicates variations in overall blood pressure within the cluster.
- Diastolic arterial blood pressure varies from 31.27 to 38.01, with a mean of 64.57. Shows a range of diastolic blood pressure levels.
- Temperature: Consistent at a mean of 36.52, reflecting stability in body temperature within the cluster.
- Oxygen saturation levels range from 49.18 to 99.13, with a mean of 96.04. Indicates variability in oxygen saturation.
- Heart rate varies between 47.72 and 137.94, with a mean of 107.05. Shows a range of heart rates within the cluster.
- The laboratory results display a wide range, spanning from -0.6 to 88033, with an average value of 109.55. This considerable variability suggests a spectrum of diverse clinical conditions. However, for a more precise discussion, it is essential to consider the specific category of the laboratory test being analyzed, as will be further presented in subsequent discussions.

- Cluster 1

- Event Day: Similar to Cluster 0, with a mean of 1.09, indicating events spread across the observed period.

- Glasgow Coma Scale: Consistently low at 5.33, indicating a uniform level of consciousness within the cluster.
- the sepsis: Not provided, limiting insights into this variable.
- arterial blood pressure ranges from 45.15 to 88.72, with a mean of 90.43. Indicates a moderate range of blood pressure levels.
- Systolic arterial blood pressure varies between 31.65 and 60.13, with a mean of 55.58. Indicates a range of systolic blood pressure levels.
- Mean arterial blood pressure ranges from 10.83 to 53.51, with a mean of 27.02. Shows variability in overall blood pressure within the cluster.
- Diastolic arterial blood pressure varies from 23.54 to 38.64, with a mean of 35.65. Shows a range of diastolic blood pressure levels.
- Temperature: Consistent at a mean of 27.02, reflecting stability in body temperature within the cluster.
- Oxygen saturation levels range from 57.39 to 142.82, with a mean of 73.15. Indicates variability in oxygen saturation.
- Heart rate varies widely from 0 to 59065.7, with a mean of 77.68. The wide range may be due to potential outliers.

- Cluster 2

- Day of clinical event: Similar to Clusters 0 and 1, with a mean of 1.09, indicating events spread across the observed period.
-Glasgow Coma Scale: Varies between 7 and 15, with a mean of 14.52. Indicates a higher level of consciousness within the cluster.
- The sepsis: Ranges from 28 to 66, with a mean of 51.18. Shows variability in this variable within the cluster.
- arterial blood pressure ranges from 42.69 to 101.35, with a mean of 79.01. Indicates a moderate range of blood pressure levels.
- Systolic arterial blood pressure varies between 67.61 and 149.12, with a mean of 101.39. Indicates a range of systolic blood pressure levels.
- Mean arterial blood pressure ranges from 37.01 to 98.28, with a mean of 93.86. Shows variability in overall blood pressure within the cluster.
- Diastolic arterial blood pressure varies from 31.04 to 38.64, with a mean of 56.27. Shows a range of diastolic blood pressure levels.
- Temperature: Consistent at a mean of 30.01, reflecting stability in body temperature within the cluster.

- Oxygen saturation levels range from 57.85 to 142.82, with a mean of 80.84. Indicates variability in oxygen saturation.

- Heart rate varies between -1 and 44302.5, with a mean of 75.21. The wide range may be due to potential outliers.

- Cluster 3

- Day of clinical event: Like Clusters 0 and 1, with a mean of 1.22, indicating events spread across the observed period.

- Glasgow Coma Scale: Varies between 3 and 15, with a mean of 14.11. Indicates a higher level of consciousness within the cluster.

- The sepsis: Ranges from 39 to 97, with a mean of 49.97. Shows variability in this variable within the cluster.

- arterial blood pressure ranges from 55.28 to 121.17, with a mean of 74.88. Indicates a moderate range of blood pressure levels.

- Systolic arterial blood pressure varies between 58.62 and 103.86, with a mean of 128.07. Indicates a range of systolic blood pressure levels.

-The mean arterial blood pressure ranges from 43.49 to 84.62, with a mean of 95.41. Shows variability in overall blood pressure within the cluster.

- Diastolic arterial blood pressure varies from 16.75 to 32.63, with a mean of 64.98. Shows a range of diastolic blood pressure levels.

- temperature: Consistent at a mean of 36.53, reflecting stability in body temperature within the cluster.

- Oxygen saturation levels range from 86.72 to 99, with a mean of 95.41. Indicates variability in oxygen saturation.

- Heart rate varies between -0.35 and 49300, with a mean of 80.84. The wide range may be due to potential outliers.

Table 12: Analysing the behavior of data in each cluster

Numerical variables		C0	C1	C2	C3
the pulse rate of the arterial blood pressure	average	111.03	90.42	74.87	79.01
	minimum	70.54	70.34	42.69	55.27
	maximum	136.56	109.50	101.35	121.17
diastolic arterial blood pressure	average	64.56	35.64	64.98	56.27
	minimum	48.80	10.83	37.01	43.49
	maximum	104.75	53.50	98.27	84.61
systolic arterial blood pressure	average	108.64	55.57	128.06	101.38
	minimum	82.88	45.15	98.27	74.60

	maximum	172.13	88.72	178.96	158.90
mean arterial blood pressure	average	77.92	42.80	85.31	70.51
	minimum	51.20	31.65	67.61	58.62
	maximum	114.90	60.12	149.11	103.85
pulse oximetry oxygen saturation level	average	96.04	73.14	95.40	93.85
	minimum	49.18	71.31	77.85	86.71
	maximum	99.13	97.93	99.46	99.00
Heart rate	average	107.04	77.68	75.20	80.84
	minimum	47.71	57.38	49.63	54.46
	maximum	137.93	142.82	114.80	131.86
Body temperature	average	36.52	27.02	36.53	30.01
	minimum	31.27	23.53	31.03	16.74
	maximum	38	38.63	37.90	32.62
Day of clinical event	average	1.05	1.09	1.21	1.08
	minimum	1	1	1	1
	maximum	37	39	36	19
Glasgow Coma Scale	average	2.6	5.33	14.51	14.10
	minimum	3.00	5.33	7.00	3.00
	maximum	15.00	5.33	15.00	15.00

While the comprehensive interpretation of physiological indicators necessitates domain knowledge and expertise to optimize their utility in clinical decision-making, a preliminary analysis provides an overarching glimpse into the trends and variations across distinct clusters. This preliminary assessment can serve as a foundation for further, more nuanced investigations, acknowledging the importance of domain-specific insights for a more refined understanding.

5.6 Insight Into The Data Behavior Within Clusters

Below visual presentations below (Figures 58 to 67) demonstrate a general view of the mean, maximum, and minimum values associated with each numerical indicator based on each cluster. Analyzing the patterns and trends across the clusters reveals valuable insights into the characteristics of each group:

5.6.1 Glasgow Coma Scale (GSC)

Based on Figure 58, Clusters 2 consistently exhibit higher mean GCS values (15),

Cluster 1, with mean GCS values of 5, represents groups with relatively low consciousness levels.

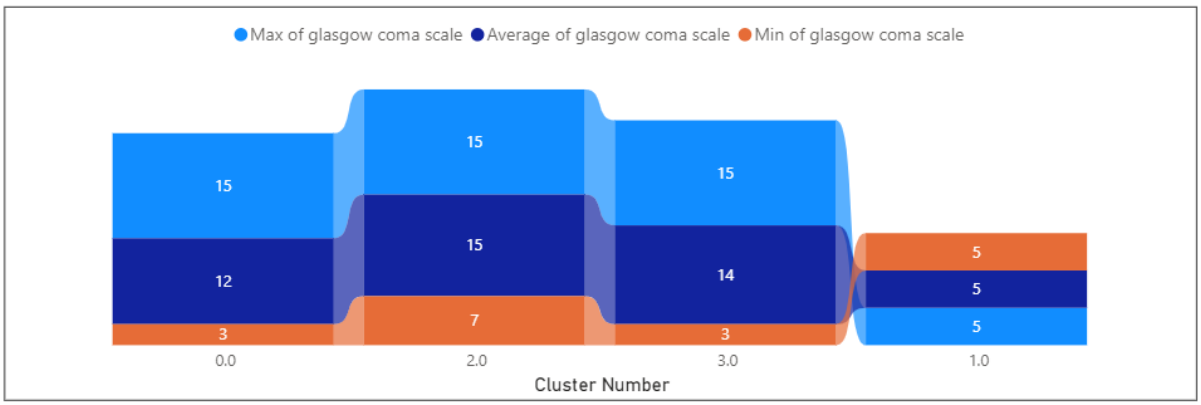


Figure 58: Behaviour of Glasgow coma in each cluster

5.6.2 Systolic Arterial Blood Pressure

According to Figure 59, cluster 2 exhibits the highest values for the minimum, maximum, and average systolic arterial blood pressure, while cluster 1 demonstrates the lowest values for these parameters.

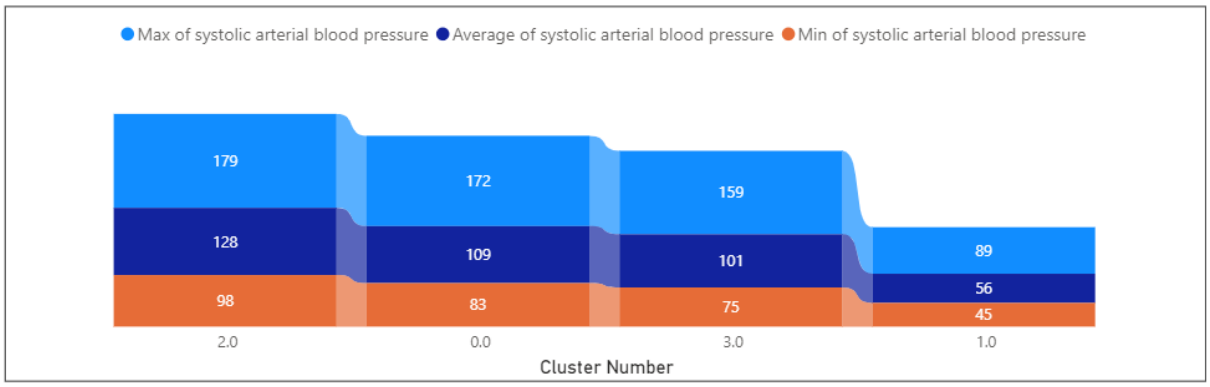


Figure 59: Behaviour of systolic arterial blood pressure in each cluster

Likewise, the mean arterial blood pressure is highest in cluster 3 and lowest in cluster 1, as illustrated in Figure 60.

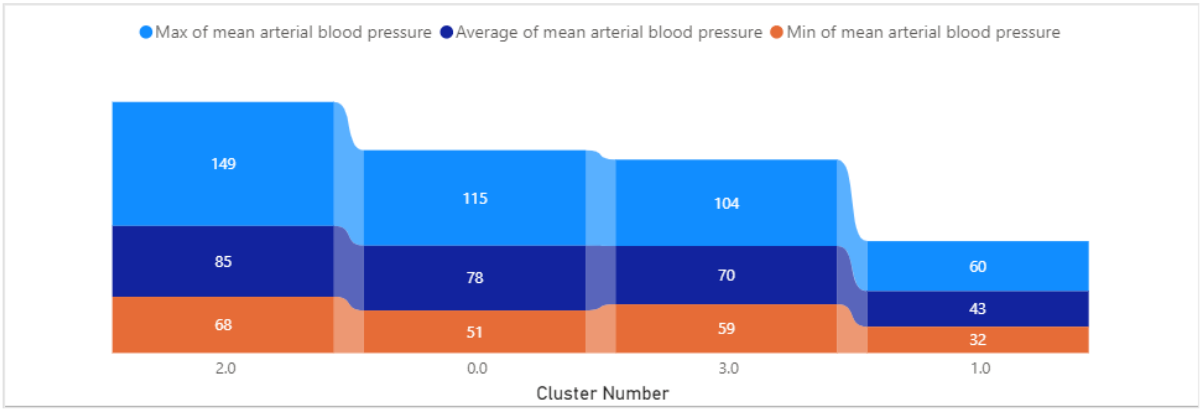


Figure 60: Behaviour of arterial (mean) blood pressure in each cluster

Furthermore, Figure 61 shows that the pulse rate of arterial blood pressure demonstrates its highest values (maximum, minimum, mean) within cluster 0, while the lowest values are observed in cluster 1.

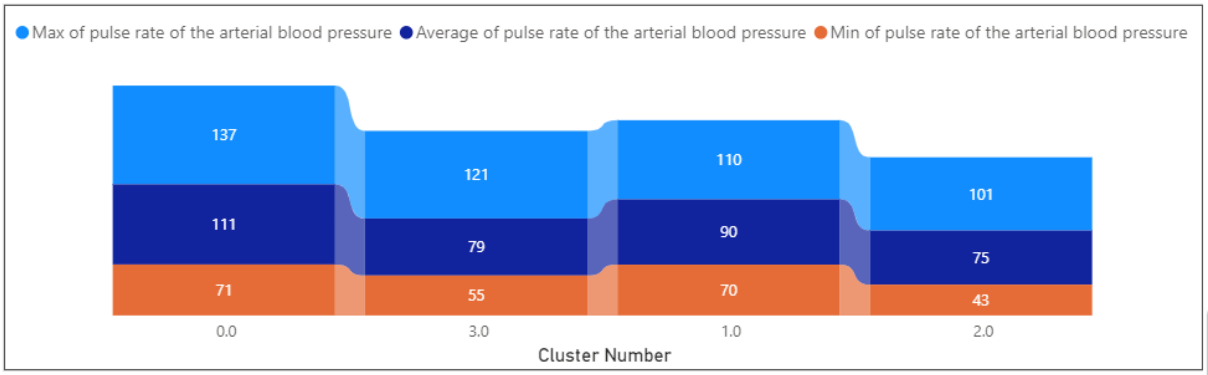


Figure 61: Behaviour of pulse rate of arterial blood pressure in each cluster

In Figure 62, Cluster 0 exhibits the highest minimum, maximum, and average values of diastolic arterial blood pressure, while these values are notably lower in Cluster 1.

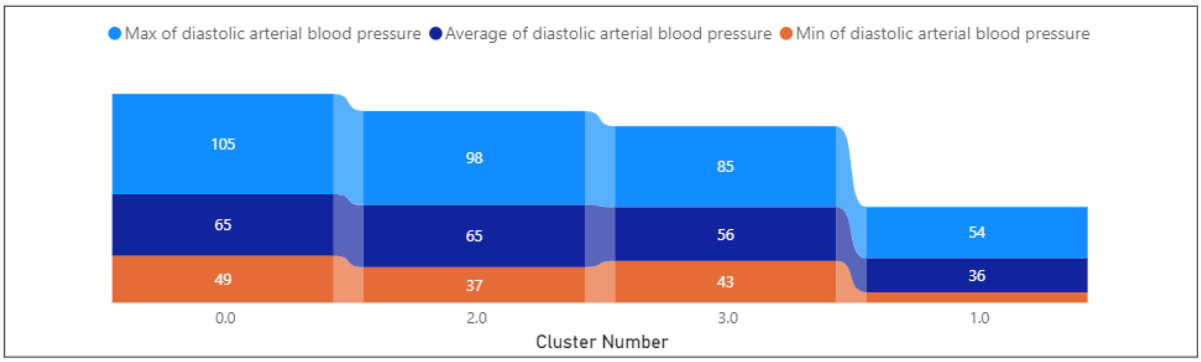


Figure 62: : Behaviour of diastolic arterial blood pressure in each cluste

5.6.3 Oxygen Saturation

Based on Figure 63, clusters 0, 2, and 3 display elevated maximum oxygen saturation levels (99), with cluster 0 boasting the highest average value (96) and the lowest minimum value (49).

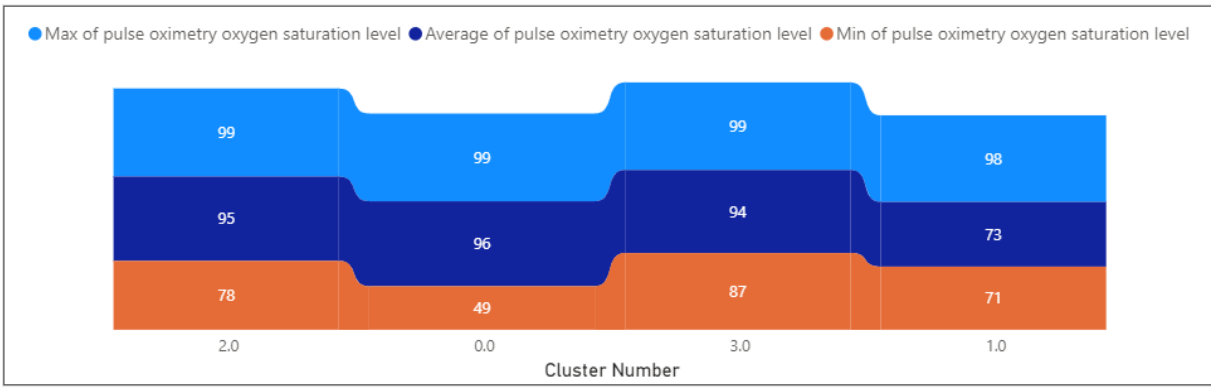


Figure 63: Behaviour of oxygen saturation level in each cluster

5.6.4 Heart Rate

In Figure 64, it's notable that Cluster 1 exhibits a higher mean heart rate of 143, along with the highest average heart rate of 107. Conversely, Cluster 0 displays the lowest minimum heart rate at 48, suggesting heightened cardiovascular activity within this cluster.

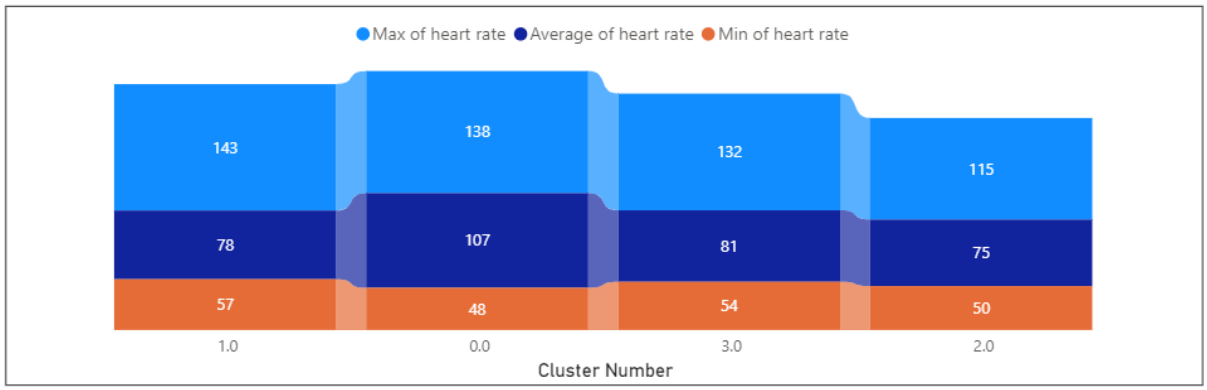


Figure 64: Behaviour of heart rate in each cluster

5.6.5 Temperature

Temperature remains relatively stable across clusters 0 and 2. However, according to Figure 65, cluster 1 displays the highest maximum value of body temperature (39), while cluster 3 exhibits the lowest minimum body temperature at 17.

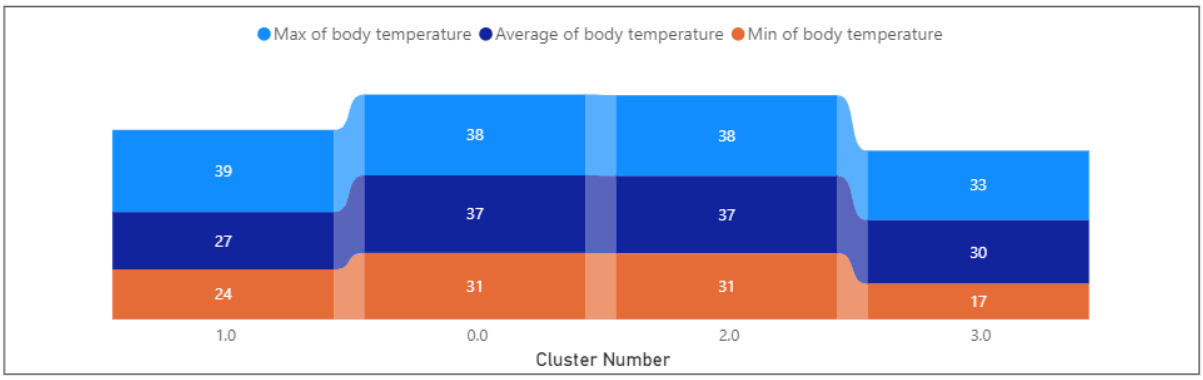


Figure 65: Behaviour of body temperature in each cluster

5.6.6 Sepsis

From Figure 66, it's evident that cluster 0 displays minimal variability in sepsis (ranging from 61 to 63), hinting at a relatively uniform response to sepsis within this group. On the other hand, cluster 2 exhibits moderate variability (ranging from 28 to 66), indicating differing degrees of sepsis conditions within the cluster. Additionally, cluster 3 demonstrates the highest maximum value of sepsis at 97, while the lowest minimum value is found in cluster 2 at 28.

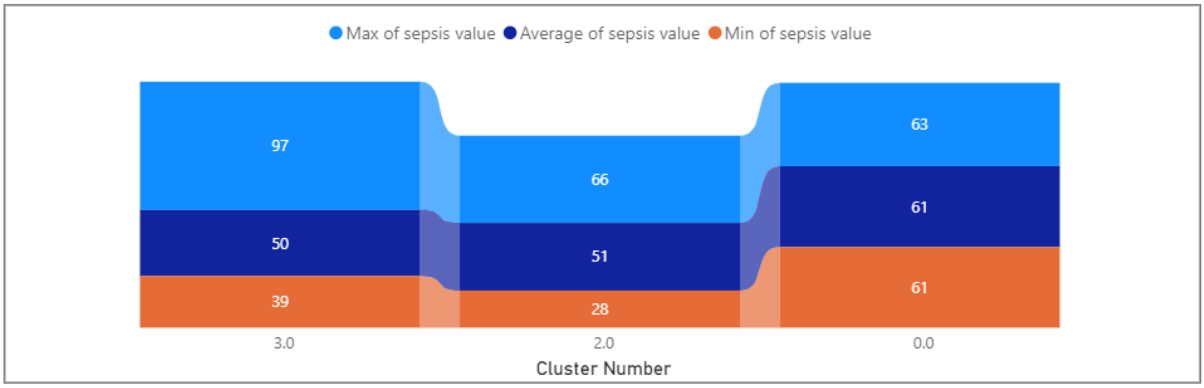


Figure 66: Behavior of sepsis in each cluster

5.6.7 Days of Clinical Events

The average day of clinical events across all clusters is approximately 1.1, suggesting a relatively even distribution of events throughout the observed period. However, as depicted in Figure 67, cluster 1 exhibits the highest number of maximum days associated with clinical transactions (39 days), while cluster 3 has the least (19 days).

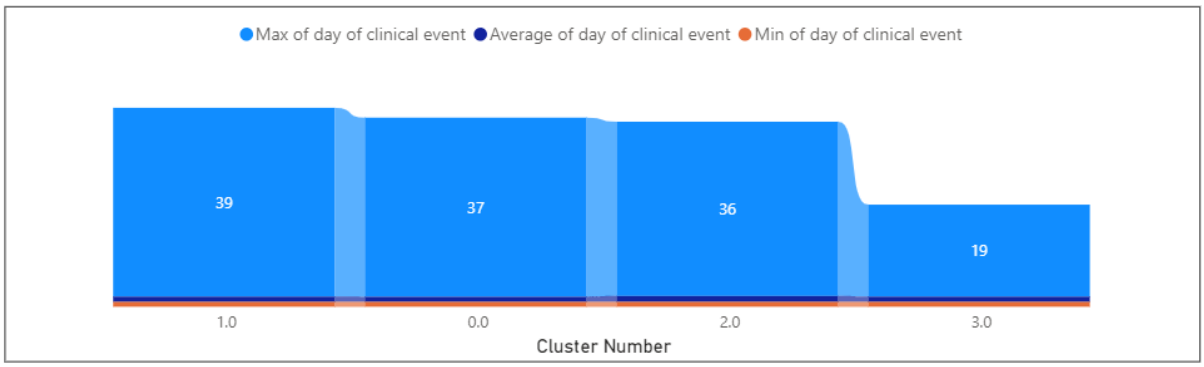


Figure 67: Analysing the total days of clinical transactions in each cluste

In summary, considering the vital signs, Clusters 0 and 3 appear to represent groups with more critical conditions, as indicated by higher GCS values, elevated blood pressure, and a wider range of laboratory results. Moreover, Cluster 1 signifies a group with consistently lower GCS, moderate blood pressure, and wider variability in heart rate and laboratory results. Additionally, Cluster 2 reflects a mix of conditions, with higher GCS, moderate blood pressure, and variability in oxygen saturation and sepsis indicators.

These trends provide a foundation for further investigation and highlight the need for domain-specific knowledge to interpret the clinical significance of the observed patterns. Additionally, outlier detection and data validation are crucial for a more accurate understanding of the dataset.

5.6.8 Temporal Distribution Of Clusters

One of the paramount advantages of temporal analysis lies in its capability to delve into the behavioral patterns of clusters over time. In this context, proposing the utilization of CEid facilitates the extraction of clinical event data daily during the clustering phase-processing. This step significantly advances the identification of patterns and trends over time, offering a more comprehensive understanding of temporal dynamics.

Figure 68 illustrates the temporal distribution of each cluster. For instance, cluster 2 is observed on all days except days 37 and 39. Another noteworthy observation is that after day 34, a singular cluster is exclusively assigned to that specific day. Additionally, cluster 3 is assigned only during the initial 19 days. In summary, temporal analysis in the context of cluster distribution provides a dynamic view of patients' health trajectories, aiding precision medicine by enabling timely interventions, enhancing predictive modeling, and fostering a nuanced understanding of the temporal aspects of clinical data.

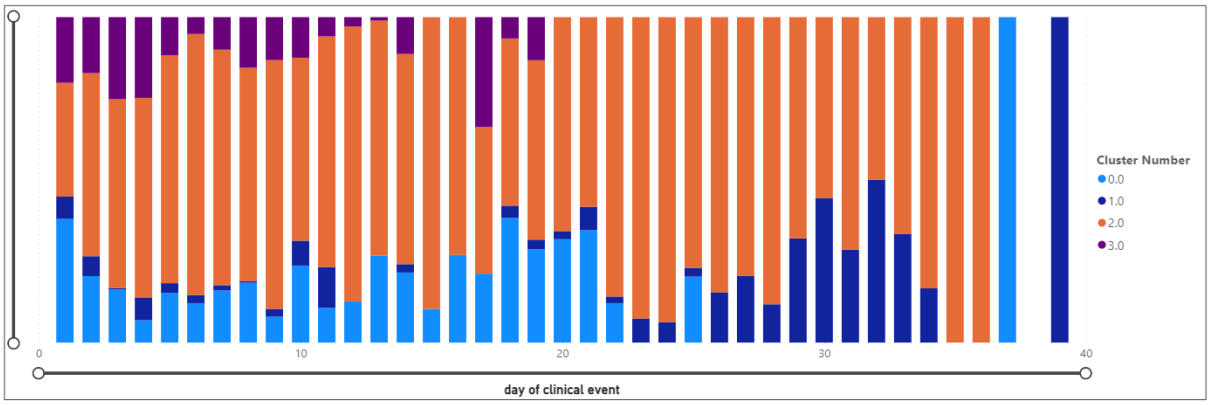


Figure 68: Temporal distribution of each cluster

5.6.9 Laboratory Exams and Results

The bar chart displayed in Figure 69 offers a comprehensive glimpse into the variety of laboratory examinations within each cluster, detailing the results as either normal or abnormal. The legends conveniently provide the codes for each examination. Hence, this visualization and analysis offer the opportunity to explore and examine the union, intersection, and distinctions of each cluster in terms of laboratory codes. This analytical insight holds paramount importance for decision-makers seeking to discern the distinct behavioral patterns of each cluster.

For instance, within this analysis, cluster 1 exhibits the lowest count of examinations yielding abnormal results, suggesting a relatively healthier profile. On the other hand, Cluster Two stands out with the highest number of examinations producing both abnormal and normal results. Such granular insights enable decision-makers to better understand and interpret the diverse dynamics within each cluster, facilitating more informed and targeted decision-making processes.

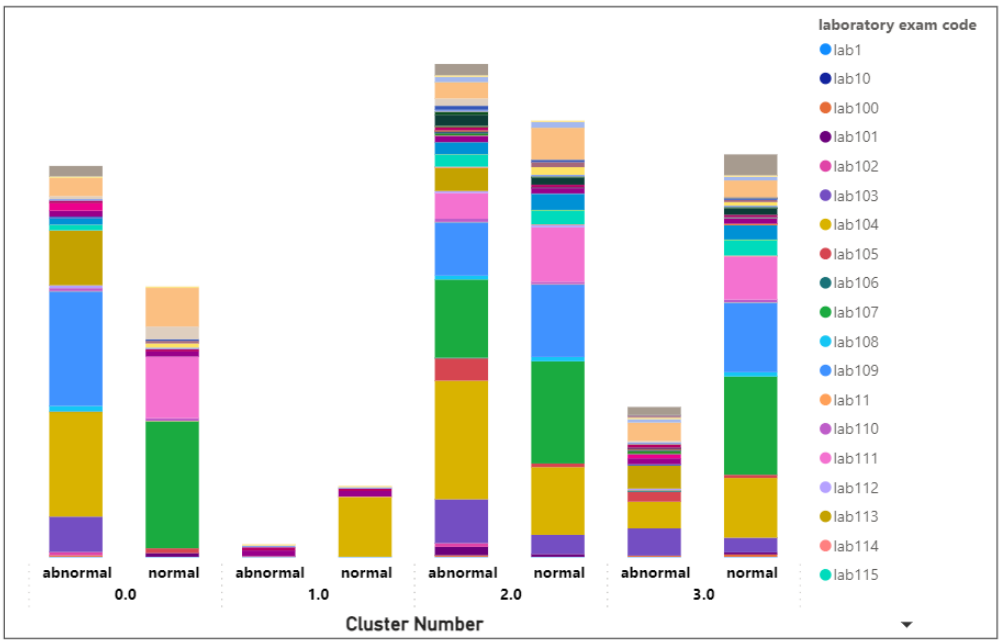


Figure 69: Behaviour of laboratory exams & results in each cluster

5.6.10 Interventions

Figure 70 provides a visual representation of the exploration into the union, intersection, and distinctions of interventions within each cluster. Remarkably, 'int9' stands out as the most recurrent intervention, observed across all clusters.

In detail, the x-axis corresponds to the intervention codes, the y-axis signifies the number of interventions, and the legend elucidates the cluster types under consideration. This graphical representation enhances our understanding of the intervention landscape, emphasizing the prevalence and distribution of intervention types across various clusters.

For instance, the “int9” is observed in all clusters.

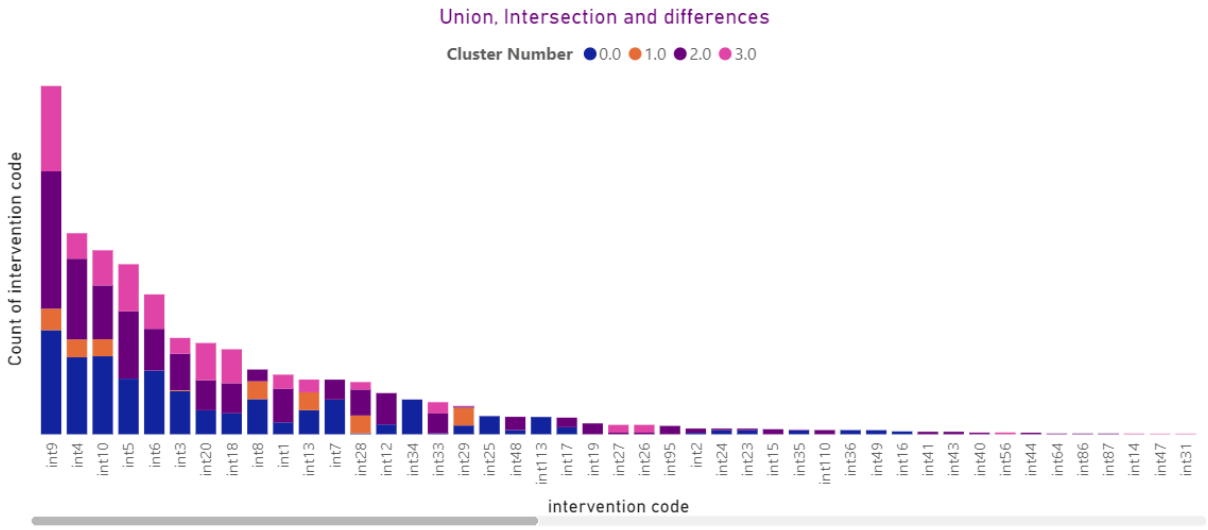


Figure 70: Distribution of interventions (codes) in each cluster

Likewise, the chart depicted in Figure 71 showcases diagnostic codes associated with their respective diagnostic types (x-axis) and their distribution across clusters. Notably, the diagnostic code 8271 is exclusively linked to cluster Zero, indicating a specific association within this cluster.

5.6.11 Diagnostic

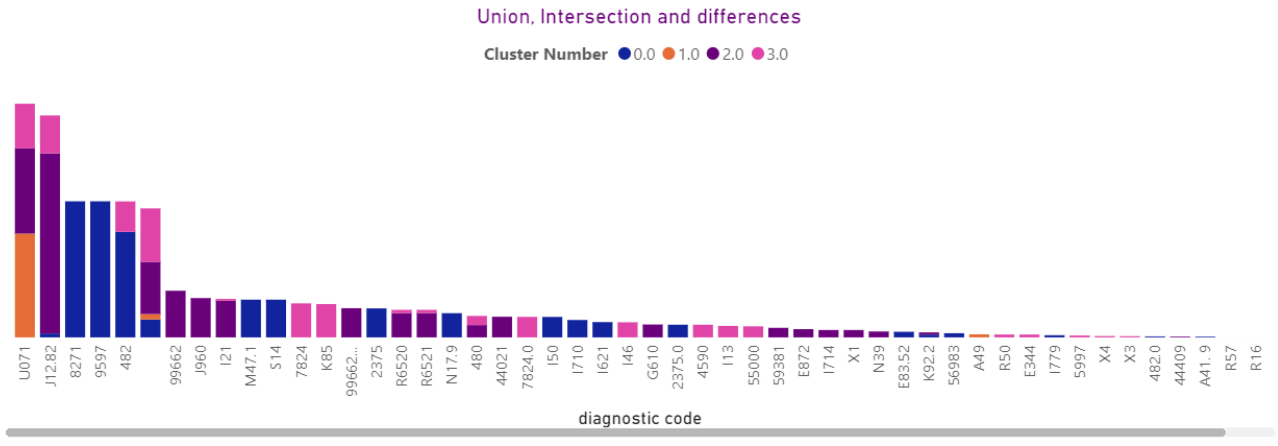


Figure 71: Distribution of diagnostic (codes) in each cluster

5.6.12 Medications

In the concurrent analysis, as portrayed in Figure 72, the visual representation delineates the distribution of medications across each cluster. While the bulk of medications is allocated to all four clusters,

discernible differences emerge in the total count of each medication type (on the x-axis) assigned to individual clusters.

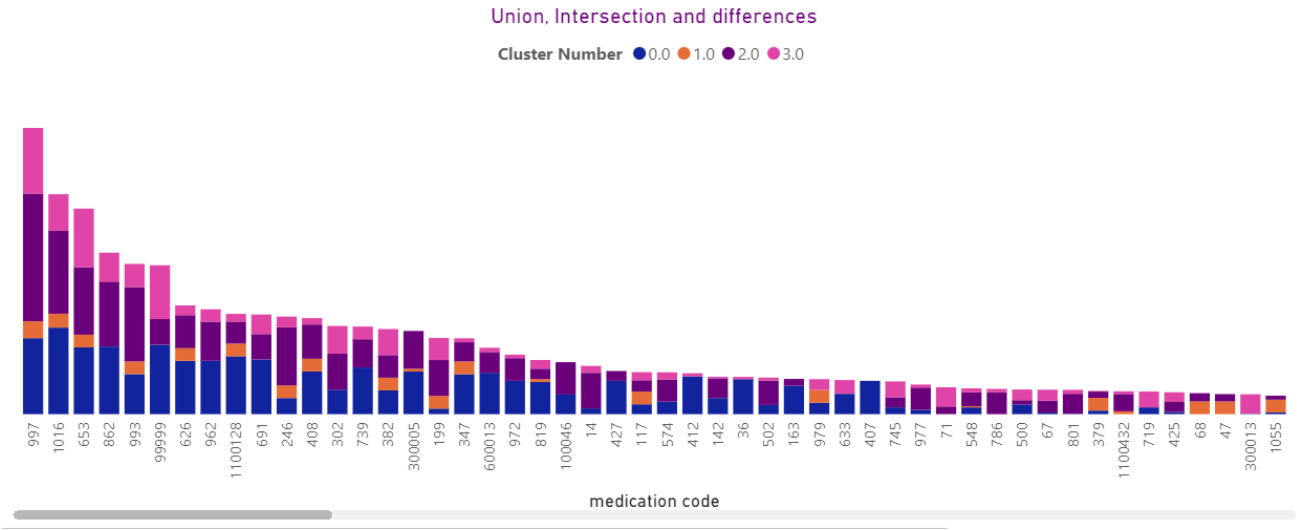


Figure 72: Distribution of medications (codes) in each cluster

5.6.13 Procedures (local, Zona)

The two Figures, 73 and 74, depict the categorization of procedures into local and Zona types across each cluster. Notably, "prcl1" is exclusively associated with clusters zero and two, while "prcl10" and "112" are present solely in cluster zero. Likewise, concerning the Zona procedures, "prcz3" is specifically assigned to cluster 3.

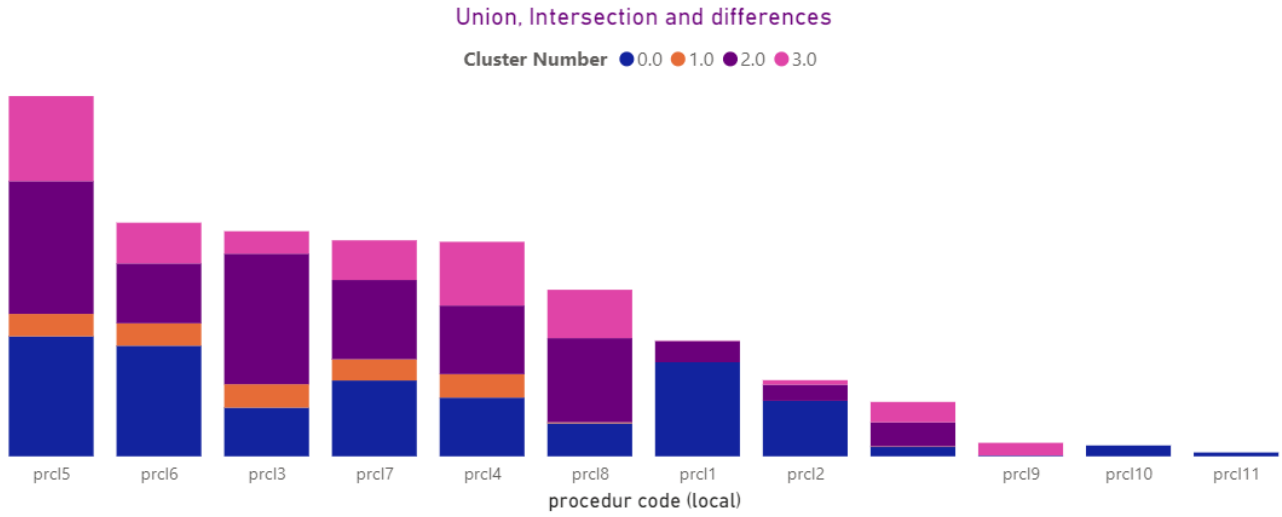


Figure 73: Distribution of procedure-local (codes) in each cluster

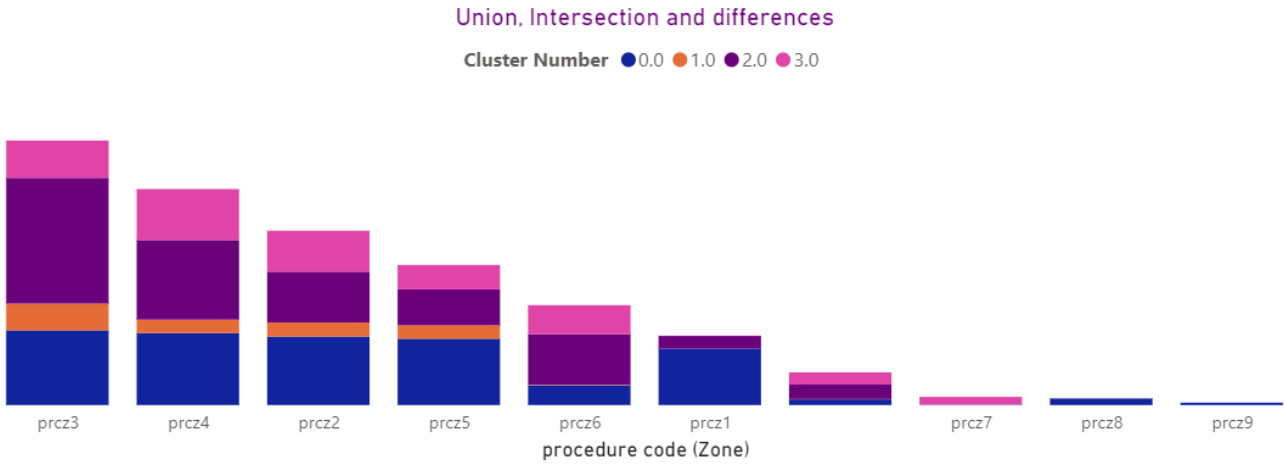


Figure 74: Distribution of procedure-zona (codes) in each cluster

5.6.14 Results and Discussion

This phase marks a crucial advancement in the clustering process. Figure 76 illustrates the contributions to the clinical decision support system by identifying similar data behaviors. Leveraging the cluster references acquired from the preceding step, we extract data from the original dataset, which encompasses diverse types of clinical data. This extraction facilitates the assignment of the whole data to the relevant cluster. In this experimental context, our approach yields grouped categorical and numerical data. Consequently, the identification of trends and analysis of data behavior in a temporal context represent sophisticated strides toward data-driven decision-making.

The temporal analysis of cluster distribution is crucial for precision medicine for several reasons. By examining how specific clusters manifest over time, we gain insights into the temporal dynamics of clinical events and their associations. This temporal perspective allows us to identify patterns, trends, and potentially critical transitions in patients' health journeys.

In precision medicine, understanding the evolution of clusters helps healthcare professionals tailor interventions based on the timing of specific clinical events. For example, if a particular cluster is prevalent during a specific time window, it may indicate a critical phase in a patient's condition that requires targeted and timely interventions.

Moreover, temporal analysis enhances the accuracy of predicting and assigning new data to relevant clusters. It enables healthcare providers to recognize patterns that may be indicative of disease progression, treatment effectiveness, or other important factors influencing patient outcomes. This, in turn, contributes to more informed decision-making and personalized treatment strategies.

The analysis presented above marks a significant stride in advancing clinical decision-making, leveraging both a data-driven perspective and its application in precision medicine. The exhaustive breakdown of categorical data, including medications, procedures, laboratory results, diagnostics, and more, linked to each cluster, imparts a nuanced understanding of the diverse clinical pathways embedded in the data. This level of granularity equips healthcare professionals with insights to comprehend the intricacies of patient conditions and treatment patterns.

Furthermore, visualizing the relationships between the union, intersection, and differences of codes across clusters unveils distinct patterns and trends within the data. This visualization aids in more informed decision-making, drawing on historical data to understand the associations between various factors.

Moreover, the identification of specific clinical types unique to certain clusters facilitates tailored and targeted interventions. Healthcare providers can utilize this detailed information to craft personalized treatment plans aligned with the specific needs and characteristics of each patient subgroup.

In essence, this analytical approach empowers clinicians with a data-driven comprehension of the intricate relationships between available historical data and patient clusters. By providing insights into both historical trends and the specific needs of patient subgroups, this method aligns closely with the principles of data-driven healthcare and precision medicine, fostering a more refined and personalized approach to clinical decision-making.

Building on these insights, the clinical decision-maker gains the ability to allocate new data into specific clusters, utilizing the outcomes of this clustering process. Furthermore, in this phase, we advanced

predictive modeling techniques to forecast the most suitable cluster for new data, ensuring a more accurate and tailored approach to decision support.

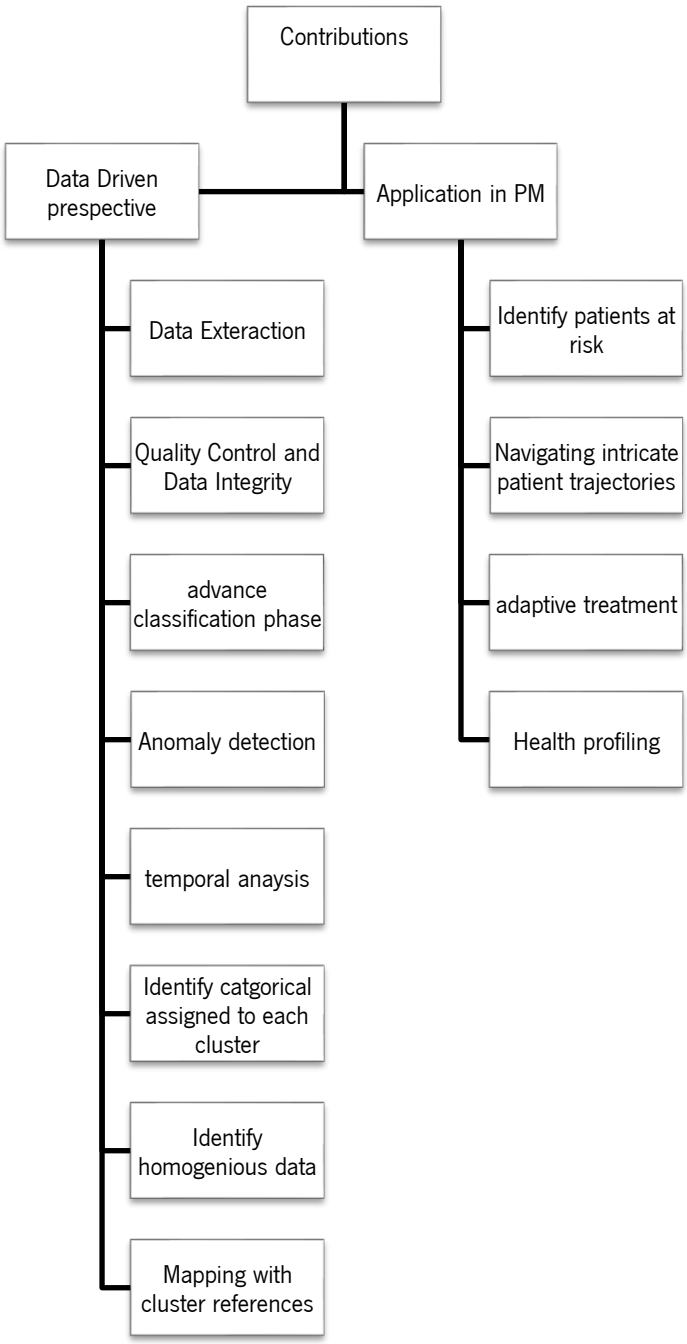


Figure 75: Contributions to IDSS4PM; Analytical Insight

5.7 Predicting Modeling through Classification

5.7.1 Classification

In this phase, we implement predictive health profiling by leveraging behavior-based classification on new data, building upon the insights gained from the temporal clustering performed in the previous phase. The objective is to apply a classification approach to train the model using 258,054 rows (after excluding missing data) and 18 predictors—a mix of categorical and numerical variables.

The predictors encompass a comprehensive range of clinical indicators, providing the capability to forecast the cluster number over the course of a clinical day. By training the classification algorithm on this dataset, we empower it to discern patterns and associations within the data, enabling the prediction of cluster numbers for each day of clinical performance.

This predictive health profiling methodology offers a dynamic and proactive approach to understanding patient trajectories, aiding healthcare professionals in anticipating and responding to potential developments in patient conditions. As new data becomes available, the classification algorithm can be applied to furnish timely predictions of cluster numbers, thereby contributing to a more informed and personalized healthcare decision-making process.

For example, we applied a random forest classifier with cross-validation techniques on a dataset of 258,054 records with 19 features. The goal was to predict the cluster using historical data and to discern the significance of each variable in the prediction process. Following the analysis, with cross-validation k=5, we obtained a mean accuracy of 0.97 and it was determined that "Temperature" was the most influential factor in predicting the cluster according to the random forest model.

The results obtained indicate the performance of your Random Forest classifier using cross-validation. Here's a summary:

a. Mean Accuracy: 0.946, with a standard deviation of 0.078. This indicates that, on average, the model correctly predicts the target class around 94.6% of the time.

Accuracy mentioned in equation 1, measures the proportion of correctly predicted instances among all instances.

$$Equation\ 1: Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Predictions}$$

Equation 1.Accuracy

b. Mean Precision: 0.961, with a standard deviation of 0.052. Precision measures the accuracy of the positive predictions. This result suggests that, on average, around 96.1% of the predicted positive cases are positive.

Precision, mentioned in equation 2, measures the proportion of true positive predictions among all positive predictions made by the model.

$$Equation\ 2: Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Equation 2.Precision

c. Mean Recall: 0.946, with a standard deviation of 0.078. Recall measures the ability of the model to correctly identify the positive cases. This value indicates that around 94.6% of the actual positive cases are correctly identified by the model.

Recall (also known as Sensitivity or True Positive Rate), mentioned in equation 3, measures the proportion of true positive predictions among all actual positive instances in the dataset.

$$Equation\ 3: Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

Equation 3.Recall

d. Mean F1 Score: 0.946, with a standard deviation of 0.077. The F1 score is the harmonic mean of precision and recall and provides a balance between them. This value suggests that the model achieves a good balance between precision and recall.

F1 Score, as mentioned in equation 4, is the harmonic mean of precision and recall, providing a balance between the two metrics.

$$Equation\ 4: F1\ Scoore = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Equation 4.F1Score

f. Mean Cohen's Kappa: 0.912, with a standard deviation of 0.126. Cohen's Kappa measures the agreement between predicted and actual class labels, considering the possibility of the agreement occurring by chance. This value indicates a substantial level of agreement beyond chance.

In equation 5, Cohen's Kappa measures the agreement between observed and predicted classifications, accounting for agreement occurring by chance. Po is the relative observed agreement and Pe is the hypothetical probability of chance agreement.

$$Equation\ 5: \text{Cohen's Kappa} = 2 * \frac{\text{Po} - \text{Pe}}{1 - \text{Pe}}$$

Equation 5.Cohen's Kappa

g. The mean AUC (Area Under the Curve) value of approximately 0.997 indicates that, on average, the classifier has an excellent ability to discriminate between different classes in the dataset. AUC values close to 1 suggest that the classifier can effectively separate positive and negative instances, making it highly reliable in distinguishing between classes.

The standard deviation of AUC, which is approximately 0.0056, represents the variability of AUC values across different folds of cross-validation. A lower standard deviation indicates that the AUC values from different folds are closer to the mean, suggesting consistency in the classifier's performance across various subsets of the data. Therefore, the high mean AUC value along with the low standard deviation indicates that the classifier performs consistently well across different cross-validation folds and demonstrates strong discriminatory power in distinguishing between classes.

The Area Under the Curve (AUC) is a metric used to evaluate the performance of a classification model. It represents the area under the Receiver Operating Characteristic (ROC) curve, which is a plot of the true positive rate (sensitivity) against the false positive rate (1 - specificity) for different threshold values. AUC provides a single scalar value that summarizes the model's ability to discriminate between classes, with higher values indicating better performance.

These metrics are commonly used to evaluate the performance of classification models, providing insights into different aspects of their predictive capabilities.

Overall, these results suggest that the Random Forest classifier is performing well in terms of accuracy, precision, recall, F1 score, and Cohen's Kappa.

Table 13: Model Performance

Metrics	Value (mean)
Accuracy	95.36%
Precision	0.961
Recall	0.946
Kappa	0.912
F1 score	0.946
AUC	0.99

5.7.2 Significance of Predictors

Based on the RF performance depicted in Figure 76, the key indicators for predicting the target (cluster number) are illustrated. Among them, "temperature," "Heart Rate," and "BLD_PULS_RATE_ART_ABP" emerge as the top predictors.

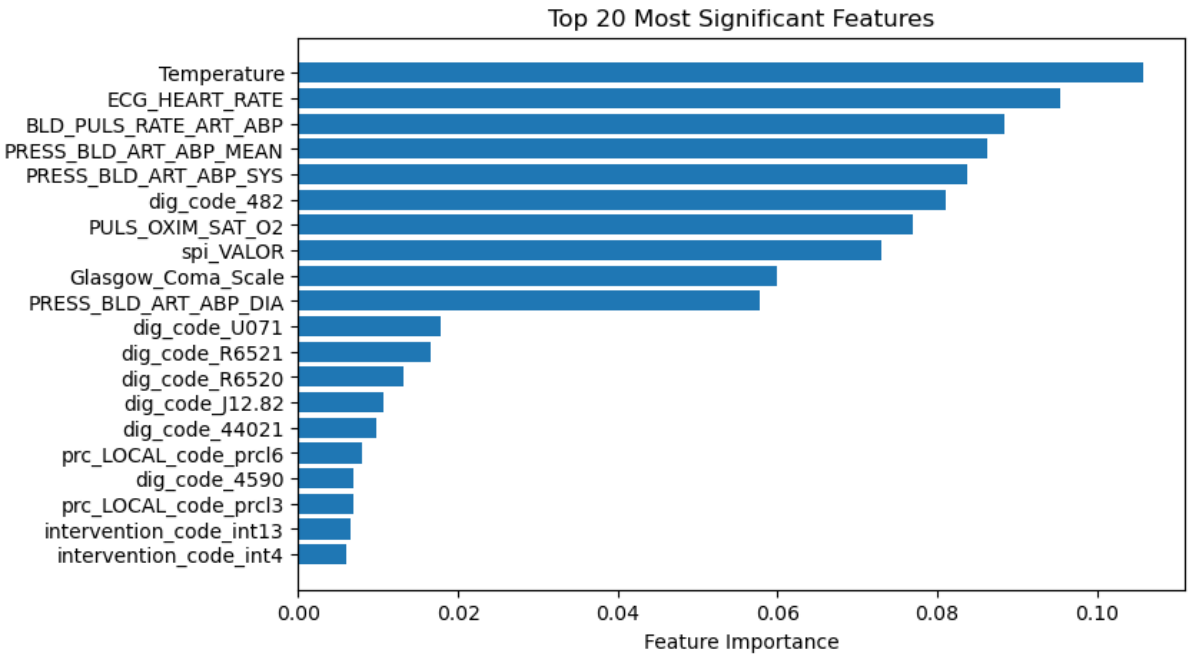


Figure 76: Significant features of RF

5.7.3 Results and Discussion

Predicting the cluster of new clinical data through classification methods represents a pivotal advancement in bolstering intelligent decision support systems for precision medicine. Figure 77 highlights the key contributions to IDSS4PM via predictive modeling (classification).

Classification models serve as invaluable tools in uncovering intricate patterns and associations within clinical data. This, in turn, facilitates the tailoring of treatment plans based on the unique characteristics of specific patient subgroups, leading to more personalized and effective interventions. Moreover, these models can unveil concealed relationships between clinical variables and health outcomes. Predicting the cluster of new data not only enables the early identification of potential health risks or complications but also empowers proactive interventions.

Furthermore, these methods significantly contribute to evidence-based decision-making by offering deep insights into patient clusters and expected outcomes. This wealth of information aids healthcare professionals in making informed choices for optimal patient care.

When it comes to continuous learning and adaptation, classification models exhibit a dynamic capability. They can be updated seamlessly with new data, enabling the system to adapt to evolving patterns in

clinical information. This continuous learning ensures that decision support systems remain pertinent and effective over time.

In terms of resource allocation optimization, understanding the probable cluster of a patient based on their clinical data becomes instrumental. This knowledge empowers healthcare providers to allocate resources—be it medical personnel, facilities, or specialized equipment—more efficiently. The result is a streamlined and targeted approach that aligns with the anticipated needs of diverse patient clusters.

In summary, the utilization of classification methods to predict clusters in clinical data elevates the capabilities of decision-support systems in precision medicine. This enhancement fosters a healthcare environment characterized by individualized, timely, and well-informed decision-making, ultimately leading to improved patient outcomes.

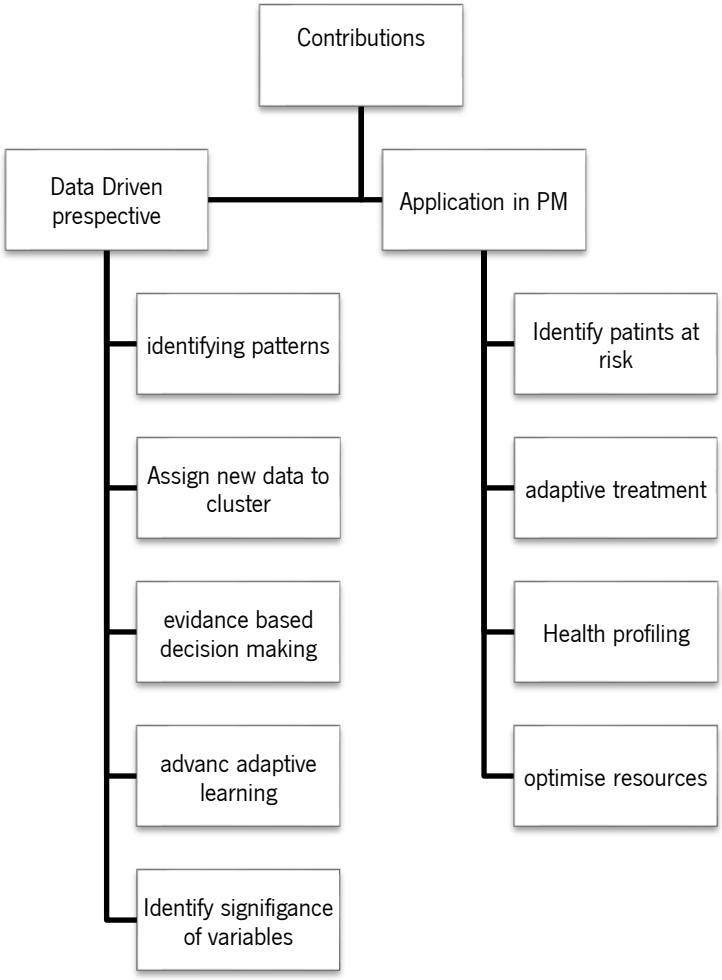


Figure 77: Contributions to IDSS4PM; Predictive analysis

6 CONTRIBUTIONS TO PRECISION MEDICINE VIA A DATA-DRIVEN APPROACH

In this chapter, the pivotal contributions to data-driven decision-making are highlighted, focusing on tailoring the most precise treatment strategies for individual projects (I and II). As it is presented in Figures 79 and 81, the analysis delves into the subject from multiple perspectives, including the utilization of data, its application in precision medicine, and its alignment with the optimal model of Simonian decision-making.

The significance of this performance is paramount, as it substantially enhances the landscape of decision-making by embracing a data-driven approach. This extension of impact reaches into practical applications, especially in fields where experts and clinical decision-makers are committed to achieving precision and optimality in the strategic planning and delivery of the most accurate treatments.

6.1 Data-Driven Psychotherapy Decision Support

i. Analytical Insight

Examining and closely observing the relationships between therapists and their clients during sessions presents a valuable opportunity to scrutinize the link between physiological and psychological indicators, as well as their mutual influence on the therapeutic alliance. Delving into the interrelated factors associated with both the client and therapist in each psychotherapy session offers a unique chance to discern patterns within the data. This meticulous analysis is crucial for identifying significant behavioral trends, which, in turn, holds substantial importance in customizing treatment approaches based on these observed patterns.

ii. Predictive Modelling

Integrating clinical and physiological data throughout sessions for both clients and therapists necessitates the application of highly sophisticated analytical models. While traditional statistical approaches enable the analysis of relationships between measures, they fall short in providing crucial insights into the specific temporal dynamics of the data, particularly physiological data, and the intricate interplay between different psychotherapy measures.

To overcome these limitations, leveraging recent advancements in computer science becomes imperative. The development of robust predictive models through the use of Data Mining (DM) and Machine Learning (ML) techniques has shown great promise in mental health care, specifically in the realm of psychotherapy.

These cutting-edge techniques not only accommodate the vast amounts of data available but also prove to be highly relevant in addressing the complexity of the psychotherapy process. By capturing the dynamic interactions between therapists and clients throughout the therapy sessions, data-driven analytical insights offer a nuanced understanding of the evolving nature of the therapeutic relationship. This, in turn, contributes significantly to the advancement of precision medicine in psychotherapy, allowing for tailored interventions that consider individual nuances and optimize therapeutic outcomes.

iii. Significance of Indicators

Feature importance in machine learning algorithms adds significant value to precision medicine by aiding in the identification of crucial factors that influence the prediction or outcome. In the context of data-driven clinical decision-making, which aims to tailor medical decisions and treatments to individual characteristics, understanding feature importance is paramount. such information is invaluable in crafting personalized treatment strategies that consider individual patient characteristics.

iv. Limitations and Challenges

In this study, we acknowledge the influence of various processes such as safety, engagement, coregulation, and cooperation on the assessment of therapeutic alliance. These interpersonal experiences are intricately linked to physiological mechanisms, as evidenced by numerous studies. However, our study encountered limitations due to restricted access to certain physiological and psychological indicators. Having access to a broader range of indicators could offer a more comprehensive analysis, enabling us to identify the significance of these indicators in predicting therapeutic alliance and understanding their relationships in greater depth.

v. Theoretical reference

Drawing from Simon's model of optimal decision-making, which serves as the theoretical foundation for our proposed framework of IDSS4PM. As it is demonstrated in Figure 79, we employed descriptive analytics to enhance our understanding through observation and analysis. This contribution aligns with the Intelligence phase of Simon's framework for optimal decision-making. Furthermore, predictive analytics, utilizing machine learning algorithms to forecast targets and identify impactful predictors, bridges both the Intelligence and Design phases. Additionally, the descriptive and predictive analyses lay the groundwork for applying prescriptive analysis, aiming to optimize outcomes through simulations and what-if scenarios, encompassing both the Intelligence and Choice phases. This exploratory study paves the way for future research on optimization and recommendations.

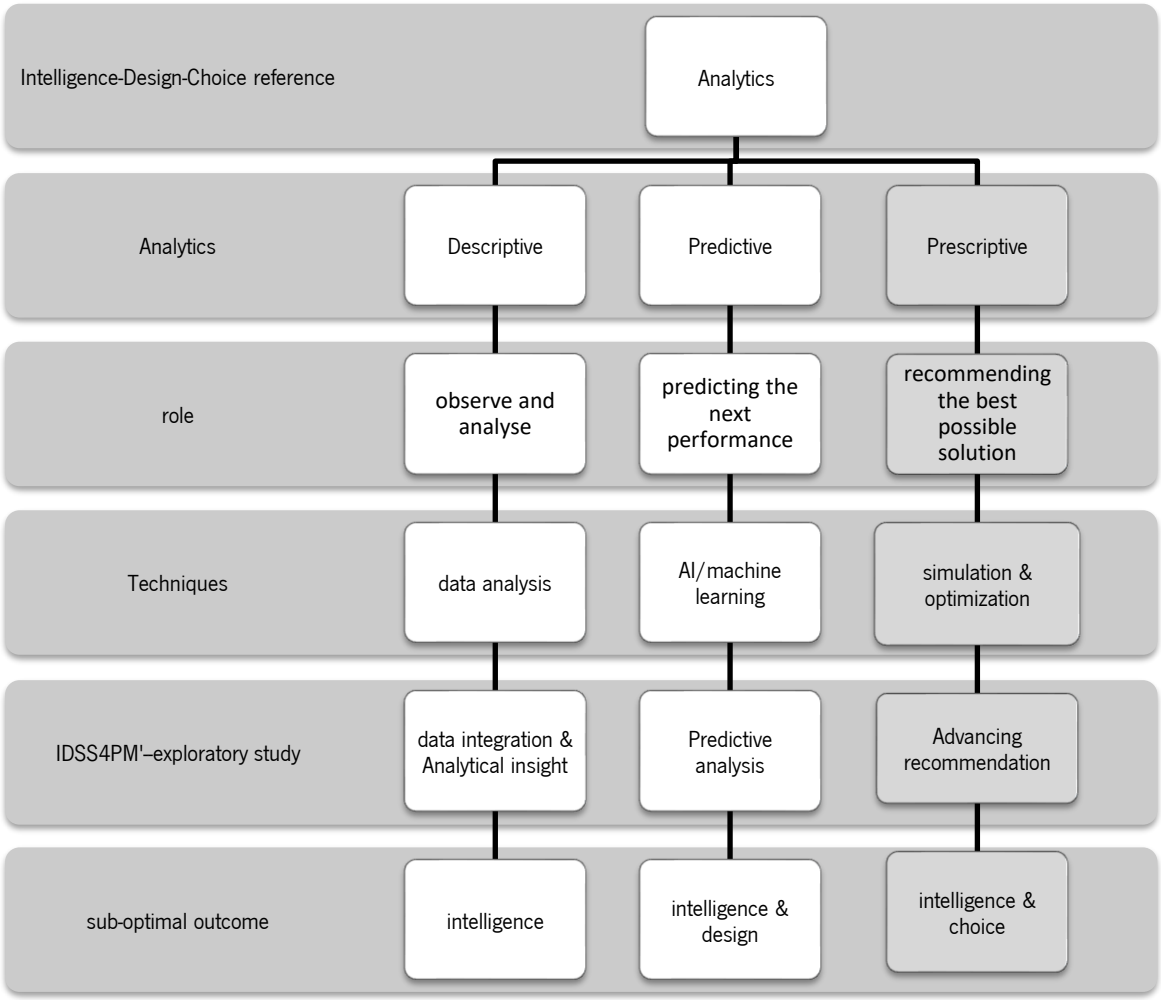


Figure 78: Theoretical reference; Project I

vi. At application level,

Analytical insight, as described in the context of examining relationships between therapists and clients in psychotherapy sessions, adds significant value to applications in precision medicine in several ways: The examination of relationships between therapists and clients provides insights into the dynamics of the therapeutic alliance. Analyzing these interactions allows for the identification of factors that positively or negatively impact the alliance. This knowledge is valuable in refining and optimizing the therapeutic relationship, which is crucial for successful precision medicine interventions. Moreover, identifying significant behavioral trends early on provides an opportunity for early intervention and preventive measures. This proactive approach aligns with the goals of precision medicine by addressing issues before they escalate, potentially improving treatment outcomes and minimizing adverse effects. Additionally, analytical insight supports continuous monitoring of patient data, allowing for dynamic adjustments to treatment plans. As patterns evolve or new trends emerge, healthcare providers can adapt interventions in real time, ensuring that the precision medicine approach remains aligned with the

patient's evolving needs.

By customizing treatment approaches based on observed patterns, analytical insight supports a patient-centered care model. Precision medicine aims to consider individual variations, and the personalized insights derived from data analysis contribute to a more patient-centric and tailored healthcare approach.

In terms of decision-making, the meticulous analysis of interrelated factors contributes to more informed decision-making in precision medicine. Healthcare professionals can make evidence-based choices, considering the nuanced relationships between physiological and psychological indicators, leading to more effective and targeted interventions.

Finally, predictive analysis and Feature importance help in assessing the risk associated with different factors. For example, in predicting the response to a particular treatment/medication, the algorithm can identify which patient-specific features are most indicative of positive or negative outcomes. This aids healthcare providers in making informed decisions about the potential risks and benefits for a specific patient. This allows for timely interventions or preventive measures, contributing to better health outcomes.

Furthermore, by understanding feature importance, precision medicine can optimize the allocation of diagnostic and therapeutic resources. Focus can be directed toward the most influential factors, improving efficiency and reducing unnecessary interventions.

Another key contribution Is about understanding complex relationships. In precision medicine, factors influencing health outcomes are often interconnected and complex. Machine learning algorithms can unravel these intricate relationships by assigning importance to each factor. This helps healthcare professionals comprehend the multifaceted nature of diseases and treatments.

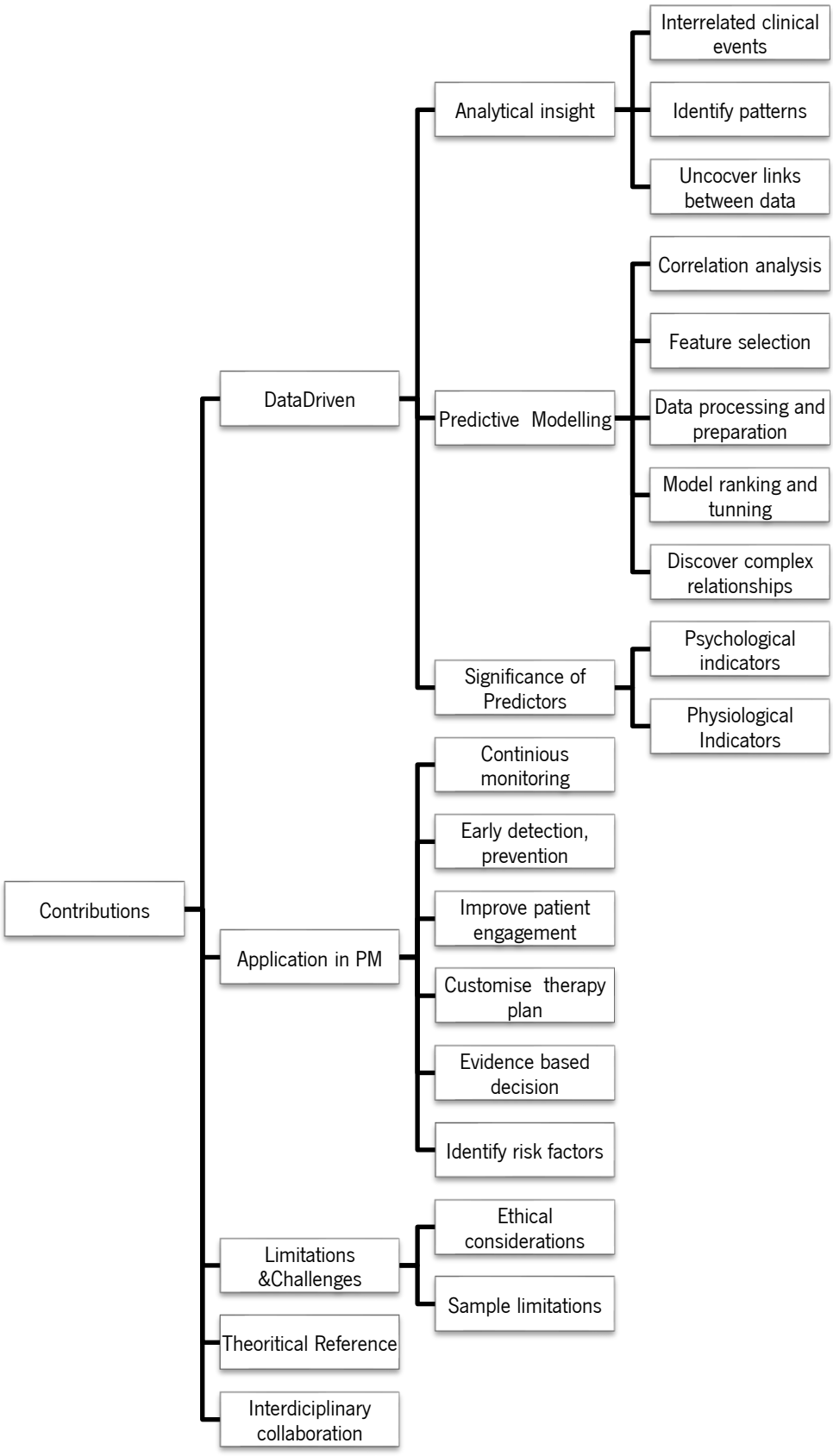


Figure 79: Contributions to IDSS4PM; Project I

6.2 Data-Driven Intensive Care Medicine Decision Support

i. CEid formulation and its transformative Impacts.

In the realm of data-driven aspects, the CEid formula stands out for its multifaceted and impactful advantages. These transformative endeavors resonate across datasets, introducing a variable that amalgamates key information. This identity plays a pivotal role in facilitating data processing and integration, essential for predictive modeling.

The sequential arrangement of events within a specific period, as captured by the event sequence, provides a comprehensive view of temporal order. This addresses challenges related to infrequent data generation and proved impactful for information exchange as a standard approach. Additionally, the CEid formula introduces a standardized method for creating keys for clinical events, ensuring uniformity and consistency in data representation. This standardization is crucial for interoperability and compatibility across healthcare systems, supporting the analysis and interpretation of temporal relationships.

Moreover, the crafted key serves as a standardized identifier for clinical events, promoting consistency and ease of information exchange. This common temporal reference facilitates smooth information exchange across different systems and platforms, providing a standardized framework for understanding and interpreting temporal relationships in clinical data.

The event sequence, part of the crafted key, furnishes a chronological order of clinical events, enabling temporal clustering. This sequential information allows for the identification of patterns or clusters of events occurring closely in time, enhancing temporal clustering by encapsulating diverse clinical events within a unified temporal framework. This aids in the recognition of meaningful clusters.

In terms of predictive modeling, the structured key, with detailed components such as the day of the event, data type, process number, episode number, and other clinical indicators, serves as a rich source for features in predictive modeling. This enables the development of predictive models that account for the sequential nature of clinical events, contributing to more accurate predictions based on temporal relationships.

The strength of the formula lies not only in its technical and data-driven aspects but also in its practical application in precision medicine. The CEid offers a dynamic and flexible means to analyze and interpret complex patient data, supporting informed decision-making and significantly enhancing the quality of patient care by providing a nuanced understanding of individual medical journeys.

This approach equips medical professionals with a versatile toolset, complete with filters facilitating a comprehensive review of diverse clinical events and their associated details, all within a unified temporal

framework. This innovative approach provides practitioners with a holistic lens through which to navigate and discern intricate patient trajectories, fostering informed decision-making and enhanced patient care.

The meticulous examination of clinical data plays a pivotal role in continuous patient monitoring. By keenly observing data fluctuations over time, healthcare providers can promptly identify deviations from a patient's baseline health and intervene early to mitigate potential complications. The unification of clinical events further empowers decision-makers, offering a comprehensive understanding of each patient's unique health journey. Armed with this information, they can tailor interventions to address the individual needs of the patient, thereby making treatment decisions more personalized. Moreover, this holistic approach to decision-making is instrumental in providing a panoramic view of a patient's health history. It not only facilitates proactive care but also contributes to an overall enhancement of patient well-being.

The capability to track and analyze clinical events longitudinally enables decision-makers to pinpoint trends and patterns that may signify the onset of health issues. Early detection through this method leads to proactive interventions, curbing the severity and cost of treatment. In terms of treatment effectiveness, clinical decision-makers leverage their ability to assess the impact of treatment strategies by analyzing patient responses over time. This valuable information guides the adjustment of treatment plans, aiming for improved outcomes and optimized resource allocation. Lastly, the focus extends beyond mere disease treatment to encompass preventive and proactive health management. The emphasis lies not only in addressing existing conditions but also in identifying risk factors and taking pre-emptive measures to prevent or delay the onset of diseases.

ii. Analytical insight

Clinical Event Identification (CEid) serves as a cornerstone in linking clinical events across a unified platform. By integrating each patient's clinical background, it enables thorough observation and analysis. This step is pivotal in identifying patterns within an integrated platform, providing decision-makers with comprehensive insights crucial for informed, evidence-based decisions. Ultimately, this approach enhances the likelihood of successful treatment outcomes and ensures patient satisfaction.

iii. Cluster analysis

As previously mentioned, the CEid plays a pivotal role in seamlessly integrating diverse data sources and types by extracting the day of the clinical event and the process number. Consequently, employing clustering techniques on the integrated data, independent of the process number, establishes an optimal platform to investigate the homogeneity of physiological data. This process involves creating cluster references to assign a combination of categorical and numerical data to each cluster. Moreover, the

standardization applied to generate custom codes for categorical variables enhances the overall performance of this procedure.

In addition, the phase of mapping and cluster assignment, with analyzing the minimum, maximum, and mean variables in each cluster provides valuable insights and benefits in several ways within the context of this work on health profiling and data-driven aspects. Here are the advantages:

a. Identification of Extreme Values:

Min and Max Values: Examining the minimum and maximum values within each cluster helps identify extreme values or outliers. These outliers may represent unique cases or anomalies that can provide valuable information about exceptional health conditions or atypical patient responses.

b. Understanding Range and Variability:

Mean and Variability: Calculating the mean and understanding the variability (min to max range) of variables within each cluster provides insights into the overall distribution of health parameters. This information is crucial for understanding the typical range of values associated with different health profiles.

c. Differential Health Characteristics:

Cluster-Specific Patterns: By comparing the minimum, maximum, and mean values across clusters, we can identify cluster-specific patterns. This helps in distinguishing the characteristic health parameters that define each cluster and contributes to the creation of more nuanced health profiles.

d. Clinical Relevance:

Identification of Clinical Significance: Extreme values or specific patterns in minimum, maximum, or mean values may have clinical significance. For instance, unusually high or low values could signal potential health risks or specific medical conditions that warrant closer attention.

e. Feature Importance in Predictive Modelling:

Variable Importance: Understanding the importance of variables based on their range and variability in different clusters can aid in feature selection for predictive modeling. Variables with substantial variations across clusters may play a crucial role in predicting health outcomes.

f. Quality Control and Data Integrity:

Detecting Data Anomalies: Discrepancies or unexpected patterns in minimum, maximum, or mean values may indicate data anomalies or errors. Regularly monitoring these statistics helps ensure the quality and integrity of the dataset.

At the end of each phase support clinical decision makers to tailoring Interventions: Insights derived from the analysis of variable ranges can contribute to the development of more precise medicine strategies. Tailoring interventions based on the specific health characteristics within each cluster allows for more targeted and effective healthcare practices.

iv. Classification:

Predictive health profiling through behavior-based classification, coupled with the utilization of temporal clustering, holds significant potential to propel the field of precision medicine forward. Here are key aspects that underscore its potential impact:

From a Data-Driven Decision-Making perspective, the incorporation of 18 predictors, encompassing a mix of categorical and numerical variables, reflects a comprehensive approach to capturing diverse facets of patient health. This rich dataset empowers healthcare professionals with the tools for more informed and data-driven decision-making, thereby elevating the precision and accuracy of medical interventions and treatment.

In Application-Level Insights: Leveraging behavior-based classification and temporal clustering enhances the model's capability to discern individual variations in patient health trajectories. This nuanced profiling contributes to a more personalized understanding of patient conditions, facilitating tailored medical interventions for better outcomes (Individualized Patient Profiling). Moreover, resulted in early detection and Intervention. The model's capacity to predict cluster numbers throughout the clinical day signifies an augmented ability to detect subtle changes in a patient's health status early on. This early detection lays the groundwork for timely interventions, potentially averting adverse health events and improving overall patient outcomes.

In addition, demonstrating real-time adaptability, the model's capacity to work with new data and provide daily predictions is crucial in a clinical setting where patient conditions may evolve rapidly. This feature enables continuous monitoring and the adaptation of treatment plans as needed.

In conclusion, the synergy between behavior-based classification, temporal clustering, and advanced data analytics holds transformative potential for precision medicine, promising more personalized, timely, and effective healthcare interventions.

v. Limitations and Challenges

We encountered significant limitations and challenges in advancing our analytical insights, including issues with data quality, diverse datasets, and infrequent data collection from sensors. Leveraging the CEid, we were able to address some of these challenges partially.

To create the artifact, we faced limitations in accessing specific types of data. For instance, datasets like SOAP contained predominantly unstructured text data, lacking the structured format required for predictive modeling. Additionally, ethical concerns regarding the accessibility of genetic profiles posed constraints on sample availability, further limiting our ability to gather diverse and comprehensive data for analysis.

vi. Theoretical reference - Suboptimal Decision-Making

Figure 80 illustrates the pivotal role of analytics and applied techniques on clinical data to achieve suboptimal outcomes. As discussed earlier, the complexity of individual patient data arises from diverse sources and formats, leading to dynamic, semi/unstructured, multi-dimensional, and fragmented data (Kaur & Mann, 2018). In the context of healthcare, suboptimal decision-making refers to instances where the chosen course of action does not yield the best possible outcome for the patient. This can occur due to various factors, including incomplete information, cognitive biases, or limitations in analytical tools. Therefore, advanced analytics, particularly predictive and prescriptive analytics, become essential for optimizing treatment performance (Palaniappan & Awang, 2008). By leveraging past patient data and sophisticated machine learning algorithms, predictive analytics can anticipate treatment features and potential outcomes. However, despite the predictive capabilities of these techniques, decision-makers may still encounter challenges in identifying the most appropriate course of action.

Prescriptive analytics, on the other hand, offers a decision-focused approach to healthcare decision-making. By simulating various treatment scenarios and optimizing for the desired outcome, prescriptive analytics can guide clinicians toward more effective interventions. Nonetheless, the application of prescriptive analytics in real-world settings may be constrained by factors such as resource limitations or organizational constraints.

Overall, the integration of advanced analytics into clinical decision-making processes holds significant promise for improving patient outcomes and enhancing healthcare delivery. However, it is essential to recognize the inherent challenges and limitations associated with suboptimal decision-making in healthcare settings and to continue refining analytical approaches to address these complexities.

To advance the sub-optimal model of clinical decision-making and develop the IDSS4PM artifact, we introduced the CEid and analytical insights to grasp the underlying problem. This involved employing descriptive analytics to glean valuable knowledge, thus representing the intelligent phase. Furthermore,

predictive modeling contributed to both the intelligence and design phases. In Simon's model, decision-makers design alternative courses of action based on predictions—a process mirrored in predictive analytics, which assists in devising treatment plans by suggesting strategies informed by projected outcomes. This iterative approach aligns with the iterative process of generating and evaluating multiple treatment scenarios based on predictive models.

Finally, Choice Phase: Prescriptive analytics corresponds to the choice phase of Simon's model, where decision-makers select the most optimal course of action among the alternatives generated in the design phase. Prescriptive analytics guides decision-makers by recommending specific treatment options or interventions that are expected to yield the best outcomes based on simulation and optimization techniques.

Furthermore, prescriptive analytics supports intelligent adaptation by continuously refining treatment recommendations based on new data or changing patient conditions. This reflects Simon's notion of intelligent adaptation, where decision-makers adjust their strategies in response to feedback and environmental changes to achieve better outcomes over time.

In this context, building upon the preceding discussion, our contribution was directed towards enhancing intelligence, refining the design process, and advancing the decision-making choices.

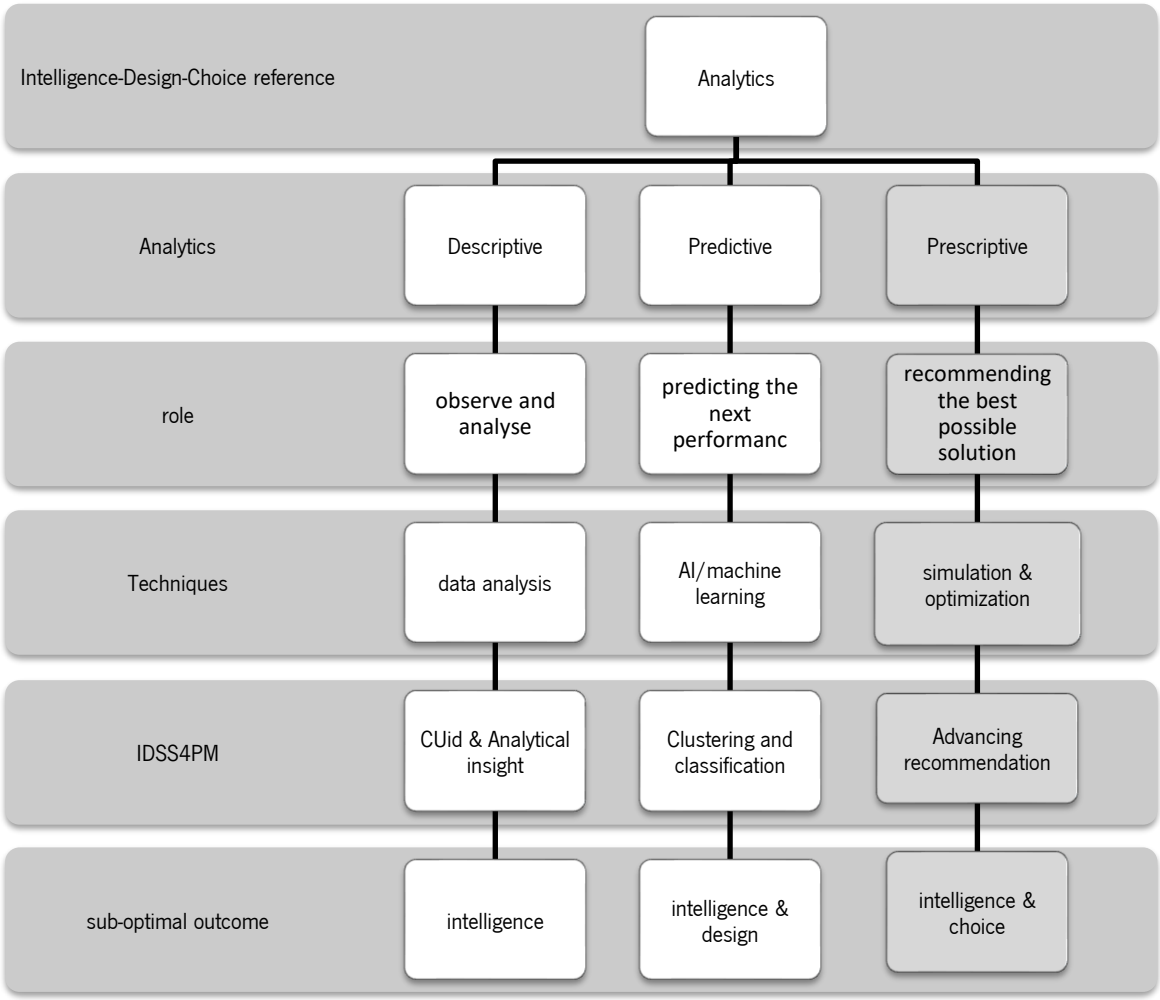


Figure 80: Theoretical reference; Project II

I. Application in PM:

In addition to our contribution to data-driven aspects, including the development of an intelligent decision support system for precision medicine with implications on the application level, the outcomes indirectly lead to:

Patients witnessing their healthcare providers making data-informed decisions based on a unified view of their clinical events are more likely to actively engage in their care. This increased engagement often results in better adherence to treatment plans and subsequently leads to improved health outcomes (Improved Patient Engagement).

Moreover, Decision-makers gain the ability to allocate healthcare resources more efficiently by prioritizing patients who require immediate attention, this strategic allocation enhances resource utilization, ultimately leading to improved patient outcomes. (Optimized Resource Allocation) and Understanding the progression of a patient's health condition and the impact of various clinical events empowers decision-

makers to avoid unnecessary and costly diagnostic tests and treatments that may not be effective. This not only leads to cost savings but also promotes a more streamlined and effective healthcare approach. In summary, scientific endeavors not only contribute to advancing data-driven aspects but also bring about positive changes in patient engagement, resource allocation efficiency, and cost-effectiveness within the realm of precision medicine.

This project was a collaborative interdisciplinary effort with the ICU of the hospital, enriching our ability to showcase the value of our contribution in creating the artifact.

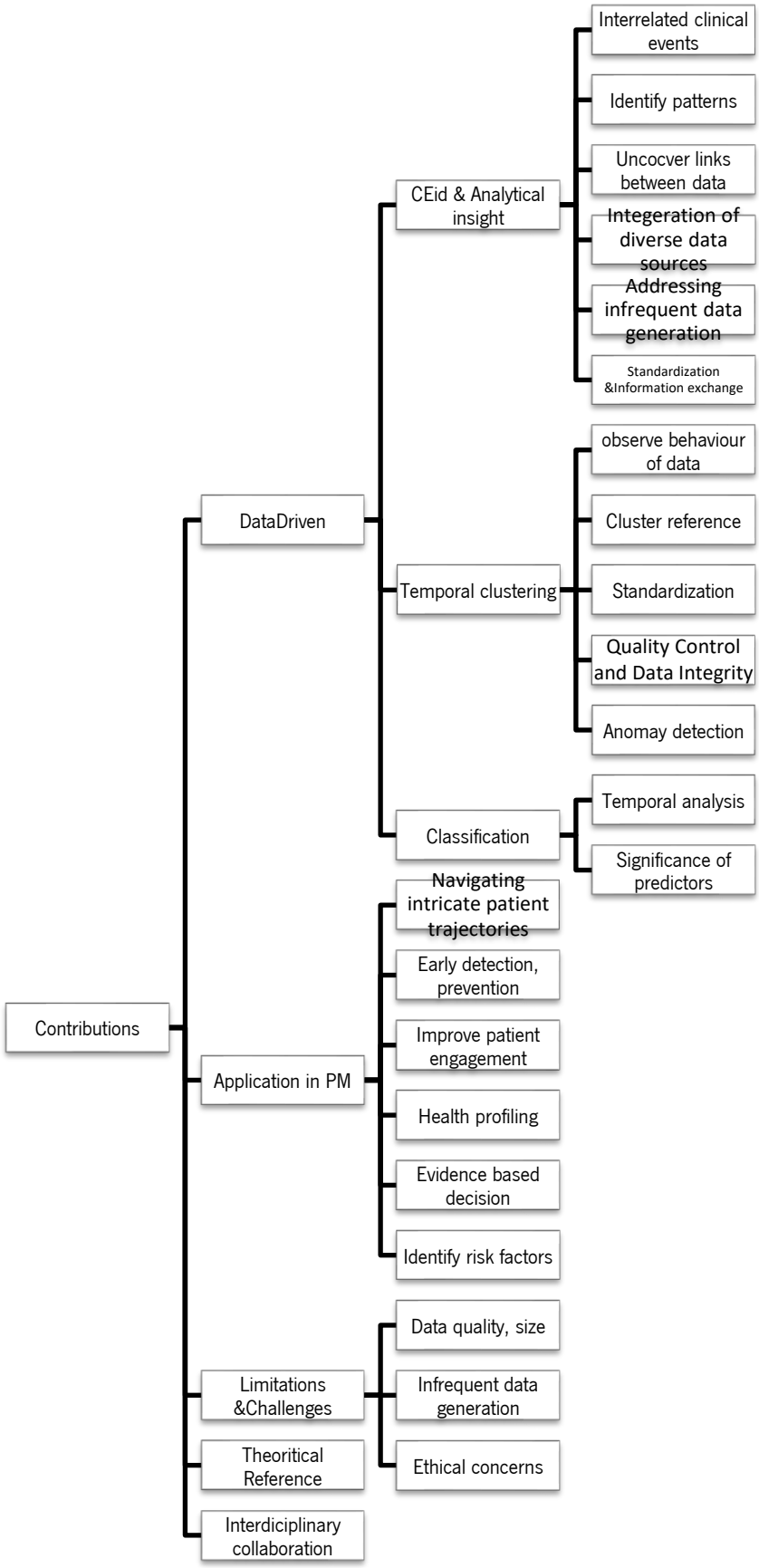


Figure 81: :Contribution to IDSS4PM; Project II

7 Conclusion, Limitations, and Future Work

This study aimed to critically analyze existing frameworks for Intelligent Decision Support Systems for Precision Medicine (IDSS4PM) by exploring literature on intelligent decision support systems and precision medicine. The evolving concept of precision medicine and key limitations and challenges in data aspects were extensively discussed, with a particular focus on data-driven aspects.

To address the research question of "How can data-driven insights shape the framework of Intelligent Decision Support Systems and advance optimal clinical decision-making?" we conducted two interdisciplinary research projects. Project I utilized psychotherapy data collected from the Department of Psychology at the University of Minho. Our exploratory study provided valuable insights into the development of IDSS4PM, aiming to uncover hidden patterns within datasets, particularly focusing on understanding therapeutic alliance in psychotherapy. Our exploration revealed significant findings regarding the complex factors influencing therapeutic alliance and the potential of data-driven methodologies to enhance treatment outcomes and support clinical decision-making in psychotherapy. By identifying effective machine learning algorithms and integrating diverse datasets, we paved the way for more precise decisions tailored to individual patients' unique circumstances. Leveraging integrated clinical events and predictive modeling techniques, we aim to further enhance treatment efficacy and patient outcomes.

However, it's essential to acknowledge the limitations of our study. Psychotherapy is a complex interpersonal process, and disentangling relational and concurrent factors remains challenging. Additionally, causal relationships between variables should be interpreted cautiously, considering bidirectional associations and interactions with therapeutic processes. Furthermore, our study's scope was limited to clients diagnosed with major depression or social anxiety, primarily treated using cognitive behavior therapy. Generalizations should be cautious, and future research could explore a broader range of therapeutic approaches and patient populations. Improvements in physiological measures and experimental designs could enhance the validity and generalizability of our findings.

In summary, this study contributes to precision medicine by identifying key factors through machine learning algorithms, enabling personalized treatment approaches, assessing risks, facilitating early detection, optimizing resource allocation, and fostering continuous learning. By addressing limitations and exploring future avenues, we aim to advance the effectiveness of medical interventions and improve patient outcomes in a tailored and individualized manner.

The novelty of this study is illuminated through Project II, where we utilized various datasets from seventy

patients in the ICU, comprising a mixture of clinical and physiological data with diverse and infrequent data generation patterns. In the data processing phase, we proposed a novel method to integrate datasets, addressing aspects such as infrequent data generation, standardization improvement, interoperability, and information exchange.

One impactful artifact resulting from this approach is the Clinical Event Identity, which encapsulates each clinical transaction with meaningful information. This identity serves as a pointer to specific clinical events and facilitates information exchange, thereby enhancing analytical insights. Based on this, we presented the results of knowledge discovery through the analytical insight phase, providing crucial support for decision-makers to observe and analyze the clinical background, health status, and risky conditions of each patient across multidimensional aspects.

Additionally, the temporal clustering phase not only provided remarkable observations to analyze data behavior in identified clusters but also proposed cluster references where we mapped those references on the original dataset, including missing cells. This mapping enabled us to identify patterns or health profiles. The results of this phase were discussed and analyzed from data-driven aspects and their application in precision medicine.

To advance predictive analytics on new data, we performed classification to demonstrate the possibility of predicting temporal cluster numbers. Project II faced challenges and limitations in terms of data quality, lack of standardization, and infrequent data generation, which resulted in major challenges in data integration for modeling.

Future works potentially involve applying prescriptive analytics via simulation and what-if scenarios to advance recommendations and propose the best possible results (treatment/actions) for each cluster, considering temporal aspects. This would be a novel approach to support clinical decision-making in the context of precision medicine.

Each project was widely discussed in terms of the data-driven contributions to the IDSS4PM, and the limitations we faced, and we also justified the work with theoretical references, in this case, the optimal decision-making model proposed by Herbert Simon.

8 List of Publication

1. Mosavi, N. S., & Santos, M. F. (2024). Enhancing Clinical Decision Support for Precision Medicine: A Data-Driven Approach. (Submited to Q1 SCimago / JCR Jornal)

2. Mosavi, N. S., Ribeiro, E., Sampaio, A., et al. (2023). Data mining techniques in psychotherapy: Applications for studying therapeutic alliance. Scientific Reports, 13, 16409. https://doi.org/10.1038/s41598-023-43366-6

3. Mosavi, N. S., & Santos, M. F. (2022). Data Engineering to Support Intelligence for Precision Medicine in Intensive Care. In 2022 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 1782-1786). Las Vegas, NV, USA. https://doi.org/ 10.1109/CSCI58124.2022.00316.

4. Mosavi, N. S., & Santos, M. F. (2023). Unveiling Precision Medicine with Data Mining: Discovering Patient Subgroups and Patterns. In 2023 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1304-1309). Mexico City, Mexico. https://doi.org/10.1109/SSCI52147.2023.10372022.

5. Mosavi, N. S., & Santos, M. F. (2022). Intelligent Decision Support System for Precision Medicine: Time Series Multi-variable Approach for Data Processing. In the International Conference on Knowledge Discovery and Information Retrieval.

6. Mosavi, N. S., & Santos, M. F. (2020). How Prescriptive Analytics Influences Decision Making in Precision Medicine. Procedia Computer Science, 177, 528-533. https://doi.org/10.1016/j.procs.2020.10.073.

7. Mosavi, N. S., & Santos, M. F. (2022). Internet of things for precision intensive medicine. Procedia Computer Science, 201, 732-737. https://doi.org/10.1016/j.procs.2022.03.099.

8. Mosavi, N. S., & Santos, M. F. (2022). To what extent healthcare analytics influences decision-making in precision medicine. Procedia Computer Science, 198, 353-359. https://doi.org/10.1016/j.procs.2021.12.253.

9. Mosavi, N. S., & Santos, M. F. (2021). Adoption of Precision Medicine; Limitations and Considerations. ArXiv, abs/2105.15115.

10. Mosavi, N. S., & Santos, M. F. (2022). Characteristics of the Intelligent Decision Support System for Precision Medicine (IDSS4PM). In X. S. Yang, S. Sherratt, N. Dey, & A. Joshi (Eds.), Proceedings of Sixth International Congress on Information and Communication Technology (pp. 703-714). Lecture Notes in Networks and Systems, Vol. 217. Springer, Singapore. https://doi.org/10.1007/978-981-16-2102-4_61.

11. Mosavi, N. S., & Santos, M. F. (2022). Intelligent Decision Support System for Precision Medicine (IDSS4PM). In T. Wang, S. Patnaik, A. W. Ip, & M. Tavana (Eds.), Advances in Decision Science and Management. ICDSM 2021. (Advances in Intelligent Systems and Computing, Vol. 1391). Springer, Singapore. https://doi.org/10.1007/978-981-16-2502-2_4.

12. Mosavi, N. S., & Santos, M. F. (2022). Implementation Considerations for the Applied Business Intelligence in Healthcare. In T. Wang, S. Patnaik, A. W. Ip, & M. Tavana (Eds.), Advances in Decision Science and Management. ICDSM 2021. (Advances in Intelligent Systems and Computing, Vol. 1391). Springer, Singapore. https://doi.org/10.1007/978-981-16-2502-2_19.

13. Mosavi, N. S., Freitas, F., Pires, R., Rodrigues, C., Silva, I., Santos, M., & Novais, P. (2022). Intelligent energy management using data mining techniques at Bosch Car Multimedia Portugal facilities. Procedia Computer Science, 201, 503-510. https://doi.org/10.1016/j.procs.2022.03.065.

BIBLIOGRAPHY

Abubakar, A. M., Elrehail, H., Alatailat, M. A., & Elçi, A. (2019). Knowledge management, decision-making style and organizational performance. *Journal of Innovation and Knowledge*, *4*(2), 104–114. https://doi.org/10.1016/j.jik.2017.07.003

Adiba, F. I., Sharwardy, S. N., & Rahman, M. Z. (2021). Multivariate time series prediction of pediatric ICU data using deep learning. *2021 International Conference on Innovative Trends in Information Technology, ICITIIT 2021*. https://doi.org/10.1109/ICITIIT51526.2021.9399593

Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, *2020*, 1–35. https://doi.org/10.1093/database/baaa010

Alharthi, H. (2018). Healthcare predictive analytics: An overview with a focus on Saudi Arabia. *Journal of Infection and Public Health*, *11*(6), 749–756. https://doi.org/10.1016/j.jiph.2018.02.005

Al-Nbhany, W. A. N. A., Zahary, A. T., & Al-Shargabi, A. A. (2024). Blockchain-IoT Healthcare Applications and Trends: A Review. *IEEE Access*, *12*. https://doi.org/10.1109/ACCESS.2023.3349187

Alqenae, F. A., Steinke, D., & Keers, R. N. (2020). Prevalence and Nature of Medication Errors and Medication-Related Harm Following Discharge from Hospital to Community Settings: A Systematic Review. In *Drug Safety* (Vol. 43, Issue 6). https://doi.org/10.1007/s40264-020-00918-3

Alzheimer's and Dementia. (2017). 2017 Alzheimer's disease facts and figures. *Alzheimer's and Dementia*, *13*(4). https://doi.org/10.1016/j.jalz.2017.02.001

Aristodemou, L., & Tietze, F. (2018). The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. *World Patent Information*, *55*(July), 37–51. https://doi.org/10.1016/j.wpi.2018.07.002

Arnott, D., & Pervan, G. (2005). A critical analysis of decision support systems research. *Journal of Information Technology*, *20*(2), 67–87. https://doi.org/10.1057/palgrave.jit.2000035

Arnott, D., & Pervan, G. (2008). Eight key issues for the decision support systems discipline. *Decision Support Systems*, *44*(3), 657–672. https://doi.org/10.1016/j.dss.2007.09.003

Asemi, A., Safari, A., & Asemi Zavareh, A. (2011). The Role of Management Information System (MIS) and Decision Support System (DSS) for Manager's Decision Making Process. *International Journal of Business and Management*, *6*(7). https://doi.org/10.5539/ijbm.v6n7p164

Awwalu, J., Garba, A. G., Ghazvini, A., & Atuah, R. (2015). Artificial Intelligence in Personalized Medicine Application of AI Algorithms in Solving Personalized Medicine Problems. *International Journal of Computer Theory and Engineering*, *7*(6), 439–443. https://doi.org/10.7763/ijcte.2015.v7.999

Bae, C. J., Griffith, S., Fan, Y., Dunphy, C., Thompson, N., Urchek, J., Parchman, A., & Katzan, I. L. (2015). The Challenges of Data Quality Evaluation in a Joint Data Warehouse. *EGEMs (Generating Evidence & Methods to Improve Patient Outcomes)*, *3*(1). https://doi.org/10.13063/2327-9214.1125

Banappagoudar, S. B., Santhosh, S. U., Thangam, M. M. N., David, S., & Shamili, G. S. (2023). The Advantages of Precision Medicine over Traditional Medicine in Improving Public Health Outcomes. *Journal for ReAttach Therapy and Developmental Diversities*, *6*(5).

Baxter, R., Hastings, N., Law, A., & Glass, E. J. . (2008). Cognitive Computing: Theory and Applications. In *Animal Genetics* (Vol. 39, Issue 5).

Beckmann, J. S., & Lew, D. (2016). Reconciling evidence-based medicine and precision medicine in the era of big data: Challenges and opportunities. *Genome Medicine*, *8*(1). https://doi.org/10.1186/s13073-016-0388-7

Behera, R. K., Bala, P. K., & Dhir, A. (2019). The emerging role of cognitive computing in healthcare: A systematic literature review. *International Journal of Medical Informatics*, *129*(February), 154–166. https://doi.org/10.1016/j.ijmedinf.2019.04.024

Bekbolatova, M., Mayer, J., Ong, C. W., & Toma, M. (2024). Transformative Potential of AI in Healthcare: Definitions, Applications, and Navigating the Ethical Landscape and Public Perspectives. In *Healthcare (Switzerland)* (Vol. 12, Issue 2). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/healthcare12020125

Bini, S. A. (2018). Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *Journal of Arthroplasty*, *33*(8), 2358–2361. https://doi.org/10.1016/j.arth.2018.02.067

Bonkhoff, A. K., & Grefkes, C. (2022). Precision medicine in stroke: Towards personalized outcome predictions using artificial intelligence. In *Brain* (Vol. 145, Issue 2). https://doi.org/10.1093/brain/awab439

Borovska, P., Ivanova, D., & Draganov, I. (2018). Internet of Medical Imaging Things and Analytics in Support of Precision Medicine for Early Diagnostics of Thyroid Cancer. *Serdica Journal of Computing*, *12*(1–2). https://doi.org/10.55630/sjc.2018.12.47-64

Brown, S. A. (2016). Patient similarity: Emerging concepts in systems and precision medicine. *Frontiers in Physiology*, *7*(NOV), 1–6. https://doi.org/10.3389/fphys.2016.00561

Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine Learning for Precision Psychiatry: Opportunities and Challenges. In *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* (Vol. 3, Issue 3). https://doi.org/10.1016/j.bpsc.2017.11.007

Carra, G., Salluh, J. I. F., da Silva Ramos, F. J., & Meyfroidt, G. (2020). Data-driven ICU management: Using Big Data and algorithms to improve outcomes. *Journal of Critical Care*, *60*, 300–304. https://doi.org/10.1016/j.jcrc.2020.09.002

Cazzola, M., Calzetta, L., Rogliani, P., & Matera, M. G. (2017). The Challenges of Precision Medicine in COPD. *Molecular Diagnosis and Therapy*, *21*(4). https://doi.org/10.1007/s40291-017-0266-z

Chang, D., Chang, D., & Pourhomayoun, M. (2019). Risk prediction of critical vital signs for ICU patients using recurrent neural network. *Proceedings - 6th Annual Conference on Computational Science and Computational Intelligence, CSCI 2019*, 1003–1006. https://doi.org/10.1109/CSCI49370.2019.00191

Chase, J. G., Preiser, J. C., Dickson, J. L., Pironet, A., Chiew, Y. S., Pretty, C. G., Shaw, G. M., Benyo, B., Moeller, K., Safaei, S., Tawhai, M., Hunter, P., & Desaive, T. (2018). Next-generation, personalised, model-based critical care medicine: A state-of-the art review of in silico virtual patient models, methods, and cohorts, and how to validation them. *BioMedical Engineering Online*, *17*(1), 1–29. https://doi.org/10.1186/s12938-018-0455-y

Chen, R., & Snyder, M. (2013). Promise of personalized omics to precision medicine. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *5*(1), 73–82. https://doi.org/10.1002/wsbm.1198

Cirillo, D., & Valencia, A. (2019). Big data analytics for personalized medicine. *Current Opinion in Biotechnology*, *58*, 161–167. https://doi.org/10.1016/j.copbio.2019.03.004

Del Re, A. C., Flückiger, C., Horvath, A. O., & Wampold, B. E. (2021). Examining therapist effects in the alliance–outcome relationship: A multilevel meta-analysis. *Journal of Consulting and Clinical Psychology*, *89*(5). https://doi.org/10.1037/ccp0000637

Delen, D. (2020). Prescriptive Analytics The Final Frontier for Evidence-Based Management and Optimal Decision. In *Library*. Pearson Education, Inc.

Diouf, P. S., Boly, A., & Ndiaye, S. (2018). Variety of data in the ETL processes in the cloud: State of the art. *2018 IEEE International Conference on Innovative Research and Development, ICIRD 2018*. https://doi.org/10.1109/ICIRD.2018.8376308

Duffy, D. J. (2016). Problems, challenges and promises: perspectives on precision medicine. *Briefings in Bioinformatics*, *17*(3), 494–504. https://doi.org/10.1093/bib/bbv060

Dulcic, Z., Pavlic, D., & Silic, I. (2012). Evaluating the Intended Use of Decision Support System (DSS) by Applying Technology Acceptance Model (TAM) in Business Organizations in Croatia. *Procedia - Social and Behavioral Sciences*, *58*, 1565–1575. https://doi.org/10.1016/j.sbspro.2012.09.1143

Felsberger, A., Oberegger, B., & Reiner, G. (2017). A review of decision support systems for manufacturing systems. *CEUR Workshop Proceedings*, *1793*.

Fenning, S. J., Smith, G., & Calderwood, C. (2019). Realistic Medicine: Changing culture and practice in the delivery of health and social care. *Patient Education and Counseling*, *102*(10), 1751–1755. https://doi.org/10.1016/j.pec.2019.06.024

Fentahun Moges Kasie, Glen Bright, A. W. (2017). Decision support systems in manufacturing: a survey and future trends. *Journal of Modelling in Management,*. https://doi.org/10.1108/JM2-02-2016-0015

Flückiger, C., Del Re, A. C., Wlodasch, D., Horvath, A. O., Solomonov, N., & Wampold, B. E. (2020). Assessing the alliance–outcome association adjusted for patient characteristics and treatment processes: A meta-analytic summary of direct comparisons. *Journal of Counseling Psychology*, *67*(6). https://doi.org/10.1037/cou0000424

Francis S. Collins, H. V. (2015). A commentary on "A new initiative on precision medicine." *Frontiers in Psychiatry*, *6*(MAY), 88. https://doi.org/10.3389/fpsyt.2015.00088

Ginsburg, G. S., & Phillips, K. A. (2018). Precision medicine: From science to value. *Health Affairs*, *37*(5), 694–701. https://doi.org/10.1377/hlthaff.2017.1624

Glen H. Murata, Albuquerque, N. (2014). *(12) Patent Application Publication (10) Pub. No.: US 2014/0236630 A1*. *1*(19).

Gligorijević, V., Malod-Dognin, N., & Pržulj, N. (2016). Integrative methods for analyzing big data in precision medicine. *PROTEOMICS*, *16*(5), 741–758. https://doi.org/10.1002/PMIC.201500396

Gopal, G., Suter-Crazzolara, C., Toldo, L., & Eberhardt, W. (2019). Digital transformation in healthcare - Architectures of present and future information technologies. *Clinical Chemistry and Laboratory Medicine*, *57*(3). https://doi.org/10.1515/cclm-2018-0658

Grady, N. W. (2016). Knowledge Discovery in Data KDD Meets Big Data. *Archives of Civil Engineering*, *62*(2), 217–228. https://doi.org/10.1515/ace-2015-0076

Gupta, N. S., & Kumar, P. (2023). Perspective of artificial intelligence in healthcare data management: A journey towards precision medicine. In *Computers in Biology and Medicine* (Vol. 162). https://doi.org/10.1016/j.compbiomed.2023.107051

Gupta, S., Kar, A. K., Baabdullah, A., & Al-Khowaiter, W. A. A. (2018). Big data with cognitive computing: A review for the future. *International Journal of Information Management*, *42*(April), 78–89. https://doi.org/10.1016/j.ijinfomgt.2018.06.005

Haque, M., Islam, T., Sartelli, M., Abdullah, A., & Dhingra, S. (2020a). Prospects and challenges of precision medicine in lower-and middle-income countries: A brief overview. *Bangladesh Journal of Medical Science*, *19*(1), 32–47. https://doi.org/10.3329/bjms.v19i1.43871

Haque, M., Islam, T., Sartelli, M., Abdullah, A., & Dhingra, S. (2020b). Prospects and challenges of precision medicine in lower-and middle-income countries: A brief overview. *Bangladesh Journal of Medical Science*, *19*(1), 32–47. https://doi.org/10.3329/bjms.v19i1.43871

Hasan, M. K., Alkhalifah, A., Islam, S., Babiker, N. B. M., Habib, A. K. M. A., Aman, A. H. M., & Hossain, M. A. (2022). Blockchain Technology on Smart Grid, Energy Trading, and Big Data: Security Issues, Challenges, and Recommendations. In *Wireless Communications and Mobile Computing* (Vol. 2022). https://doi.org/10.1155/2022/9065768

Hasenfuß, G., & Vogelmeier, C. F. (2019). Digital medicine. In A. André (Ed.), *Springer* (Vol. 60, Issue 4). https://doi.org/10.1007/s00108-019-0594-7

Hee Lee, D., & Yoon, S. N. (2021). Application of artificial intelligence-based technologies in the healthcare industry: Opportunities and challenges. *International Journal of Environmental Research and Public Health*, *18*(1). https://doi.org/10.3390/ijerph18010271

Hevner, A. R., March, S. T., Park, J., Ram, S., & Ram, S. (2004). Research Essay Design Science in Information. *MIS Quarterly*, *28*(1), 75–105. https://doi.org/10.2307/25148625

Horvitz, E. J., Breese, J. S., & Henrion, M. (1988). Decision theory in expert systems and {AI}. *Intl. J. Approximate Reasoning*, *2*(3), 247–302.

Hulsen, T., Jamuar, S. S., Moody, A. R., Karnes, J. H., Varga, O., Hedensted, S., Spreafico, R., Hafler, D. A., & McKinney, E. F. (2019). From big data to precision medicine. *Frontiers in Medicine*, *6*(MAR), 1–14. https://doi.org/10.3389/fmed.2019.00034

Jabbar, M. A., Samreen, S., & Aluvalu, R. (2018). The future of health care: Machine learning. *International Journal of Engineering and Technology(UAE)*, *7*(4). https://doi.org/10.14419/ijet.v7i4.6.20226

Jameson, J. L., & Longo, D. L. (2015). Precision medicine - Personalized, problematic, and promising. *New England Journal of Medicine*, *372*(23), 2229–2234. https://doi.org/10.1056/NEJMsb1503104

Jean-Charles Pomerol. (1997). Artificial Intelligence and human decision making. *European Journal of Operational Research*. https://doi.org/10.1145/2987491.2987493

Johnson, K. B., Wei, W. Q., Weeraratne, D., Frisse, M. E., Misulis, K., Rhee, K., Zhao, J., & Snowdon, J. L. (2021). Precision Medicine, AI, and the Future of Personalized Health Care. In *Clinical and Translational Science* (Vol. 14, Issue 1). https://doi.org/10.1111/cts.12884

Johnson, N., Parbhoo, S., Ross, A. S., & Doshi-Velez, F. (2021). *Learning Predictive and Interpretable Timeseries Summaries from ICU Data*. https://arxiv.org/abs/2109.11043v1

Kalantari, B. (2010). Herbert A. Simon on making decisions: enduring insights and bounded rationality. *Journal of Management History*, *16*, 509–520. https://doi.org/10.1108/17511341011073988%0ADownloaded

Karafiloski, E., & Mishev, A. (2017). Blockchain solutions for big data challenges: A literature review. *17th IEEE International Conference on Smart Technologies, EUROCON 2017 - Conference Proceedings*. https://doi.org/10.1109/EUROCON.2017.8011213

Khadanga, S., Aggarwal, K., Joty, S., & Srivastava, J. (2019). Using Clinical Notes with Time Series Data for ICU Management. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 6432–6437. https://doi.org/10.18653/v1/d19-1678

Kherdekar, V. A., & Metkewar, P. S. (2016). A technical comprehensive survey of ETL tools. *International Journal of Applied Engineering Research*, *11*(4). https://doi.org/10.37622/ijaer/11.4.2016.2557-2559

Khnaisser, C., Lavoie, L., Diab, H., & Ethier, J. F. (2015). Data warehouse design methods review: Trends, challenges and future directions for the healthcare domain. *Communications in Computer and Information Science*, *539*. https://doi.org/10.1007/978-3-319-23201-0_10

Kirs, P. J., Sanders, G. L., Cerveny, R. P., & Robey, D. (2006). An Experimental Validation of the Gorry and Scott Morton Framework. *MIS Quarterly*, *13*(2), 183. https://doi.org/10.2307/248926

Kizer, K. W. (2001). Establishing health care performance standards in an era of consumerism. *JAMA*, *286*(10). https://doi.org/10.1001/jama.286.10.1213

König, I. R., Fuchs, O., Hansen, G., von Mutius, E., & Kopp, M. V. (2017). What is precision medicine? *The European Respiratory Journal*, *50*(4), 1–12. https://doi.org/10.1183/13993003.00391-2017

Kosorok, M. R., & Laber, E. B. (2019a). Annual Review of Statistics and Its Application Precision Medicine. *Annual Review of Statistics and Its Application*, *6*.

Kosorok, M. R., & Laber, E. B. (2019b). Precision Medicine. *Annual Review of Statistics and Its Application*, *6*, 263–286. https://doi.org/10.1146/annurev-statistics-030718-105251

Kraus, J. M., Lausser, L., Kuhn, P., Jobst, F., Bock, M., Halanke, C., Hummel, M., Heuschmann, P., & Kestler, H. A. (2018). Big data and precision medicine: challenges and strategies with healthcare data. *International Journal of Data Science and Analytics*, *6*(3). https://doi.org/10.1007/s41060-018-0095-0

Kuceba, R., & Kieltyka, L. (2009). Industrial Paper Criteria Classification Intelligent Decision Support Systems SYSTEMS CLASSIFICATION WITH REFERENCE TO DECISION MAKING. *Quality*, 5351–5355.

Lasorsa, I., D'Antrassi, P., Ajčević, M., Stellato, K., Di Lenarda, A., Marceglia, S., & Accardo, A. (2016). Personalized support for chronic conditions. *Applied Clinical Informatics*, *07*(03). https://doi.org/10.4338/aci-2016-01-ra-0011

Latimer, T., Roscamp, J., & Papanikitas, A. (2017). Patient-centredness and consumerism in healthcare: an ideological mess. In *Journal of the Royal Society of Medicine* (Vol. 110, Issue 11). https://doi.org/10.1177/0141076817731905

Leff, D. R., & Yang, G. Z. (2015). Big Data for Precision Medicine. *Engineering*, *1*(3), 277–279. https://doi.org/10.15302/J-ENG-2015075

Leone, D., Schiavone, F., Appio, F. P., & Chiao, B. (2021). How does artificial intelligence enable and enhance value co-creation in industrial markets? An exploratory case study in the healthcare ecosystem. *Journal of Business Research*, *129*. https://doi.org/10.1016/j.jbusres.2020.11.008

Lintz, J. (2020). Identification of Master Patient Index Record Challenges from Healthcare Professionals' Perspectives. *European Journal of Public Health*, *30*(Supplement_5). https://doi.org/10.1093/eurpub/ckaa166.032

Liu, S., Duffy, A. H. B., Whitfield, R. I., & Boyle, I. M. (2010a). Integration of decision support systems to improve decision support performance. *Knowledge and Information Systems*, *22*(3), 261–286. https://doi.org/10.1007/s10115-009-0192-4

Liu, S., Duffy, A. H. B., Whitfield, R. I., & Boyle, I. M. (2010b). Integration of decision support systems to improve decision support performance. *Knowledge and Information Systems*, *22*(3), 261–286. https://doi.org/10.1007/s10115-009-0192-4

Liu, X., Luo, X., Jiang, C., & Zhao, H. (2019). Difficulties and challenges in the development of precision medicine. *Clinical Genetics*, *95*(5), 569–574. https://doi.org/10.1111/CGE.13511/

Lytras, M., Visvizi, A., Zhang, X., & Aljohani, N. R. (2020). Cognitive computing, Big Data Analytics and data driven industrial marketing. *Industrial Marketing Management*, *90*, 663–666. https://doi.org/10.1016/j.indmarman.2020.03.024

Maceachern, S. J., & Forkert, N. D. (2021). Machine learning for precision medicine. In *Genome* (Vol. 64, Issue 4, pp. 416–425). Canadian Science Publishing. https://doi.org/10.1139/gen-2020-0131

Maglaveras, N., Vassiis, K., Koutkias, V., & Chouvarda, I. (2016). Integrated care and connected health approaches leveraging personalised health through big data analytics. *Studies in Health Technology and Informatics*, *224*(May), 117–122. https://doi.org/10.3233/978-1-61499-653-8-117

Martínez-García, M., & Hernández-Lemus, E. (2022). Data Integration Challenges for Machine Learning in Precision Medicine. In *Frontiers in Medicine* (Vol. 8). https://doi.org/10.3389/fmed.2021.784455

McDonald, C. J., Weiner, M., & Hui, S. L. (2000). Deaths due to medical errors are exaggerated in Institute of Medicine Report. In *JAMA* (Vol. 284, Issue 1). https://doi.org/10.1001/jama.284.1.93

McPadden, J., Durant, T. J. S., Bunch, D. R., Coppi, A., Price, N., Rodgerson, K., Torre, C. J., Byron, W., Hsiao, A. L., Krumholz, H. M., & Schulz, W. L. (2019). Health care and precision medicine research: Analysis of a scalable data science platform. *Journal of Medical Internet Research*, *21*(4), 1–11. https://doi.org/10.2196/13043

Menezes, B. C., Kelly, J. D., Leal, A. G., & Le Roux, G. C. (2019). Predictive, prescriptive and detective analytics for smart manufacturing in the information age. *IFAC-PapersOnLine*, *52*(1), 568–573. https://doi.org/10.1016/j.ifacol.2019.06.123

Mesko, B. (2017a). The role of artificial intelligence in precision medicine. In *Expert Review of Precision Medicine and Drug Development* (Vol. 2, Issue 5, pp. 239–241). https://doi.org/10.1080/23808993.2017.1380516

Mesko, B. (2017b). The role of artificial intelligence in precision medicine. *Expert Review of Precision Medicine and Drug Development*, *2*(5), 239–241. https://doi.org/10.1080/23808993.2017.1380516

Mesko, B. (2017c). The role of artificial intelligence in precision medicine. *Http://Dx.Doi.Org/10.1080/23808993.2017.1380516*, *2*(5), 239–241. https://doi.org/10.1080/23808993.2017.1380516

Michael A. Eierman , Fred Niederman, C. A. c. (1995). sunx DSS theory : A model of constructs and relationships. *Decision Support Systems*, *9236*(94).

Milewa, T. (2009). Health care, consumerism and the politics of identity. In *The New Sociology of the Health Service*. https://doi.org/10.4324/9780203879740-13

Mosavi, N. S., Ribeiro, E., Sampaio, A., & Santos, M. F. (2023). Data mining techniques in psychotherapy: applications for studying therapeutic alliance. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-43366-6

Mosavi, N. S., & Santos, M. F. (n.d.). *Characteristics of the Intelligent Decision Support System for Precision Medicine (IDSS4PM)* (N. D. and A. J. Xin-She Yang, Simon Sherratt, Ed.; pp. 1–8). Springer Nature.

Mosavi, N. S., & Santos, M. F. (2020). How prescriptive analytics influences decision making in precision medicine. *Procedia Computer Science*, *177*, 528–533. https://doi.org/10.1016/j.procs.2020.10.073

Mosavi, N. S., & Santos, M. F. (2021a). To what extent healthcare analytics influences decision making in precision medicine. *Procedia Computer Science*, *198*(2021), 353–359. https://doi.org/10.1016/j.procs.2021.12.253

Mosavi, N. S., & Santos, M. F. (2021b). To what extent healthcare analytics influences decision making in precision medicine. *Procedia Computer Science*, *198*(2021), 353–359. https://doi.org/10.1016/j.procs.2021.12.253

Mosavi, N. S., & Santos, M. F. (2022a). *Intelligent Decision Support System for Precision Medicine (IDSS 4 PM)*. *3*(Ic3k), 29–36. https://doi.org/10.1007/978-981-16-2502-2_4

Mosavi, N. S., & Santos, M. F. (2022b). Internet of things for precision intensive medicine. *Procedia Computer Science*, *201*(C), 732–737. https://doi.org/10.1016/j.procs.2022.03.099

Nacional, U. T. (2003). Development of a Relational Database Management System . *JCS&T*, *3*(2).

Naik, K., Goyal, R. K., Foschini, L., Chak, C. W., Thielscher, C., Zhu, H., Lu, J., Lehár, J., Pacanoswki, M. A., Terranova, N., Mehta, N., Korsbo, N., Fakhouri, T., Liu, Q., & Gobburu, J. (2023). Current Status and Future Directions: The Application of Artificial Intelligence/Machine Learning for Precision Medicine. In *Clinical Pharmacology and Therapeutics*. https://doi.org/10.1002/cpt.3152

Naithani, N., Sinha, S., Misra, P., Vasudevan, B., & Sahu, R. (2021). Precision medicine: Concept and tools. In *Medical Journal Armed Forces India* (Vol. 77, Issue 3). https://doi.org/10.1016/j.mjafi.2021.06.021

Nakache, D. (2003). PROBLEMS IN DESIGNING HUGE DATAWAREHOUSES AND DATAMARTS. *9th Americas Conference on Information Systems, AMCIS 2003*.

P. Bernus, P. Nemes, G. S. (1946). Conference on congestive heart failure; case history. In *Master schedule of professional staff proceedings ... United States. Army. Halloran general hospital, St. George, Staten island, N. Y* (Vol. 1, Issue 9).

Palanisamy, V., & Thirunavukarasu, R. (2019). Implications of big data analytics in developing healthcare frameworks – A review. *Journal of King Saud University - Computer and Information Sciences*, *31*(4), 415–425. https://doi.org/10.1016/j.jksuci.2017.12.007

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Pelter, M. N., & Druz, R. S. (2024). Precision medicine: Hype or hope? In *Trends in Cardiovascular Medicine* (Vol. 34, Issue 2). https://doi.org/10.1016/j.tcm.2022.11.001

Pencina, M. J., & Peterson, E. D. (2016). Moving from clinical trials to precision medicine: The role for predictive modeling. *JAMA - Journal of the American Medical Association*, *315*(16), 1713–1714. https://doi.org/10.1001/jama.2016.4839

Phillips-Wren, G. (2008). Intelligent decision support to assist real-time collaboration. *2008 International Symposium on Collaborative Technologies and Systems, CTS'08*, 375. https://doi.org/10.1109/CTS.2008.4543953

PHILLIPS-WREN, G. (2012). Ai Tools in Decision Making Support Systems: a Review. *International Journal on Artificial Intelligence Tools*, *21*(02), 1240005. https://doi.org/10.1142/s0218213012400052

Phillips-Wren, G., Daly, M., Power, D., & Adam, F. (2017). A Critical Review of Decision Support Systems Foundational Articles. *AMCIS 2017 Proceedings*, *2005*, 1–10. https://aisel.aisnet.org/amcis2017/AdvancesIS/Presentations/18

Power, D. (2017). Supporting Decision-Makers: An Expanded Framework. *Proceedings of the 2001 InSITE Conference*, *June*. https://doi.org/10.28945/2384

Power, D. J. (2000). Web-Based and Model-Driven Decision Support Systems: Concepts and Issues. *AMCIS 2000 Proceedings*, 352–355. http://aisel.aisnet.org/amcis2000/387

Power, D. J. (2002). Decision Support Systems: Concepts and Resources for Managers. In *Information Systems Management* (Vol. 20, Issue 4).

Power, D. J. (2008). Understanding Data-Driven Decision Support Systems. *Information Systems Management*, *25*(2), 149–154. https://doi.org/10.1080/10580530801941124

Praful Bharadiya, J. (2023). A Comparative Study of Business Intelligence and Artificial Intelligence with Big Data Analytics. *American Journal of Artificial Intelligence*. https://doi.org/10.11648/j.ajai.20230701.14

Prakash, N., & Sarkar, A. (2015). Development of an intelligent decision support system for an hierarchical business organization. *2015 International Conference and Workshop on Computing and Communication, IEMCON 2015*, 1–8. https://doi.org/10.1109/IEMCON.2015.7344440

Quazi, S. (2022). Artificial intelligence and machine learning in precision and genomic medicine. In *Medical Oncology* (Vol. 39, Issue 8). https://doi.org/10.1007/s12032-022-01711-1

Ramesh Sharda, Dursun Delen, E. T. (2014). Business intelligence and analytics. In *Higher Education Strategy and Planning*. https://doi.org/10.4324/9781315206455-12

Republic Of Korea AI Strategy. (2019). "Toward AI World Leader, beyond IT" National Strategy for Artificial intelligence. In *The Government of the Republic of Korea*.

Reza Kheirandish, S. M. (2019). Herbert Simon, innovation, and heuristics. *Mind and Society*, *17*(1), 97–109. https://doi.org/10.1007/s11299-019-00203-6

Richard, U., & Averweg, F. (n.d.). *Decision-making support systems : Theory & practice*.

Sadat Mosavi, N., & Filipe Santos, M. (2021). *Adoption of Precision Medicine; Limitations and Considerations* (David C. Wyld et al. (Eds), Ed.; pp. 13–24). AIRCC Publishing Corporation. https://doi.org/10.5121/csit.2021.110302

Sanchez-Pinto, L. N., Bhavani, S. V., Atreya, M. R., & Sinha, P. (2023). Leveraging Data Science and Novel Technologies to Develop and Implement Precision Medicine Strategies in Critical Care. In *Critical Care Clinics* (Vol. 39, Issue 4). https://doi.org/10.1016/j.ccc.2023.03.002

Schork, N. J. (2015). Personalized medicine: Time for one-person trials. *Nature*, *520*(7549), 609–611. https://doi.org/10.1038/520609a

Shim J. P., Warkentin M., Courtney J. F., Power D. J., Sharda R., & Carlsson C. (2002). Past, {P}resent, and {F}uture of {D}ecision {S}upport {T}echnology. *Decision Support Systems*, *33*(2), 111–126. www.elsevier.com/locate/dsw

Shrank, W. H. (2017). Primary Care Practice Transformation and the Rise of Consumerism. *Journal of General Internal Medicine*, *32*(4). https://doi.org/10.1007/s11606-016-3946-1

Simon, H. A. (1959). Theories of Decision-Making and Behavioral Science. *The American Economic Review*, *49*(3), 253–283.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, *70*, 263–286. https://doi.org/10.1016/j.jbusres.2016.08.001

Skulimowski, A. M. J. (2016). *Knowledge, Information and Creativity Support Systems: Recent Trends, Advances and Solutions*. *364*(February 2016). https://doi.org/10.1007/978-3-319-19090-7

Sousa, M. J., Pesqueira, A. M., Lemos, C., Sousa, M., & Rocha, Á. (2019). Decision-Making based on Big Data Analytics for People Management in Healthcare Organizations. *Journal of Medical Systems*, *43*(9). https://doi.org/10.1007/s10916-019-1419-x

Sprague, R. H. (1980). A framework for the development of decision support systems. *MIS Quarterly: Management Information Systems*, *4*(4), 1–26. https://doi.org/10.2307/248957

Sriram, R. D., & Subrahmanian, E. (2020). Transforming Health Care through Digital Revolutions. In *Journal of the Indian Institute of Science* (Vol. 100, Issue 4). https://doi.org/10.1007/s41745-020-00195-0

Srivani, M., Murugappan, A., & Mala, T. (2023). Cognitive computing technological trends and future research directions in healthcare — A systematic literature review. In *Artificial Intelligence in Medicine* (Vol. 138). https://doi.org/10.1016/j.artmed.2023.102513

Stanek, S., Drosio, S., & Namyslo, J. (2014). Experiences with building modern communication-driven decision support. *Frontiers in Artificial Intelligence and Applications*, *261*. https://doi.org/10.3233/978-1-61499-399-5-483

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. In *npj Digital Medicine* (Vol. 3, Issue 1). https://doi.org/10.1038/s41746-020-0221-y

Tarafdar, M., Beath, C. M., & Ross, J. W. (2017). Enterprise Cognitive Computing Applications: Opportunities and Challenges. *IT Professional*, *19*(4). https://doi.org/10.1109/MITP.2017.3051321

Tarassoli, S. P. (2019). Artificial intelligence, regenerative surgery, robotics? What is realistic for the future of surgery? *Annals of Medicine and Surgery*, *41*(January), 53–55. https://doi.org/10.1016/j.amsu.2019.04.001

Tran, T. Q. B., du Toit, C., & Padmanabhan, S. (2021). Artificial intelligence in healthcare—the road to precision medicine. *Journal of Hospital Management and Health Policy*, *5*(September), 29–29. https://doi.org/10.21037/jhmhp-20-132

Vegter, M. W. (2018). Towards precision medicine; a new biomedical cosmology. *Medicine, Health Care and Philosophy*, *21*(4), 443–456. https://doi.org/10.1007/s11019-018-9828-z

Velmovitsky, P. E., Bublitz, F. M., Fadrique, L. X., & Morita, P. P. (2021). Blockchain applications in health care and public health: Increased transparency. In *JMIR Medical Informatics* (Vol. 9, Issue 6). https://doi.org/10.2196/20713

Wagner, J. B. (2019). Genomics and precision medicine to direct statin use in the young. *Progress in Pediatric Cardiology*, *54*(August), 101145. https://doi.org/10.1016/j.ppedcard.2019.101145

Wang, L. (2017). Big Data in Healthcare: A New Frontier in Personalized Medicine. *Open Access Journal of Translational Medicine & Research*, *1*(1), 15–18. https://doi.org/10.15406/oajtmr.2017.01.00005

Watson, H. J. (2017a). Preparing for the cognitive generation of decision support. *MIS Quarterly Executive*, *16*(3).

Watson, H. J. (2017b). The Cognitive Decision-Support Generation. *BUSINESS INTELLIGENCE JOURNAL*, *22*(2).

Wilkinson, J., Arnold, K. F., Murray, E. J., van Smeden, M., Carr, K., Sippy, R., de Kamps, M., Beam, A., Konigorski, S., Lippert, C., Gilthorpe, M. S., & Tennant, P. W. G. (2020). Time to reality check the promises of machine learning-powered precision medicine. In *The Lancet Digital Health* (Vol. 2, Issue 12). https://doi.org/10.1016/S2589-7500(20)30200-4

Wong, A. K. C., & Wang, Y. (2003). Pattern discovery: A data driven approach to decision support. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, *33*(1). https://doi.org/10.1109/TSMCC.2003.809869

Woodside, A. G., Lai, W., Kim, K. H., Jung, D. K., Woodside, A. G., Lai, W., Kim, K. H., Jung, D. K., Fiori, S., Dyner, I., Franco, C. J., Herling, R. W., Popescu, M., Keller, J. M., March, J. G., The, S., Journal, B., Autumn, N., March, J. G., … Schmidt-Daffy, M. (2016). Bounded Rationality , Ambiguity , and the Engineering of Choice Published by : RAND Corporation Stable URL : http://www.jstor.org/stable/3003600 REFERENCES Linked references are available on JSTOR for this article : Bounded rationality , ambiguity , and t. *Systems Research and Behavioral Science*, *23*(4), 493–513. https://doi.org/10.1016/j.aap.2013.10.028

Yang, X., Huang, K., Yang, D., Zhao, W., & Zhou, X. (2024). Biomedical Big Data Technologies, Applications, and Challenges for Precision Medicine: A Review. In *Global Challenges* (Vol. 8, Issue 1). https://doi.org/10.1002/gch2.202300163

Zhou, F., Yang, B., Li, L., & Chen, Z. (2008). Overview of the new types of intelligent decision support system. *3rd International Conference on Innovative Computing Information and Control, ICICIC'08*, 1–4. https://doi.org/10.1109/ICICIC.2008.412