Emotion Detection in School Failure Prevention

Renata Sofia Vieira Magalhães

UMinho | 2024

Universidade do Minho
Escola de Letras, Artes e Ciências Humanas

Renata Sofia Vieira Magalhães

Emotion Detection in School Failure Prevention

march 2024

University of Minho
School of Arts and Humanities

Renata Sofia Vieira Magalhães

Emotion Detection in School Failure Prevention

march 2024

University of Minho
School of Arts and Humanities

Renata Sofia Vieira Magalhães

Emotion Detection in School Failure Prevention

Master's Dissertation in Digital Humanities

Dissertation supervised by
Sílvia Lima Gonçalves Araújo
Dalila Alves Durães

march 2024

Acknowledgements

First and foremost, I would like to thank my supervisors, Professor Sílvia Araújo and Professor Dalila Durães for believing in me and guiding me through my academic journey.

Secondly, I would like to thank the ISLab and more specifically professor Francisco Marcondes for all the guidance throughout the writing of this dissertation. Your help and encouragement have been vital.

To all the professors whom I have encountered throughout my master's I would like to show my deepest gratitude for having taught me all the skills necessary that led me to where I am now. It has been a pleasure to have attended every single one of your classes.

To my colleagues, I could not have done this without you guys. Thank you for always believe in me, even and especially when I struggled to.

Last but not least, I am also immensely grateful for my family and friends. This would not have been possible without your continuous support and motivation.

Statement of Integrity

I hereby declare having conducted this academic work with integrity.

I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

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

University of Minho, Braga, march 2024

Renata Sofia Vieira Magalhães

Abstract

The issue of school failure and dropout is persisting and alarming as it presents repercussions to society on the whole. Emotion detection, also known as affective computing, is an evolving field of research focused on enabling intelligent systems to detect, perceive, understand, and interpret human emotions. This has great potential in many fields, including education.

This dissertation presents an emotion detection approach through tools that were not developed for that purpose. By annotating a Likert-Scale questionnaire, student emotional traits were extracted and analyzed, including correlation to students' risk of failing. Additionally, four machine learning models were tested and Support Vector Machine showed the best performance in accurately predicting student emotions.

This approach was applied to a sample of 845 students from northern Portugal, within the context of the the Northern Regional Operational Programme (NORTE 2020), under Portugal 2020 within the scope of the project "Hello: Plataforma inteligente para o combate ao insucesso escolar", Ref. NORTE-01-0247-FEDER-047004, which aimed to develop a digital platform comprising a conversational agent and an education intelligence system capable of predicting students' risk of failure, allowing for a timely detection and consequent action.

Findings report that anxiety was the emotion that presented the highest scores, with male students scoring slightly higher than female students for this particular emotion. It was also found that there was a gradual decrease in positive emotions (happiness, trust, optimism, interest) as students progressed from the fifth to the ninth grade, with the eighth grade showing the lowest scores overall, though with positive emotions declining more than negative emotions (boredom, anxiety, distraction, shame). Students at low risk of failure presented higher scores for all emotions compared to other risk clusters, and higher optimism was associated with lower failure risk.

This emotion detection model presents itself as an approach to combating school failure, contributing to the identification of students emotional traits.

Keywords Emotion Detection, School Failure, Machine Learning, Natural Language Processing

Resumo

O problema do insucesso e do abandono escolar é persistente e alarmante, pois tem repercussões na sociedade em geral. A deteção de emoções, também conhecida como computação afectiva, é um campo de investigação em evolução que visa permitir que os sistemas inteligentes detectem, percebam, compreendam e interpretem as emoções humanas. Isto tem um grande potencial em muitos domínios, incluindo a educação.

Esta dissertação apresenta uma abordagem de deteção de emoções através de ferramentas que não foram desenvolvidas para esse fim. Através da anotação de um questionário de escala de Likert, foram extraídos e analisados os traços emocionais dos alunos, incluindo a correlação com o risco de reprovação. Além disso, foram testados quatro modelos de aprendizagem automática e o Support Vector Machine apresentou o melhor desempenho na previsão exacta das emoções dos alunos.

Estas técnicas foi aplicada a uma amostra de 845 alunos do norte de Portugal, no âmbito do Northern Regional Operational Programme (NORTE 2020), under Portugal 2020 within the scope of the project "Hello: Plataforma inteligente para o combate ao insucesso escolar", Ref. NORTE-01-0247-FEDER-047004, que teve como objetivo desenvolver uma plataforma digital composta por um agente conversacional e um sistema de inteligência educacional capaz de prever o risco de insucesso dos alunos, permitindo a sua deteção atempada e consequente atuação.

Os resultados indicam que a ansiedade foi a emoção que apresentou as pontuações mais elevadas, com os alunos do sexo masculino a obterem pontuações ligeiramente superiores às das alunas nesta emoção específica. Verificou-se também que houve um decréscimo gradual das emoções positivas (felicidade, confiança, otimismo, interesse) à medida que os alunos progrediam do quinto para o nono ano, com o oitavo ano a apresentar as pontuações mais baixas em geral, embora com as emoções positivas a diminuírem mais do que as emoções negativas (aborrecimento, ansiedade, distração, vergonha). Os alunos com baixo risco de insucesso apresentaram pontuações mais elevadas para todas as emoções, em comparação com outros grupos de risco, e um maior otimismo foi associado a um menor risco de insucesso.

Este modelo de deteção de emoções apresenta-se como uma abordagem de combate ao insucesso escolar, contribuindo para a identificação das características emocionais dos alunos.

Palavras-chave Deteção de Emoção, Insucesso Escolar, Machine Learning, Processamento de Linguagem Natural

Contents

List of Figures

List of Tables

Acronyms

AI Artificial Intelligence. ix, xii, 1, 11, 12, 13, 19, 35

DL Deep Learning. ix, xii, 12, 18, 24

DM Data Mining. xii, 2, 8, 11, 18

DT Decision Tree. xii, 16

EDM Educational Data Mining. xii, 18

LR Linear Regression. xii, 16

MAE Mean Absolute Error. x, xii, 40, 54, 55, 57, 59

ML Machine Learning. ix, xii, 2, 8, 11, 12, 13, 14, 18, 22, 24, 39, 41, 43, 61

MSE Mean Squared Error. x, xii, 40, 54, 55, 57, 59

NLP Natural Language Processing. ix, xii, 2, 11, 12, 19, 20, 21, 22, 23

SA Sentiment Analysis. ix, xii, 11, 22, 23, 24, 43

SVM Support Vector Machine. xii, 17

Chapter 1

Introduction

The first chapter serves as an introduction to the research carried out in this dissertation. It aims to give a clear overview of the context, reasons behind the research, and the objectives it seeks to achieve.

1.1 Context and Motivation

The present dissertation is integrated into the Hello project - "HELLO: Plataforma inteligente para o combate ao insucesso escolar", with the reference HELLO - 047004. This project aims to develop a digital platform in the field of Education which acts as mediator between all the parties in the school domain: Student, School, Family and Community.

The issue of school failure and dropout is concerning as it has reaching implications for individuals and society on the whole. The consequences of these educational challenges extend beyond the academic realm, influencing future employment opportunities, socio-economic status, and overall well-being.

With the need of it being considered as a long-term process rather than an event, literature claims that the issue of school dropout marks the conclusion of a long trajectory of academic disengagement and failure that often begins in early childhood, and sometimes even prior to a child's entry into school (Samuel and Burger, 2020).

Despite the prevalent focus on higher education or online learning in existing studies, there is a crucial need to extend research efforts to mandatory education. Exploring this aspect not only allows for a deeper understanding of the problem of school failure but also paves the way for effective preventive measures.

In this context, **Artificial Intelligence (AI)** emerges as a promising ally in the quest to mitigate school failure and dropout rates, allowing for the prediction and consequent prevention of such. A digital platform such as the one developed in this project's context gathers all the necessary tools to make that happen.

Witnessing firsthand the challenges that students face and the complexities involved in identifying early signs of academic struggle has fueled my commitment to finding innovative solutions. Drawing on my work experience as a teacher, I recognized the need for a proactive approach to address these issues, prompting me to contribute my insights and expertise to this transformative initiative.

The advances in **Machine Learning (ML)**, **Data Mining (DM)** and **Natural Language Processing (NLP)** can help identify students at risk in a timmely manner and allow for preventive action.

1.2 Hello Project

The HELLO project centers on the collaboration with a public government initiative dedicated to preventing school failure and dropout in K-12 education. The focus is on developing a software platform integrating an innovative risk matrix and advanced artificial intelligence algorithms. The platform prioritizes two critical areas: identifying school failure or dropout risks in a timely manner to facilitate prompt intervention and promoting effective study and learning practices to enhance overall school success. Importantly, the platform tailors its recommendations to individual student situations, encompassing both performance and behavior. Students are categorized into four distinct risk clusters, enabling a personalized approach to follow-up and support.

The platform integrates the following and distinctive elements:

- Innovative Risk Assessment Matrix: The platform employs a risk assessment matrix that informs the system, generating profiles and risk clusters across dimensions relevant to individual students. This phase relies on primary causes and indicators of school failure, facilitating the early detection of risk indicators for timely intervention.

- Predictive System with Artificial Intelligence Algorithms: The platform incorporates a predictive system based on artificial intelligence algorithms to identify risks promptly. The research explores and compares various machine learning classification algorithms, aiming to establish a decision support system. Additionally, the system creates profiles for each student, enabling the early detection of risky behavior.

- Conversational Agent: An integral component of the platform is a conversational agent that accompanies, motivates, and advises students based on profile and risk cluster. It promotes appropriate study strategies and motivation for learning, and overall school success.

This platform targets middle school students, recognizing their increased vulnerability to mental health issues, depression, and involvement in criminal activities upon dropping out (Freeman et al., 2015) (Peguero, 2011), seeking to provide a competitive technological solution, optimizing analysis time and data collection efficiency, ultimately granting educators and psychologists more time for direct intervention with students.

The team working on the project is multidisciplinary, including educational psychologists, a technology company and IT researchers. The first were in charge of crafting age-appropriate scripts to ensure the conversational flow suits children of varying ages and educational levels. These meticulously designed scripts serve as a guide for the team of computer science researchers tasked with constructing the conversational agent. Concurrently, a technology company specializing in software solutions for the Education and Training sectors is entrusted with deploying the agent. Through collaborative efforts, this diverse team aims to harness the strengths of each discipline to develop a digital platform capable of identifying students at risk of academic failure and dropout.

1.2.1 Pilot Study

This research initially sought to empirically validate a theoretical model for school failure, established through a meticulous literature review, psychometric analysis of instruments, and expert discussions, carried out by researchers in the field of Educational Psychology. The analysis focused on assessing the relevance of various predictors (individual, school, family) and their impact on indicators of school failure, such as academic performance in Portuguese and Mathematics, satisfaction with performance, perception of alignment between performance and perceived ability, and student involvement in school.

Data collection occurred at the onset of the third term in the 2021-2022 school year, encompassing students from the 2nd and 3rd cycles of basic education in public and private schools in the northern region of the country. The sample comprised 845 students, predominantly male (n=430, 50.9%), aged 10 to 17 (M=12.25, SD=1.541), with approximately half in the 2nd Cycle (n=422, 49.9%).

The theoretical model underwent empirical testing using Structural Equation Models and the IBM SPSS Analysis of Moment Structures (AMOS) software, v.28. The fit of the proposed model to the student database obtained from school administrations was favorable ($X^2/(9) = 1.708$, $p = .081$; CFI = .998; TLI = .986; RMR = .031; RMSEA = .029). All variables demonstrated significant explanatory power over school failure outcomes, except for the family support variable, which did not exhibit significant effects.

To contribute to a broader discussion and construct a risk matrix for students' likelihood of school failure, cluster analyses were conducted. Cluster analysis, an exploratory multivariate technique, identified groups of individuals prone to school failure based on specific predictors and defined distinct levels of risk.

In an initial analysis, the program defined clusters autonomously, resulting in 3 clusters. The distribution of samples across these clusters was considered adequate, and the quality of the cluster model was assessed as average. The most influential predictors in defining the clusters were, in decreasing order of importance: the Maths grade, followed by the student's Portuguese grade in the previous term, the student's satisfaction with their grades, the student's involvement in school and the perception of whether their grades reflect their ability.

In a second analysis, the Educational Psychology researchers specifically defined 4 clusters to ensure theoretical and practical consistency in the interpretation of the clusters identified. In this analysis, a maximum limit of 4 was set by the researchers, seeking theoretical and practical consistency in the interpretation of the identified clusters. The distribution of samples among the clusters was adequate and more balanced than in the previous analysis, resulting in an average quality of the cluster model. The most crucial predictors for defining the clusters, in descending order of importance, were the student's Portuguese grade in the previous term, the student's Maths grade in the previous term, the student's satisfaction with their grades, the student's involvement in school and, finally, the student's perception of whether their grades reflect their ability. This risk cluster distribution can be seen in 15.

- Cluster 1. High risk of school failure.
- Cluster 2. Medium-high risk of school failure.
- Cluster 3. Medium-low risk of school failure.
- Cluster 4. Low risk of school failure.

1.2.2 Conversational Agent

The first ever thought of a chatbot occured in 1950 when Alan Turing wondered if a computer program could engage in conversation with humans without them recognizing that the communicator was artificial. This is known as the Turing Test (Adamopoulou and Moussiades, 2020). However, it was not until 1966 that the first chatbot came to life. ELIZA emulated the functions of a psychotherapist by rephrasing user sentences in the form of questions and while its communicative capabilities were constrained, ELIZA served as a source of inspiration for the subsequent evolution of other chatbots, varying in features, methodologies and capabilities (Shawar and Atwell, 2015).

The framework chosen for this project was RASA. It employs a cutting-edge Transformer-based architecture to discern intricate relationships between words (Rasa, 2023) and is renowned for its Natural Language Processing versatility and learning capabilities, enhancing "pattern matching" through extrapolation from instances (Devi et al., 2021).

Rasa's framework general structure comprises components like Rasa NLU, Rasa rules, Rasa stories, actions, and domain.yml (see Listings 1.1, 1.2, 1.3).

Listing 1.1: Example of an action.

```
class AskForMaterialAction(Action):
    def name(self) -> Text:
        return "action_ask_material"
    def run(self,
        dispatcher: CollectingDispatcher,
        tracker: Tracker,
        domain: DomainDict,
    ) -> List[EventType]:

        dispatcher.utter_message(image = "images/hello/student_with_good_grades/traz_o_material.png")
        dispatcher.utter_message(text=f"Brings␣school␣supplies.")
        dispatcher.utter_message(buttons= [{"title" : "👍", "payload" : "👍"},
                                           {"title" : "👎", "payload" : "👎"}])
```

Listing 1.2: Example of an rule.

```
- rule: Say goodbye anytime the user says goodbye
  steps:
  - intent: goodbye
  - action: utter_goodbye
```

Listing 1.3: Example of an intent.

```
- intent: request_help
  examples: |
    - talk to teacher
    - speak to teacher
    - want to contact teacher
    - want to get in touch with teacher
    - is the teacher there?
    - may I speak with the teacher?
    - when can the teacher talk to me?
```

The NLU component is responsible for understanding user input and extracting relevant information, such as user intent and entities. Rules define conditional statements to control conversation flow, stories represent predefined sequences of actions, and actions specify tasks the chatbot can perform. The domain file defines the chatbot's vocabulary, actions, and responses, encompassing intents, entities, actions, and templates.

In this project, the conversational agent incorporates 502 actions, 63 intents, 10 rules, and 4 stories. It features button-type answers and emojis to engage K-12 students. Additionally, the chatbot can detect inappropriate words, redirect conversations, and suggest seeking help from the class director in case of confusion.

The chatbot is developed in both Portuguese, its original language, and English, with equivalent characteristics and language registers in both versions, ensuring a consistent experience for users.

Chatbot Sessions

A total of four activities were scripted and tested in a controlled environment with 51 participants.

Objectives were set and specific activities encompassing closed and open questions, emojis, and images were designed based on each risk cluster. The aim was to create an interactive context aligned with students' reality, fostering engagement and thereby enhancing response effectiveness. This means that elements that foster the participants' relationship with the platform were included, such as humor, playful activities or community role models to reinforce targeted objectives or behaviors.

For each of the four sessions where students engaged with the chatbot, general and specific objectives and activities were designed. Key objectives included fostering motivation for learning, developing study strategies, anticipating consequences, reinforcing the instrumentality of learning, and providing guidance in setting concrete, realistic, and assessable goals. Bases on the four risk clusters, activities were developed and adjusted accordingly. Throughout the sessions, interactions aligned with predefined options ensured the activities' consistency. Examples of activities include "exploring talents" [Activity 1], "How to assist a classmate with English?" [Activity 2], and "strategies for success" [Activity 3].

- First activity. The initial activity aims to establish a connection with students through the chatbot. It begins by introducing itself and prompting students to share positive characteristics about themselves, along with a brief physical description. The interaction proceeds with a questionnaire designed by the project's psychological team, addressing topics such as study habits, views on education, and family/friends support. The conversational agent then discusses the students' hobbies, illustrating how these activities can benefit academic performance. For example, if a student enjoys sports, the agent highlights the teamwork involved, similar to group projects in school. This activity not only helps categorize students based on questionnaire responses but also seeks to show that school can be enjoyable, encouraging the integration of daily and leisure activities with academic knowledge and strategies.

- Second activity. In the second activity, the chatbot initiates the interaction by greeting students and inquiring about their week. To create a comfortable and engaging atmosphere, a light-hearted joke is shared. Subsequently, a scenario is presented where a classmate is facing challenges in class, and the student is asked about strategies to improve their grade. Several strategies are explored in-depth, allowing students to understand how to enhance their own academic performance. It's noteworthy that the activity aligns with the assigned cluster, tailoring specific strategies for high-risk students and considering alternative approaches for low-risk individuals. The objective is to help students anticipate potential outcomes and empower them to improve their academic performance.

- Third activity. In the third activity, a more detailed exploration of study techniques is provided, illustrating the potential consequences of applying or not applying these techniques. The activity includes a practical exercise where students plan a birthday party using the outlined steps. The chatbot demonstrates how systematically following these steps can lead to long-term benefits. This real-life example aims to show students how they can implement these strategies in their academic pursuits, ultimately improving their performance and grades.

- Fourth activity. Moving on to the fourth activity, it reflects on the third activity by asking students if they have broken down their goals into smaller steps and attempted to achieve each step. The chatbot then discusses the students' role models or individuals they look up to, providing examples of how these figures succeeded in school to reach their current status. Celebrities from various fields, including music and sports, are mentioned. The activity concludes by presenting part of the questionnaire from the first activity, allowing for a comparison of results between the beginning and end of the school year. The chatbot bids farewell to the students, leaving them with a motivational message.

1.3 Objectives and Research Questions

This dissertation aims to classify and analyse students' emotional traits through data retrieved from a digital assistant that monitors student data to create a personalized profile of the student that will help in the identification of school failure and dropout prevention.

The main objective lies in the development of an emotion detection method using tools that were not intended for that purpose.

To achieve the general objective described in the previous paragraph, several tasks need to be carried out:

- Study of user profiles;

- *Profiling* - Preparation of data sources and data extraction;

- Studying, understanding and visualising data;
- Modelling and data preparation and pre-processing;
- Study of suitable *machine learning* algorithms for predictive analysis of emotions;
- Comparison of the various *machine learning* models for predictive analysis;
- Study and analysis of emotions;

To achieve the general objective described in the previous paragraph, it aims to answer the following research questions:

- **RQ1.** Is it possible to detect students' emotional traits through tools that were not intended for that purpose?
- **RQ2.** Is it possible to relate school failure and emotions?
- **RQ3.** How is school dropout and/or school failure correlated to emotions?

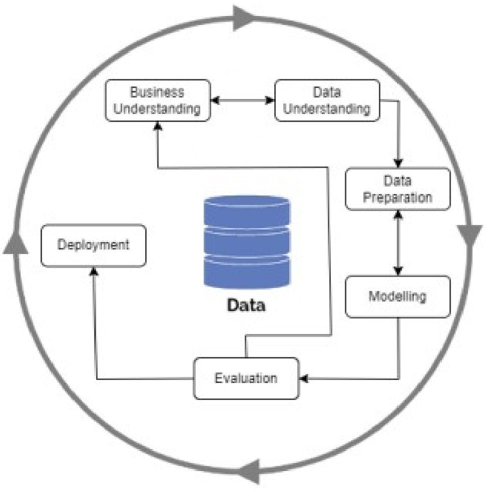
1.4 Methodology

In this research, the CRISP-DM methodology was employed to guide the investigation, ensuring a systematic and comprehensive approach to addressing all relevant contents and techniques at the appropriate stages. Originating from previous attempts to define knowledge discovery methodologies (Wirth and Hipp, 2000), CRISP-DM offers a comprehensive view of the Data Mining (DM) project lifecycle. It not only outlines the various phases of a project but also elucidates the tasks within each phase and their interrelations. It emphasizes that the relationships between **Data Mining (DM)** tasks can vary based on project objectives, user interests, and, most importantly, the nature of the data involved (Chapman et al., 2000).

As illustrated in Figure 1, the methodology consists of six distinct phases, serving as the foundational model for executing **Machine Learning (ML)** projects. Each phase delineates a clear path to be followed. However, it is crucial to acknowledge that the application of this technology must be adapted to the specific characteristics of each project. The utilization of CRISP-DM as a guiding framework ensures the proper contextualization of the methodology within the unique parameters of this research.

- Business Understanding: This phase directs efforts towards understanding the project objectives and requirements from a business perspective. Subsequently, the knowledge gained is converted into formulating a Data Mining problem and a preliminary plan designed to achieve the objectives.

Figure 1: CRISP-DM methodology lifecycle (Martínez-Plumed et al., 2019)



.

- Data Understanding: The understanding phase begins with data collection and continues with an initial study to gain insight into the quality of the data. It aims to identify interesting patterns that may form hypotheses about hidden information.

- Data Preparation: In the data preparation phase, a set of activities is performed on the initial data to construct the final dataset. These activities typically include selecting tables, records, and attributes, as well as transforming and cleaning the data.

- Modeling: In this phase, modeling techniques are selected and applied, and their parameters are calibrated. Considering that there are various techniques for the same type of problem, and some may have specific requirements regarding data formats, it may be necessary to revisit the data preparation phase.

- Evaluation: The evaluation phase expects the constructed model to have high quality. However, before proceeding with the final implementation, a thorough evaluation is necessary. This involves reviewing the executed steps to ensure that the model meets the previously stipulated business expectations. A crucial aspect of this phase is to determine if any key objectives have been left unanswered.

- Deployment: Creating the model does not mark the end of the project, as the knowledge gained must be organized and presented in a way that the client can use and truly derive value from the developed solution. In the deployment phase, models are often applied "live" within the decision-making processes of a specific organization. This phase can be executed simply through a report or more complexly, involving the implementation of a Data Mining process throughout the entire organization. It is essential for the client to understand the actions required to effectively utilize the created models.

1.5 Dissertation Structure

This dissertation is divided into 7 main chapters:

Introduction: The first chapter describes the context and motivation for carrying out this work, the general objective, as well as the necessary tasks that will need to be followed in order to achieve the objectives.It also presents the research questions that will be the object of this research, as well as the work methodology that will be used throughout the development of this dissertation.

Concepts and Technologies: This chapter contains the concepts and technologies that will form the theoretical basis for the project. Of particular note are the concepts of Machine Learning, Deep learning, Natural Language Processing and explainability of artificial intelligence.

Literature Review: consists of a literature review on school failure and dropout, artificial intelligence in education, emotion detection and a brief summary.

Methods and Methodologies: aims to describe the dataset used, as well as as the data understanding, data preparation, modeling and evaluation metrics.

Emotion Detection: includes a detailed description of the proposed emotion detection approach.

Results and Discussion: shows the main findings of the work and presented the emotion classification, correlation to risk failure, machine learning models and ends with a discussion.

Conclusions and future work: presents the main conclusions, the answer to the research questions, the contributions, as well prospects for future work.

Chapter 2

Concepts and Technologies

This section aims to describe concepts such as **Artificial Intelligence**, **Machine Learning**, Deep Learning, **Data Mining**, **Natural Language Processing**, **Sentiment Analysis** and Affective Computing.

2.1 Artificial Intelligence

Artificial Intelligence (AI) refers to the capacity of machines to execute tasks that traditionally demand human intelligence, such as reasoning, learning, and problem-solving (Swathi et al., 2019; Mondal, 2020).

Given its multidisciplinary nature, as it comprises several fields such as philosophy, mathematics, neuroscience, psychology and computer engineering, it can be difficult to find a consensual definition for AI, as all these fields have different methodologies and paradigms. Consequently, it is not easy to form a common language and understanding across all fields.

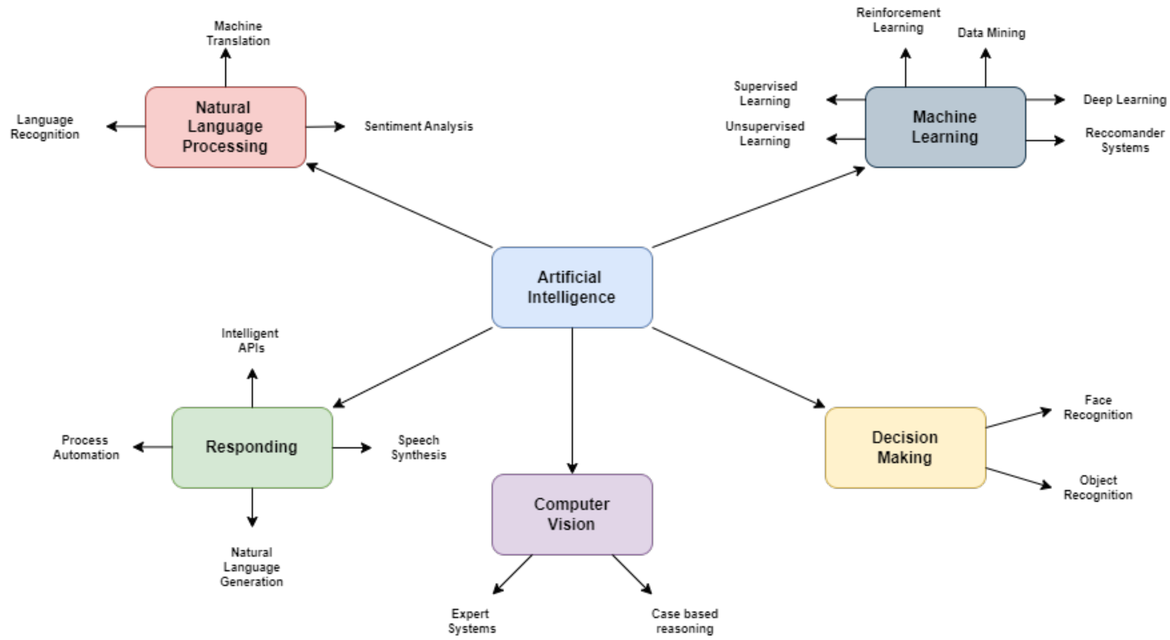
What is more and contrary to common perception, the concept of **AI** is not a recent development; its origins can be traced back to the 1950s, when the famous "Turing Test" first took place, when Alan Turing questioned the ability of machines to think with the query "Can machines think?".

Its current surge in popularity is attributed to its substantial economic impact and its transformative influence on diverse facets of human existence, extending to fields such as healthcare, customer service, education, and transportation (Kelly et al., 2023; Mondal, 2020) .

On the whole, **Artificial Intelligence** can be divided into into 5 main sub-fields: Natural Language Processing, Machine Learning, Decision Making, Computer Vision, and Responding, as represented in Figure 2.

It is common for these sub-fields to overlap amongst the application of **Artificial Intelligence (AI)** in the world. A good example of this is Robotics, a multidisciplinary field that draws from Natural Language Processing, Machine Learning, Decision Making, Computer Vision, and Responding **AI**. The field of robotics

Figure 2: A few subfields of **Artificial Intelligence**.



is an application area within **AI**, encompassing various **AI** sub-fields to create intelligent and adaptive machines. Robotics involves the integration of **AI** technologies to enable robots to perceive, learn, make decisions, and interact with the environment and humans in a purposeful and autonomous manner. The synergy between **AI** and robotics continues to the improvement of solutions in fields such as self-driving vehicles, industrial automation, healthcare, amongst others.

Figure 3: Overlapping of **AI**,**ML**, **DL** and **NLP**.

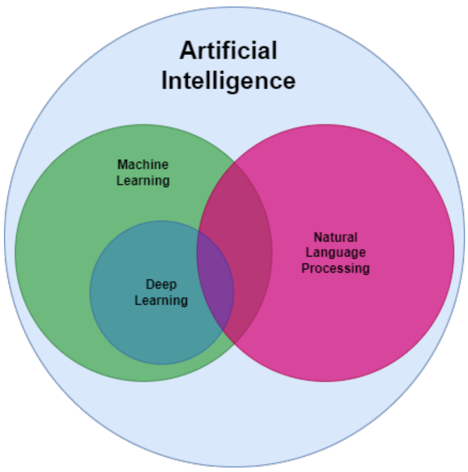
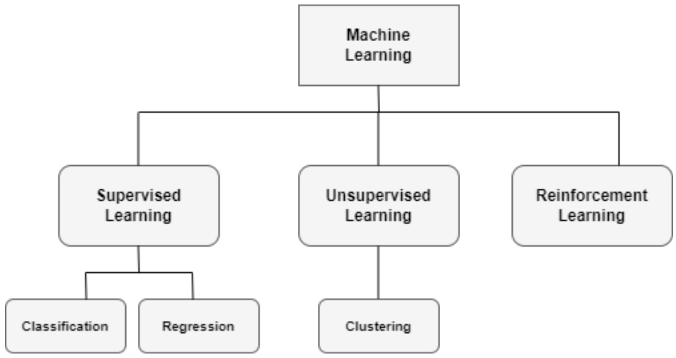


Figure 4: Types of learning in **Machine Learning**.



2.2 Machine Learning

Fundamentally, **ML** is the field of artificial intelligence with the purpose of developing computational learning models.

From a computational standpoint, machine learning is referred to as a machine's ability to improve its performance based on previous results. Currently, this field of artificial intelligence has various applications such as automatic language translation, facial recognition, voice recognition, recommendation systems, or object classification problems. It is through **ML** that machines learn how to deal with data more efficiently (Mahesh, 2020). Figure 3 present overlapping between **AI**, **ML**, and

With so many available datasets in the world today, there is a growing demand for machine learning. In situations where data is vast and exceeds human capacity for comprehension, or when it's challenging to discern what information to extract and analyze by mere observation, machine learning is used.

The concept of **ML** can be subdivided into types of learning: Supervised Learning, Unsupervised Learning, and Reinforcement Learning, as represented in Figure 4. This dissertation follows Supervised Learning.

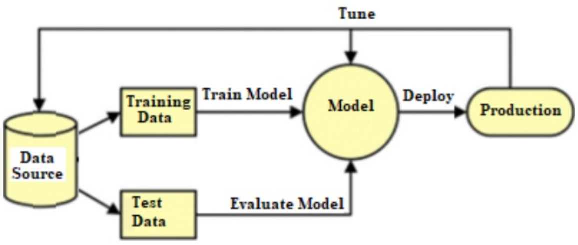
2.2.1 Supervised Learning

Supervised learning methods need external supervision to train machine learning models, hence the name supervised. They need guidance and additional information to return the result.

Supervised machine learning is employed to make predictions from data. To achieve this, it is essential to identify what needs to be predicted, referred to as the target variable.

This involves training models with labeled data, consisting of input-output pairs. Labeled data means that the output is already known. The models learn patterns from the data, enabling them to make

Figure 5: Supervised Learning Workflow (Mahesh, 2020).



predictions or classify new, unseen inputs. In this learning process, human intervention is essential in providing accurately labeled data and refining models based on known parameters (Talaei Khoei and Kaabouch, 2023). The worflow of a supervised learning model can be seen in Figure 5.

Methods like classification and regression are used to build predictive models, depending on the nature of the data available. Briefly, if the goal is to predict discrete categories or classes, such as a yes-no question or a categorical matter, a classification model would be more appropriate. As an example, if we aim to forecast whether a student will pass or fail by analyzing their historical profile, the resulting prediction will be either "pass" or "fail."

Regression models are well-suited for tasks where the target variable is a continuous quantity. For instance, if we are attempting to predict whether a student will pass or fail based on their extensive past records, the prediction output will be in numeric form, such as indicating a likelihood of "73%" to succeed.

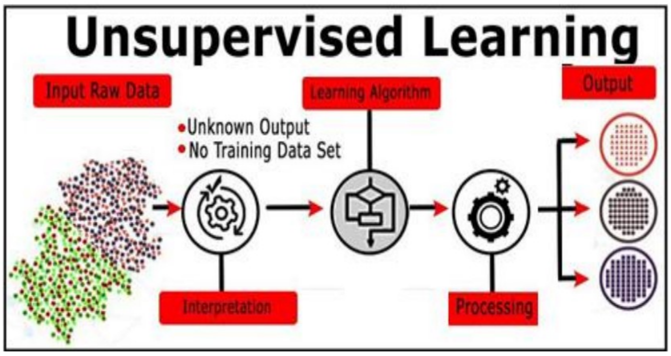
However, Regression models are well-suited for tasks where the target variable is a continuous quantity. For instance, if we are attempting to predict whether a student will pass or fail based on their extensive past records, the prediction output will be in numeric form, such as indicating a likelihood of "73%" to succeed.

Choosing the correct type of model depends on the nature of the data and the goal of the prediction.

Some of the commonly used supervised learning algorithms are linear regression, logistic regression, support vector machines, k -nearest neighbors, decision tree, random forest and knife base.

Recommender systems are a great example of supervised learning. Spotify and Netflix, for instance, utilize recommender systems to provide tailored recommendations to their user base, based on the user's previous consumption and interests. By knowing what the customer likes and consumes (labeled data), we use supervised learning **Machine Learning** algorithms to predict what the customer likes and provide recommendations.

Figure 6: Unupervised Learning Workflow (Mahesh, 2020).



2.2.2 Unsupervised Learning

Unsupervised learning techniques operate without the need for supervision during the training of models. Algorithms like K-means clustering, neural networks, principal component analysis, and hierarchical clustering analyze and group unlabeled data based on provided inputs. Unlabeled data does not have any specific targets or labels for prediction and only includes features representing the data.

In the absence of predefined labels, unsupervised learning algorithms sift through this type of data to uncover patterns, correlations, or clusters, without prior guidance on what to look for. The grouping, or clustering, of data samples is commonly utilized to explore and identify hidden patterns (Talaei Khoei and Kaabouch, 2023). Figure 6 shows the workflow of a unsupervised learning model.

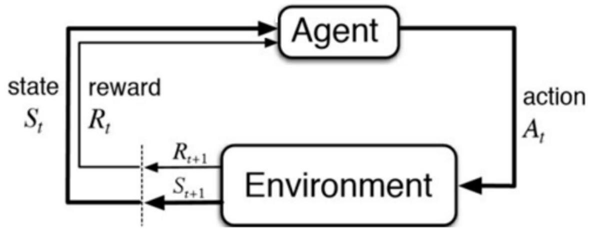
Unsupervised learning employs unlabeled data for machine training, where the model learns from the data, identifying patterns and features without predefined output variables. The algorithm to use depends on the specific problem being addressed.

These algorithms learn independently and make predictions without external guidance. An illustrative application of unsupervised learning is fraud detection. In this scenario, banks and financial institutions can employ unsupervised learning to identify irregular spending patterns and transactions that may indicate fraudulent or malicious activities.

2.2.3 Reinforcement Learning

Reinforcement learning is a type of machine learning where a model, called an agent, learns by interacting with its environment and getting feedback from its own actions. The goal is for the agent to learn the best actions to take in the environment to maximize rewards. Unlike other types of learning, there is no supervisor; the agent explores on its own to figure out the best actions that lead to the most rewards. Reinforcement learning is good for handling dynamic and complex situations, making it useful for things

Figure 7: Reinforcement Learning Workflow (Mahesh, 2020).



like controlling robots and playing games. However, it requires a lot of training, can struggle with complex situations, and needs careful planning for rewards and exploration (Talaei Khoei and Kaabouch, 2023).

The workflow of this type of learning is represented in Figure 7.

It is anticipated that reinforcement learning will become more widely used in various applications, such as robotics and self-driving technology, in the foreseeable future.

An example of reinforcement learning is to train a machine that can identify the shape of an object given a list of different objects such as square, triangle, rectangle or a circle. Some of the important reinforcement learning algorithms are Q -learning, Monte Carlo, Sarsa and DeepQ network. reinforcement learning methods do not need any supervision to train machine learning model

2.2.4 Machine Learning Algorithms

Linear Regression

Linear Regression (LR) is the simplest and most common statistical model when the goal is to measure the relationship between continuous variables.

In **Linear Regression**, there are input variables, represented by the variable x, and an output variable, represented by y. When x is a single variable, the model is called simple linear regression, and when there are multiple inputs, it is referred to as multiple linear regression (Ray, 2019).

Decision Tree

Decision Tree (DT) are a supervised machine learning prediction algorithm that can be used for classification or regression. They represent a hierarchical model of possible decisions and their respective consequences. This model is characterized by nodes and branches in each tree. Each node represents a feature of a category to be classified, and each subset defines a value that the node can take. Thus, observations about an item can be converted into conclusions through the analysis of the **DT**. Taking this into account, the simplicity of analyzing a **DT**, as well as its accuracy, have been factors justifying the widespread use and extension of this model (Mahesh, 2020; Ray, 2019).

Random Forest

The Random Forest algorithm is composed of a set of Decision Trees, each trained on a random subset of the training set. This way, the model makes predictions based on multiple decision trees, enhancing classification performance compared to individual decision trees. The aim is to combine predictions from multiple decision trees to obtain the average of the final prediction results. It can be used both for classification and regression problems (Das et al., 2015; Ray, 2019).

Support Vector Machine

The **Support Vector Machine (SVM)** operates with numeric input and binary output, aiming to find a linear plane with the maximum margin to effectively separate two output classes. When dealing with categorical input, it can be converted into numeric input, and categorical output can be represented as multiple binary outputs. An advantageous feature of **SVM** is its ability to handle a large number of dimensions. Additionally, the use of a kernel function enables **SVM** to address non-linear relationships (Das et al., 2015; Ray, 2019).

Boosting

Boosting reduces bias by overcoming underfitting issues. It is an iterative technique that adjusts the weight of an observation based on the last assigned classification. Therefore, if an observation was misclassified, it is assigned a higher weight, and vice versa. However, it can lead to overfitting problems (Mahesh, 2020; Ray, 2019).

- LightGBM: (Light Gradient Boosting Model) is an algorithm based on decision trees that applies gradient boosting. Its advantages over similar models include speed and efficiency in the training process, minimal memory usage, support for big data, and good results.
- XGBoost: (EXtreme Gradient Boosting) is also an algorithm based on decision trees where gradient boosting is applied. Unlike the previous one, it is not as fast and memory-efficient. However, it manages to be more consistent in terms of results.
- AdaBoost: (Adaptive Boosting) is an algorithm where boosting is applied using decision trees.

2.2.5 Data Mining

A subfield of **ML**, **Data Mining (DM)** can be quickly defined as "mining information and knowledge from large databases or information repositories", according to Han (1996). As an illustration, **DM** applications can help e-commerce systems analyse customer behaviours and, as a result, increase sales and improve their sales strategy.

Though evident that data mining techniques in education haven't advanced to the extent seen in the business sector, for instance, it has gathered the interest of the scientific community in the past few years. So, a new term was born. **Educational Data Mining (EDM)**. The fundamental aim of (**EDM**) is to extract valuable insights from educational data, including student records. This acquired knowledge holds the potential to improve the teaching and learning processes within the educational system (Algarni, 2016). A good example of this would be to use **ML** and consequently **EDM** to identify students at-risk, improve graduate rates and improve the schooling experience as is the main goal of the project this dissertation is inserted in.

Dealing with huge amounts of data can be hard, even with the help of machine learning and data mining.

2.2.6 Deep Learning

Briefly, despite being a branch of **Machine Learning**, **Deep Learning (DL)** differs from other areas of Machine Learning in the data it uses and how it leverages that data for learning. Unlike classification algorithms that use structured and labeled data for learning, **Deep Learning** utilizes unstructured and unlabeled data in its training. Learning is achieved through multi-layered neural networks that, as they iterate over the training data, become more effective in the classification task at hand.

These networks are designed to emulate how scientists believe the human brain operates during learning. The algorithm processes and reprocesses data, gradually refining the analysis and results to effectively classify object classes. Neural network layers consist of interconnected nodes, each progressively using a more complex algorithm to extract and identify attributes and patterns in the data, ultimately calculating the algorithm's confidence in the assigned classification.

Among the set of **Deep Learning** algorithms are convolutional neural networks, which are used for image classification in computer vision, and recurrent neural networks, which are employed in models where attributes and patterns change over time.

An example of **Deep Learning** application is Google's Deep Dream software, which possesses the capability not only to categorize images but also to produce unique and artificial artworks derived from its own knowledge Shinde and Shah (2018).

2.3 Natural Language Processing

Though it is quite hard to achieve a single definition of **Natural Language Processing (NLP)**, a vast majority of researches define it as consisting of creating computational models for tasks that rely on information expressed in natural language, such as translating and interpreting text, and facilitating human-machine interaction (Norvig and Intelligence, 2002).

On the whole, **NLP** draws from various disciplines, including computer science, linguistics, logic, and psychology (Joshi, 1977). While there are a few challenges due to the complex nature of human language, recent progress in deep learning has demonstrated potential in overcoming these obstacles (Rebala et al., 2019)

The current trend in **NLP** research focuses on unsupervised or semi-supervised learning algorithms, capable of learning from unannotated data or a mixture of annotated and non-annotated data. Before the 1980s, **NLP** systems relied on handcrafted, rule-based approaches. However, a revolution occurred in the 1980s with the introduction of machine learning techniques based on Chomskyan linguistic theories and Moore's law. Decision trees and statistics-based techniques replaced handcrafted rules, particularly in machine translation, where complex statistical models yielded success. Despite earlier reliance on large textual corpora, recent research aims to effectively learn from smaller datasets (Gupta, 2014).

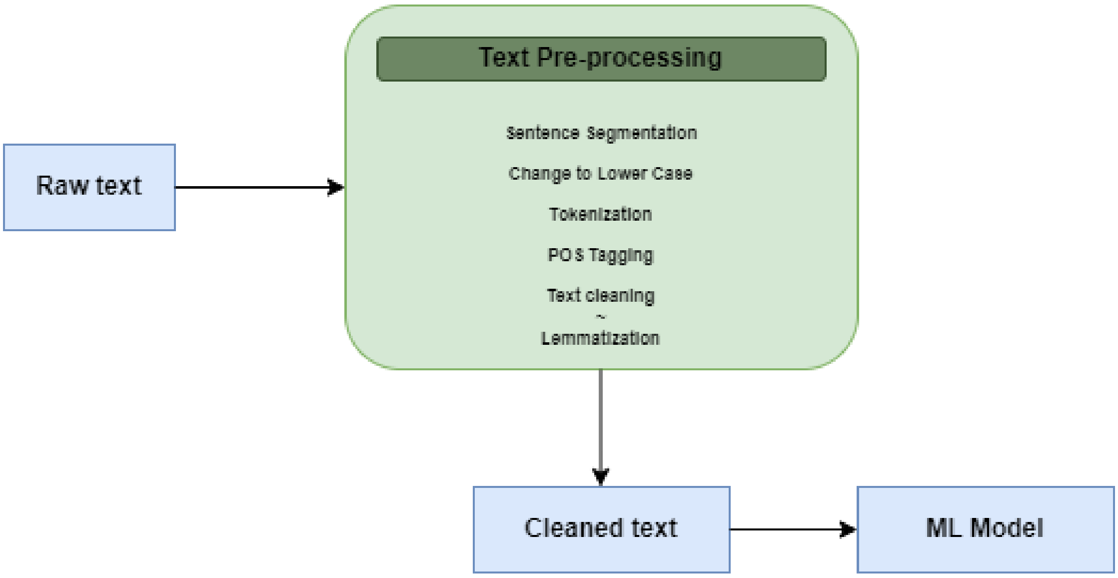
A few **NLP** applications include:

- Machine Translation: Machine Translation automatically translates text between human languages. This **AI**-complete task demands a deep understanding of semantics, grammar, and real-world concepts.

- Analysis of Discourse: Discourse analysis determines the structure of connected text, identifying relationships like contrast and explanation. It also categorizes speech acts, including content questions, yes/no questions, assertions, and statements.

- Morphological Splitting: Morphological splitting separates terms into morphemes, recognizing corresponding categories. The challenge lies in complex term structures, especially in languages like Turkish with multiple forms for each dictionary entry.

Figure 8: **NLP** pipeline.



- Named Entity Recognition (NER): Identifying terms labeled as named entities, like places or people names, presents challenges, including determining entity types and overcoming inconsistent capitalization across languages.
- Part-of-Speech: Marking part of speech involves identifying grammatical categories for each term. Ambiguity in languages, where common terms may have multiple parts of speech, adds complexity, especially in languages like English with minimal inflectional morphology.
- Text Parsing: Parsing grammatically analyzes sentences, creating parse trees. Ambiguity in natural language grammars leads to numerous possible parses for a single line, many lacking meaningful interpretations.
- Speech Recognition: Speech recognition identifies spoken words by listening to sound clips, a challenging task, particularly in languages with coarticulation where consecutive word sounds mix.
- Sentiment Analysis: Sentiment analysis extracts subjective information from text, used to determine public sentiments about a varitety of subjects.

A basic **NLP** pipeline can be seen in Figure 8.

The pre-processing method holds significant importance in text mining techniques and applications, as the characters, words, and sentences identified during this phase serve as the foundational units for all subsequent processing stages. This is essential because text data often contains special formats like number and date formats, and common words such as prepositions, articles, and pronouns, which are unlikely to contribute to text mining, can be removed.

Sentence segmentation, also known as sentence boundary detection, involves dividing a text document or corpus into individual sentences. This process is crucial for identifying word boundaries, enabling subsequent processing on each sentence.

What is more, raw text often includes abbreviations and words in all capital letters. While this step is commonly overlooked, it is a straightforward and effective text pre-processing measure, particularly in scenarios with a notably scattered dataset. It has been observed that variations in capitalization, such as 'Portugal' versus 'portugal' show distinct results. This implies that when a word is written in uppercase or lowercase, the computer treats them as two different words, resulting in separate word vectors in subsequent stages, like word embeddings. Consequently, a widely adopted best practice in text pre-processing is to convert all words to lowercase. This, however, might complicate named entity recognition or sentiment analysis, for instance.

Tokenization consists of breaking down words and symbols in a sentence into smaller parts called tokens. The idea is to look at the words in a sentence one by one. The resulting list of tokens becomes input for subsequent processing, such as parsing or text mining. This is important in both language studies and computer science (Tabassum and Patil, 2020).

The main purpose of tokenization is to make sure that the words in documents are treated consistently.

POS tagging involves labeling each word in a sentence with its suitable part of speech, such as nouns, verbs, adverbs, adjectives, pronouns, conjunctions, and their sub-categories (Tabassum and Patil, 2020)

In the text cleaning stage certain words and elements are removed from a collection of text data to improve the efficiency of a machine-learning model. The process involves deleting numbers, capitalization, punctuation, and stopwords from the text data. Regular expressions are used to carry out the text cleaning process.

Stop words are extremely common words like 'and', 'are', 'this', that are not helpful in classifying documents. Therefore, they should be eliminated. This procedure trims down the text data and enhances system performance. Every text document contains these unnecessary words, which are not essential for text mining applications, decreasing both data size and model training time (Vijayarani et al., 2015).

Lemmatization involves either removing or substituting the suffix of a word to derive its base form, known as the lemma. Unlike a stemmed word, the lemma always retains its meaningfulness. For instance, the lemma of the word 'Caring' is 'Care,' which remains a meaningful term (Tabassum and Patil, 2020).

Overall, the aforementioned steps are some of the basic tasks of text processing through **Natural**

Language Processing. It is important to note that the sequence in which these steps are executed holds significance. For instance, if stopwords like "is", "the", and "in" are eliminated before Part-of-Speech tagging, the POS tagger might inaccurately label the sentence, resulting in erroneous outcomes (Tabassum and Patil, 2020).

2.3.1 Sentiment Analysis

Sentiment Analysis (SA) constitutes a domain within text mining research dedicated to extracting and categorizing sentiments, emotions, and attitudes conveyed in textual data (Medhat et al., 2014), as well as summarizing opinions, analyzing reviews and detecting sarcasm.

It can help us interpret emotions in unstructured texts as positive, negative, or neutral, and even calculate how strong or weak the emotions are.

Employing techniques such as **Natural Language Processing (NLP)** and **Machine Learning (ML)** facilitates the identification and examination of emotions conveyed through written or spoken language (Cui et al., 2023). This application extends across diverse fields, including healthcare (Milne-Ives et al., 2020), analysis of customer feedback (Iqbal et al., 2022), social media (Bibi et al., 2022), and education (Dake and Gyimah, 2023).

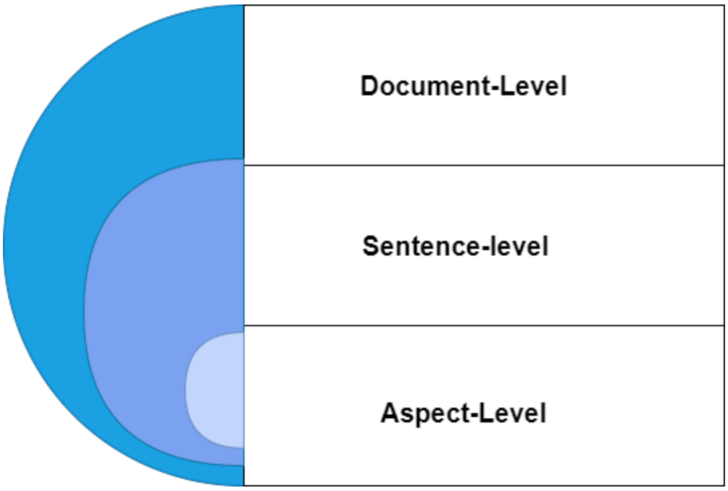
While other terms such as opinion mining, opinion extraction, subjectivity analysis and polarity analysis are used in the literature as synonyms for the , some authors differentiate them in the following way: pinion analysis extracts and analyzes people's opinions about an entity, while sentiment analysis identifies the sentiment expressed in a text and then analyzes it.

Thus, the goal of sentiment analysis is to identify opinions, the feelings they express, and ultimately define their polarity. Despite this distinction by some authors, both terms, along with those mentioned earlier, are universally categorized under the general term **Sentiment Analysis**.

Overall, given the textual nature of sentiment extraction and analysis sources, the processes involved are intrinsically related to **Natural Language Processing** processes. Sentiment analysis is indeed a computational process belonging to the NLP field. As such, **Sentiment Analysis (SA)** has the capability to address various challenges and aspects that are characteristic of NLP problems. Some of these challenges include dealing with negations, word sense disambiguation, and coreference resolution.

Despite the overlap, **Sentiment Analysis** has a more restricted focus compared to broader **NLP** tasks. **Sentiment Analysis** doesn't delve into fully analyzing and understanding the complete semantics of each sentence. Instead, it concentrates on specific aspects like determining the polarity (positive, negative, neutral) of associated sentiments.

Figure 9: Types of **Sentiment Analysis**.



In summary, sentiment analysis is a specialized application within the larger **NLP** field, addressing specific challenges related to understanding sentiments in textual data without needing to comprehensively analyze the entire linguistic structure of sentences.

There are three main levels in **Sentiment Analysis (SA)**) (Medhat et al., 2014), (Wankhade et al., 2022): document-level, sentence-level, and aspect-level, as represented in Figure 9.

- Document-level: This level looks at an entire opinion document and decides if it expresses a positive or negative sentiment about a specific topic.

- Sentence-level: Here, the focus is on each sentence in a document. It aims to classify if a sentence is talking about opinions (subjective) or just stating facts (objective), so that, IF it Is expressing opinions, it determines whether those opinions are positive or negative.

- Aspect-level: This level is more detailed. It aims to classify sentiment based on specific aspects of entities (like products or services). First, these entities and their aspects are identified. People might have different opinions about different aspects of the same thing. An example of aspect-level **SA** is reviews of a specific product, in which the user might have an overall positive review but specific opinions about different aspects of the product. For instance, one might give an overall positive review of a restaurant but highlight the poor service from the staff. In this sense, aspect-level **SA** is fine-grained process of assessing the sentiment of a text concerning a particular aspect.

2.3.2 Affective Computing

Affective computing is a developing area of study that seeks to empower intelligent systems to recognize, sense, infer, and interpret human emotions.

This interdisciplinary field spans disciplines ranging from computer science to psychology, and from social science to cognitive science. While sentiment analysis and emotion recognition are distinct research areas, they are interconnected within the realm of Affective Computing research (Poria et al., 2017).

Emotion recognition seeks to identify the emotional state of individuals, focusing on discrete emotions or dimensional emotions. It primarily emphasizes visual emotion recognition, audio/speech emotion recognition, and physiological emotion recognition. On the other hand, sentiment analysis predominantly deals with textual assessments and opinion mining, particularly in the context of social events, marketing campaigns, and product preferences, and its outcomes typically categorize sentiments as positive, negative, or neutral Wang et al. (2022).

Given that an individual in a joyful mood often exhibits a positive attitude toward their surroundings, there is an overlap between emotion recognition and sentiment analysis. For instance, there's a system created to understand and analyze what someone says or writes. This system uses a special method called "context-level inter-modal attention." Its job is to figure out if the message is positive or negative and to recognize the specific emotions being expressed, like anger, disgust, fear, happiness, sadness, or surprise. This is an example os **Sentiment Analysis** and Emotion Recognition overlapping.

Advancements in affective computing have led to the creation of publicly available benchmark databases, primarily comprising unimodal databases (textual, audio, visual, and physiological databases) and multimodal databases. The widespread use of these databases has spurred the development of affective computing approaches based on **ML** and **DL**.

It is important to note that detecting emotions from text poses significant challenges. The task is made difficult by the presence of numerous ambiguities, the introduction of new slang or terminologies regularly, and the evolving nature of language.

Chapter 3

Literature Review

This section aims to provide an overview of current literature in terms of school dropout, artificial intelligence in education, natural language processing and emotion detection.

This review followed the procedural guidelines outlined in Kitchenham's Systematic Literature Review (SLR) Methodology, as it is appropriate for and widely used in the field of software engineering (Keele et al., 2007).

Conducted in June 2023 using SCOPUS (www.scopus.com), Semantics Scholar, ACM Digital Library and Google Scholar, the literature review sought to identify studies at the intersection of emotion detection/affective computing, educational data mining, and academic achievements/student academic performance, using the query presented on Figure 10.

Figure 10: Query used for literature review.

```
( TITLE-ABS-KEY ( "emotion mining" ) OR TITLE-ABS-KEY ( "affective computing" ) OR TITLE-ABS-KEY ( "sentiment analysis" ) AND TITLE-ABS-KEY ( "school dropout" ) OR TITLE-ABS-KEY ( "academic performance" ) )
```

In screening the found articles, specific exclusion criteria were established. Therefore, documents were excluded or included based on whether they met the following criteria:

AC1 Paper is not written within the established time frame but is considered relevant due to its findings.

EC1 Papers are not open-access.

EC2 Papers are not written in English or Portuguese.

EC3 Papers are not from the last 5 years.

EC4 Papers are not from the field of computer science.

After this extensive screening through different repositories and databases, 732 articles were found and, after the inclusion and exclusion criteria were applied to this process, a total of 19 articles were selected and are detailed in Table 1. The filtering procedure can be seen in Figure 11.

Figure 11: Flow diagram of literature review screening.

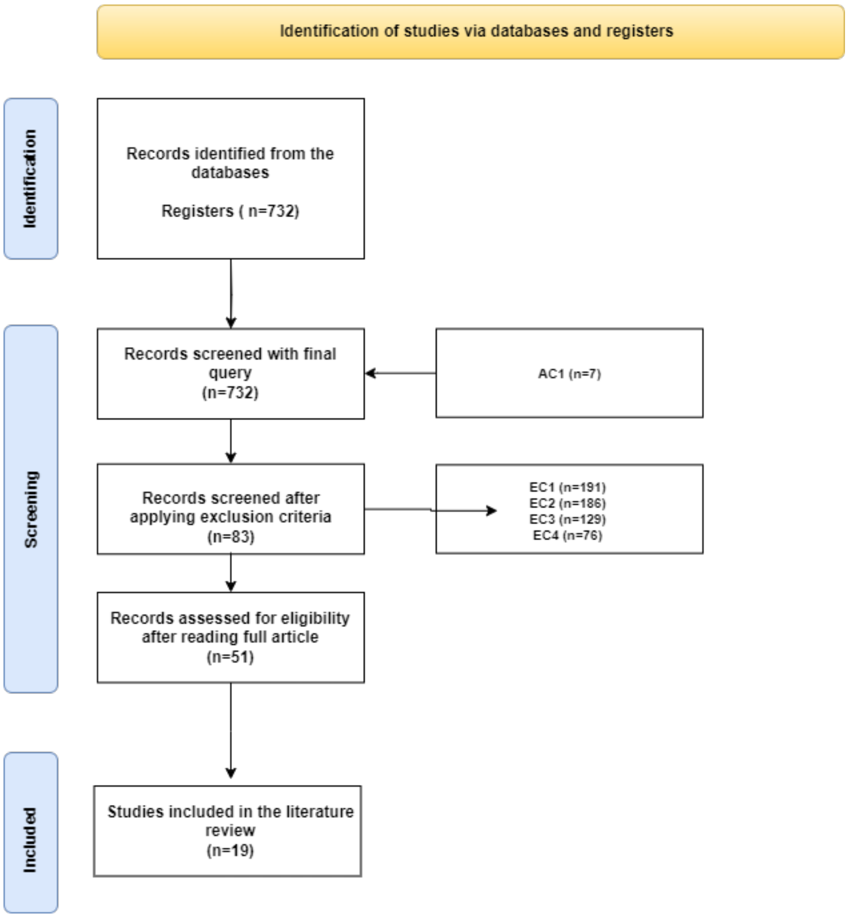


Table 1: Systematic Literature Review.

Paper	Target	Methodology/ Key Findings
Gorde et al. (2023)	Online Learning	The paper introduces a robust, real-time system for face detection and emotion recognition using machine learning and the OpenCV library. It emphasizes the use of deep learning (DL), particularly convolutional neural networks (CNN), for emotion detection, which has shown superior performance over traditional image processing methods. The system is designed to classify emotions like happiness, sadness, anger, surprise, etc., by analyzing facial expressions, thus enhancing comprehension of human interactions for various applications

N et al. (2023)	Education	The study utilized convolutional neural networks (CNN) to identify human emotions from facial expressions.
Gupta et al. (2023)	Online Learning	This paper used ensemble Convolutional Neural Networks to detect cognitive states, focusing on differentiating between attentive and inattentive states.
Bóbó et al. (2022)	Online Learning	This paper presents the SASys architecture, which employs sentiment analysis to predict student dropouts in e-learning. By analyzing text from virtual learning environments and incorporating student engagement data, SASys identifies emotional states indicative of dropout risks. The study shows that sentiment analysis can effectively gauge student satisfaction and engagement, assisting educators in early intervention to reduce dropout rates.
Hu et al. (2022)	Higher Education	The study introduces an ensemble method that extracts features of engagement semantics and sentiment and stress from an AdelaideX student dataset. This method is used to compare performance measures on how accurately these features can predict student grades. The baseline models used for validating the MOOC dataset are Naïve Bayes (NB), Random Forests (RF), and Deep Learning Artificial Neural Network (ANN). These models were implemented using Sklearn libraries and Keras
?	Higher Education	Based on the combined risk weights from the surveys and journal entries written by students, a system developed by the authors automatically flags students who may need additional support. This flagging can also be done manually by instructors or the students themselves.
Aljarrah et al. (2021)	Online Learning	The paper highlights the use of algorithms like deep learning, artificial neural networks, natural language processing, and machine learning for analyzing textual emotions. The random forest algorithm, in particular, showed notable success in this area.

Li and Xu (2014)	Higher Education	Improved BP Neural Network Based on Bayesian: The methodology involves an improved BP (Back Propagation) Neural Network optimized with Bayesian methods for mental health assessment.
Kaur et al. (2021)	Online Learning	This paper present a pictorial method with robotic characters, and students marked their mood at the beginning, middle, and end of each test session. Findings show that positive mood correlated with better performance compared to negative mood.
Baragash et al. (2022)	Higher Education	Sentiment analysis was performed using the machine learning approach with Rapid Miner software and Support Vector Machine (SVM) classifier. The analysis classified students' opinions into positive and negative classes based on sentiment and National Research Council (NRC) word-emotion lexicon.
Wang and Shi (2021)	Online Learning	Proposal of a context-aware network model based on transfer learning to predict learner performance and improve the educational process.
Bouhlal et al. (2020)	Online Learning	The paper describes a three-phase system for emotion recognition – detection, extraction, and classification of facial expressions. Eigenfaces, Fisherfaces, and Support Vector Machine (SVM) are highlighted.
Muñoz et al. (2020)	Online Learning	The paper focuses on the design of an emotion-aware learning analytics architecture and its integration into a semantic task automation platform.
Taurah et al. (2020)	Online Learning	Neural networks are used for training datasets to classify learning intervals based on the learner's average emotion and performance during specific time intervals
Xing et al. (2019)	Online Learning	The methodology includes a machine learning model to detect achievement emotions in forum posts and a survival analysis to assess their influence on student dropout.

Ranjan and Mishra (2022)	Higher Education	This study used Machine Learning models to perform sentiment analysis on Google app reviews, focusing on university students' behavior towards the app market. Among the classifiers used, the Support Vector Machine (SVM) showed best performance.
Rodriguez et al. (2012)	E-Learning	The study employs emotion extraction techniques on student essays to identify emotions like joy, anger, fear, and sadness. It explores the use of emotion dictionaries and word spotting techniques for this purpose.
Pekrun et al. (2011)	Higher Education	The objective of this paper is to develop and validate the Achievement Emotions Questionnaire (AEQ), designed to assess emotions connected to learning and academic performance of university students.

3.1 School Failure and Dropout

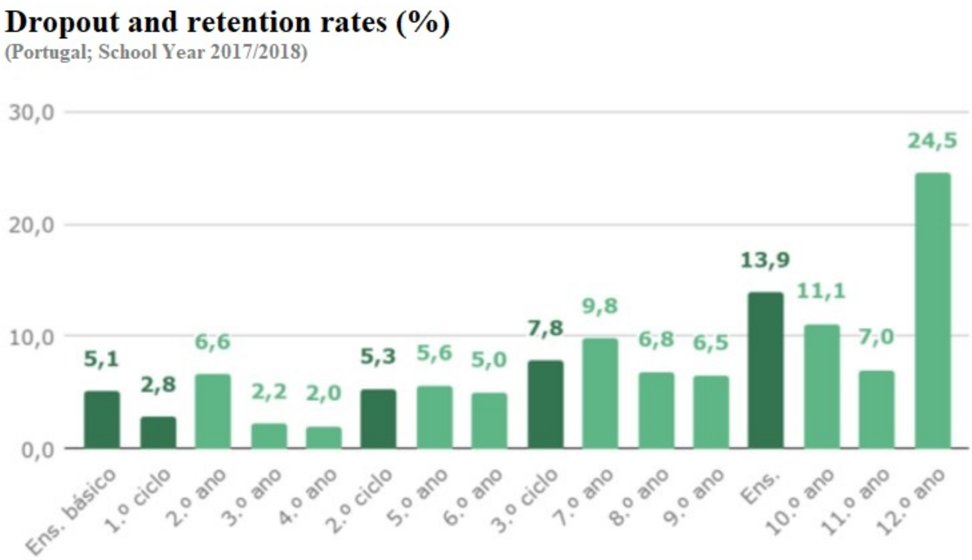
Despite the efforts made in terms of educational policies to create programs that promote success and, despite the downward trend observed in recent years, the fight against school dropout and failure remains an unavoidable challenge.

National Overview

According to the European Commission (European Commission, 2019), school dropout, which is in turn closely related to school failure and the low level of qualifications of the Portuguese population, has consequences at an individual and social level, and is associated with the following problems: unemployment, social exclusion, poverty and health problems. In fact, school failure is a factor that has been worked on by the Ministry of Education over the years.

However, and despite its downward trend, this problem has not yet been eliminated. Figure 12 shows the retention and dropout rates by level of education, cycle of studies and year of schooling in 2017/2018 (Direção Geral de Estatística da Educação e da Ciência, 2018). The values recorded for this indicator are increasing with the level of education and cycle of studies, and assume, in primary education, particular importance in the initial year of each cycle, in which students are subject to evaluation. The highest value is registered, however, in the final year of secondary education (24.5%).

Figure 12: Retention and dropout rates in the year 2017/2018 in Portugal (Direção Geral de Estatística da Educação e da Ciência, 2018).



Since Portugal joined the European Union, attempts have been made to fight school failure and dropout, through reforms and various measures implemented by successive governments, in an attempt to mitigate their dimension. According to the report issued in 2021 by the Conselho Nacional de Educação (2023) (in English, National Education Council) , the average dropout rate by young people aged 18-24 was 9.7%, with the European Union aiming to reach figures below 9% by 2030. Portugal reached a figure just above 5%, which shows a big drop compared to the information available in the 2018 report, which recorded a figure of 11.8%, while the average for the 28 countries of the European Union was 10.6%.

International Overview

At the European level, Croatia is the country with the lowest dropout rate, being only 2.4%, contrasting with Romania, which reaches 15.3%. According to data from the European Commission in 2016 there were still more than 4 million early school leavers across Europe. Globally, according to data from the UNESCO (2018)'s report, around 258 million children and young people in primary, primary and secondary education are not in school or have left school early. Studies show that attitudes towards school, substance use, emotional problems, parental involvement in school, and learning difficulties are the strongest predictors of truancy and school failure (Erdem and Kaya, 2020).

According to the study conducted by Miguel et al. (2012), one of the main challenges lies in early intervention, as it is urgent to prevent students from significantly breaking with the school system or dropping out, so that they can be identified and targeted for intervention.

In this sense, school mechanisms are needed to ensure a timely assessment and signaling of students, combined with methods capable of identifying students at risk accurately and in a timely manner. However, and despite all the research that has been done in this area, to date there is no instrument capable of performing this signaling.

3.2 Artificial Intelligence in Education

Some studies conducted mainly in the area of higher education indicate that the use of virtual assistants is becoming increasingly popular and that they can be an asset in supporting the student, contributing to the automation of tasks, offering support in terms of time management, access to information and facilitating communication. However, this technology is still only at an early stage, therefore needing some improvements and creation of new tools, in order to have more impact on student motivation and interaction (Gubareva and Lopes, 2020).

Some researchers claim that a few students see virtual assistants as useless or confusing, but they record significant improvements and even a change in perception after constant use and experience using them. These improvements even include some motivation on the part of students, who expected to get better grades after using a Virtual Assistant (Dorfman et al., 2001).

Mendoza et al. (2022) also record positive results from the use of chatbots by teachers and students at an elementary school in Mexico, who found the tool friendly and practical. Confirming that the use of virtual assistants can be a useful tool in the teaching and learning process, they add that information related to violence or depression should be reached and handled only by psychologists. This raises a relevant question: what role could digital assistants take regarding predictive and behavioral fields?

In the national market, only a single platform stands out, utilizing predictive systems and artificial intelligence – the SAPIE-EB platform. SAPIE-EB serves as a tool employing AI techniques to address academic failure and proactively identify risks associated with early school dropout (Cordeiro et al., 2020). Nonetheless, this forecast relies solely on indicators that, by themselves, may not be entirely reliable in signaling school failure. Moreover, some of the utilized measures might not be the most suitable for the targeted age group of the platform discussed in this paper. These indicators include attendance, behavior, academic performance, household income, and the educational qualifications of the mother. What is more, this platform does not include any data related to emotional states of students.

Globally, a few platforms utilizing artificial intelligence to predict instances of school failure and dropout.

However, WatsomApp, for instance, exclusively focuses on detecting bullying as a risk factor (WatsomApp, 2019). Notably, it operates through a game-like interaction with students, posing questions to identify bullying, and subsequently, the results are communicated to psychologists for intervention. However, drawbacks include its narrow focus on only one risk factor (bullying), limited interaction with psychologists exclusively, excluding other stakeholders, and the absence of predictive mechanisms for issuing recommendations.

Similarly, SAMEBullying serves as a solution exclusively for detecting cyberbullying and does not address other risk factors ENCAMINA (2019). It operates as a platform for teachers to monitor student conversations, raising concerns about data privacy. The platform's analysis is confined to identifying bullying in student conversations, presenting limitations in scope.

3.3 Emotion Detection

Within natural language processing, affective computing emerges as a distinct area, allowing the discernment of emotions based on data obtained from the target user (Li and Xu, 2014). Existing literature presents various models for emotion classification, encompassing text, audio, image, or video formats (Jayalekshmi and Mathew, 2017). While emotion lexicons, such as the NRC Emotion Lexicon and LIWC, are available, either manually annotated or automatically generated through machine learning algorithms (e.g., WordNet Affect and SentiWordNet), these were deemed inadequate for this project's context. The project necessitated spontaneous and authentic text to yield accurate results, given its objective of assessing the emotions expressed by questionnaire respondents in response to questions posed by others.

One of the main factors affecting a student's level of attention and performance is undoubtedly their emotional state. Emotion plays an extremely important role in the decision and knowledge acquisition process, and therefore influences perception, the learning process, the way people communicate, and the way decisions are made. There are already some models developed to understand how emotions influence the learning process such as Russell's Circumplex Model (Russell, 1980) and Kort's Spiral Model of Learning (Kort et al., 2001). The former is used to describe the student's emotional space, the latter to explore affective evolution during the learning process.Models such as the ones mentioned above can be studied and improved in order to analyze the evolution of the student along their school path and make predictive analyses about it.

In Bóbó et al. (2022), a system called SASys was introduced, designed for e-Learning, to recognize students' emotions using Sentiment Analysis. The system was tested in a real online learning platform, where it successfully identified students' motivation levels and analyzed sentiments in forum comments. SASys recommended motivational messages for students at risk of dropping out, and student feedback indicated that the system could help reduce dropout rates in distance education. The sentiment analysis, specifically using the FrameNet approach, showed promising results with 65% accuracy and 68% recall, aligning with expectations for this type of analysis. FrameNet, chosen for its linguistic foundation, demonstrated effectiveness in the case study with 88% accuracy, 75% precision, 84% recall, and a 78% F1-score. These results affirm FrameNet as a suitable lexical solution. In the case study, SASys successfully defined students' Emotional State by combining data from the Virtual Learning Environment with sentiment analysis results. This emotional state information was then used to recommend motivational messages, showcasing the architecture's effective performance.

The descriptive findings in Zablocki (2009) indicated noteworthy distinctions between individuals who dropped out and those who did not, particularly in terms of disability classification, race/ethnicity, gender, disciplinary school exclusion, grade retention, academic grades, and levels of emotional engagement. While showing a relatively minor impact, elevated levels of emotional engagement were linked to reduced chances of dropout. Suggestions in this study also emphasize the need for cautious consideration of school practices like disciplinary removal and grade retention. Additionally, it is recommended to implement school programs aimed at supporting at-risk youth to prevent dropout.

Using a Twelve Emotions in Academia Model as base, Ruiz et al. (2018) developed the Emotions-Module within the PresenceClick system, providing students with the ability to recognize and track their emotions through visualizations. Instructors can access an anonymized version of this data to enhance their understanding of the group's emotional state and propose effective learning strategies for overall improvement. Considering the positive correlation between attendance in face-to-face sessions and student progress, attendance information has also been incorporated. Over a span of two academic years, an experiment was conducted within a single subject in the Computer Science degree, collecting emotional data and attendance through PresenceClick. Subsequently, a correlation study and principal component analysis were performed on the compiled data, demonstrating data consistency and enabling the feeding of prediction models each academic year. With confirmed correctness and stability of data, data mining techniques were employed to generate two models based on probability tables and decision trees. These models empower both instructors and students to identify issues early on and prevent academic challenges. This research has shown proven positive emotional state of students to be essential for favoring student learning.

Arroyo et al. (2009) discusses the utilization of sensors in intelligent tutors for detecting students' emotional states and integrating emotional support. In two classroom experiments employing four sensors, the tutor dynamically gathered data streams of physiological activity and students' self-reported emotions. The findings suggest that fluctuating student emotions, observed on a state-based level, are linked to broader, more enduring affective variables such as self-concept in mathematics. Students generated self-reports of emotions, and models were developed to automatically deduce these emotions from physiological data obtained from the sensors. Summaries of students' physiological activity, particularly data streams from facial detection software, proved effective in predicting over 60% of the variance in students' emotional states. This predictive accuracy surpassed that of emotions predicted using other contextual variables from the tutor when these sensors were not present. The research also indicates that by adjusting the "context" of the tutoring system, we may enhance students' emotion reports and, subsequently, improve attitudes toward mathematics.

Bustos-López et al. (2021) states that understanding the emotions of students enables educators to effectively adjust or guide educational resources, activities, learning environments, and procedures within a specific educational community, where factors such as age, learning styles, and skills pose challenges.

These findings suggest that emotion detection can be a valuable tool in identifying and addressing potential school failure.

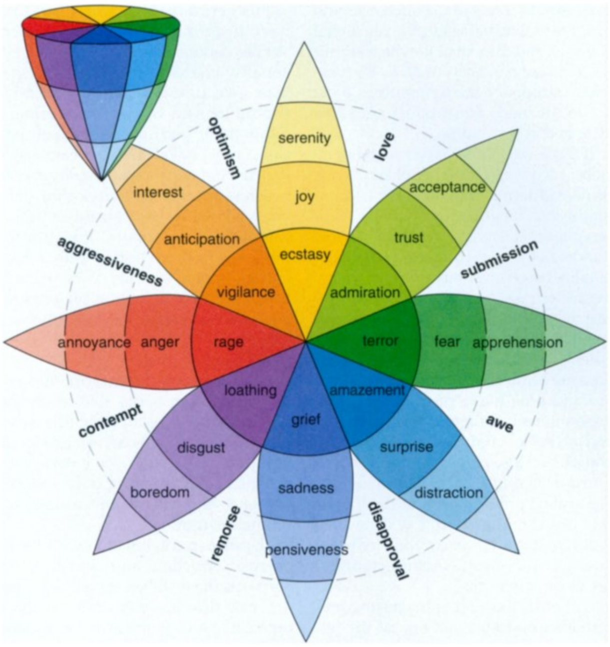
Additionally, Plutchik's dimensional model is one of the most commonly used models for emotion detection practises (Plutchik, 2001). Even though the theme of emotion detection and classification has been widely discussed by researchers in recent years, psychologists have not yet fully agreed on a single definition of emotion, but there's consensus that it is a complex concept. Plutchik defines emotions as

> *a multifaceted and inferred response to a stimulus, including subjective feelings, cognitions, impulses to actions and behaviour* (Plutchik, 2001,p.551)

His three-dimensional circumplex model, which depicts the relationships between emotions using a color wheel analogy, is one of the most used models for emotion detection and analysis. The model includes a vertical dimension for intensity, a circular dimension for similarity, and eight sectors representing primary emotions. The empty spaces within the model represent dyads, which are combinations of two primary emotions Plutchik (2001). The eight primary emotions in the model can be organized into four pairs of opposites: joy and sadness, fear and anger, anticipation and surprise, disgust and trust. This can be seen in Figure 13.

Plutchik claims that emotions can combined in the same way that blue and red make purple, and, for instance, mixing joy and anticipation produces optimism. The vertical dimension of the model allows

Figure 13: Plutchik's circumplex model of emotions (Plutchik, 2001,p.349) .



for intensity, meaning that numerous other combinations can be made, which can convolute the emotion classification process.

It is in Plutchik's dimensional model that this dissertation's emotion detection approach is based on.

3.4 Overview

This overview of current literature showed that there is a variety of studies on the application of **Artificial Intelligence** and Emotion Detection in the field of education, including the topic of school failure prevention. However, the vast majority of studies focus on online learning environments or higher education, leaving out the importance of developing tools for this signaling of students at age of mandatory education

A few studies included in this research have shown that emotion detection is linked to student performance, engagement, and retention, suggesting its pivotal role in educational success. Challenges remain in accurately detecting and interpreting emotions. The advancement in artificial intelligence, especially in natural language processing and machine learning, is central to this.

Chapter 4

Methods and Methodologies

4.1 Dataset

The dataset used was retrieved from a private school in northern Portugal, at the beginning of the third term during the 2021-2022 school year.

The sample encompassed 845 students, predominantly male (n = 430, 50.9%), ranging in age from 10 to 17 (M = 12.25, SD = 1.541), with approximately half belonging to the 2nd Cycle (n = 422, 49.9%).

It comprised 845 entries and 123 columns.

4.1.1 Data understanding

The original dataset contained 123 columns, which represented variables such as:

- Student ID;
- Authorization for project participation;
- Gender;
- Age;
- School Year;
- If the student has failed or not;
- If the student attends tutoring sessions and if so, what type of tutoring and how often;
- The students' parents' education;
- Parents' marital status;

- Student education goal;
- Student grade satisfaction;
- Student failure risk;
- Student answers questionnaire of 50 statements, scripted by a team of educational psychologists.

Distribution of students per school year

Figure 14: Distribution of students per school year.

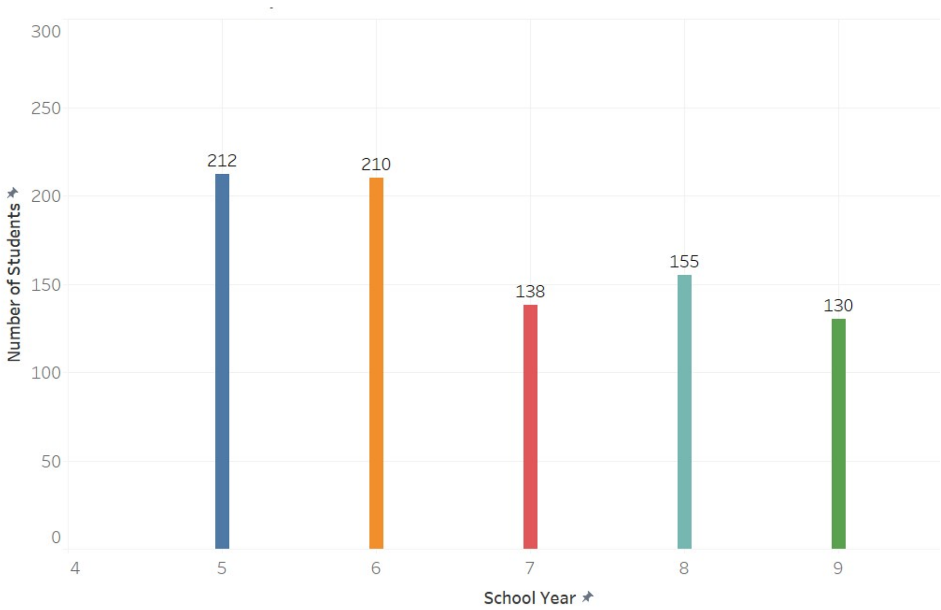
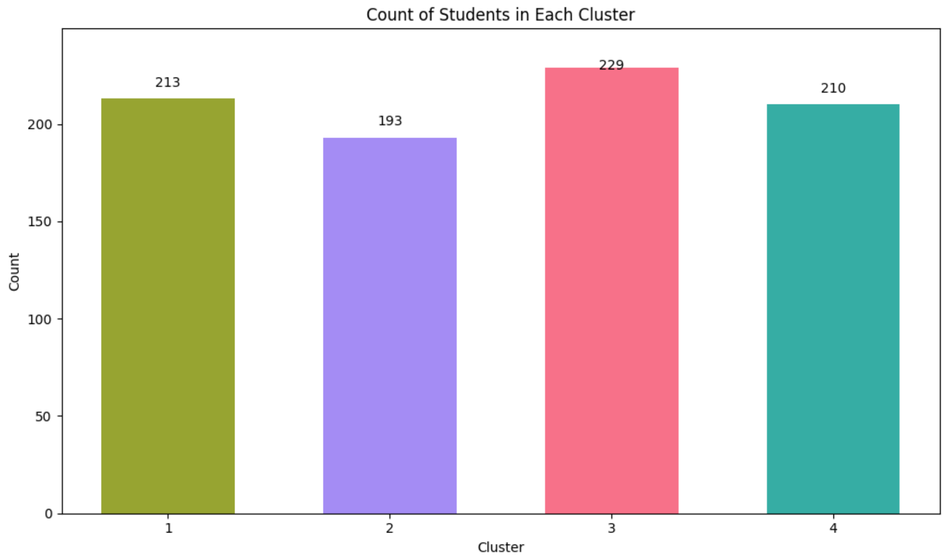


Figure 14 shows the distribution of the number of students per school year. This distribution is moderate, with grades 5 and 6 having a slightly higher number of students than grades 7, 8 and 9. However, this different is not significantly different.

Clustering of risk of failure

In an initial analysis, the program defined clusters autonomously, resulting in 3 clusters. The distribution of samples across these clusters was considered adequate, and the quality of the cluster model was assessed as average. The most influential predictors in defining the clusters were, in decreasing order of importance: the Maths grade, followed by the student's Portuguese grade in the previous term, the student's satisfaction with their grades, the student's involvement in school and the perception of whether their grades reflect their ability.

Figure 15: Distribution of risk clusters in the Pilot Study.



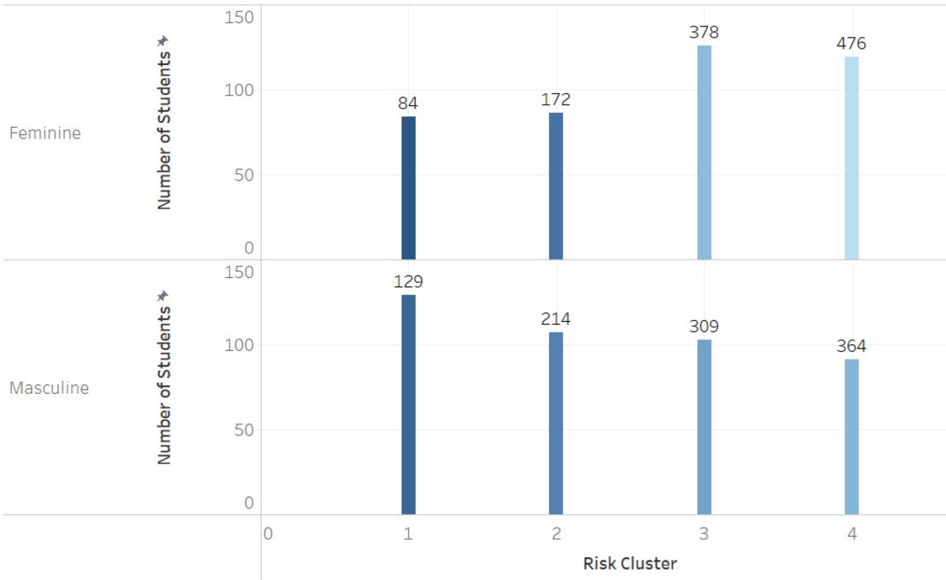
In a second analysis, the Educational Psychology researchers specifically defined 4 clusters to ensure theoretical and practical consistency in the interpretation of the clusters identified. In this analysis, a maximum limit of 4 was set by the researchers, seeking theoretical and practical consistency in the interpretation of the identified clusters. The distribution of samples among the clusters was adequate and more balanced than in the previous analysis, resulting in an average quality of the cluster model. The most crucial predictors for defining the clusters, in descending order of importance, were the student's Portuguese grade in the previous term, the student's Maths grade in the previous term, the student's satisfaction with their grades, the student's involvement in school and, finally, the student's perception of whether their grades reflect their ability. This risk cluster distribution can be seen in Figure 15.

- Cluster 1. High risk of school failure.
- Cluster 2. Medium-high risk of school failure.
- Cluster 3. Medium-low risk of school failure.
- Cluster 4. Low risk of school failure.

This distribution can be visualized on Figure 15.

Figure 16 shows the distribution of the risk of failure per gender. It is possible to observe there are more male students at a high risk of failure than female students.

Figure 16: Distribution of gender per risk of failure.



4.1.2 Data Preparation

The data preparation involves transforming the datasets so that the information contained in them is appropriately exposed to **Machine Learning** and knowledge extraction tools. Therefore, the data must be formatted for the tools, as when collected from the real world, it may be incomplete and, above all, may have inconsistencies.

The original dataset contained a lot of variables that would not be necessary for the emotion classification process and therefore only the answers to the questionnaire and the student ID were extracted from the original dataset and a new one was created, containing, then, 845 entries and 51 columns, with the two variables aforementioned.

After the emotion classification and students' emotional scores had been calculated, these were added back to the first dataset in order to allow for data analysis, as detailed in the Results section of this dissertation.

In the end, the final dataset contained 845 entries and 19 columns, with the variables represented in Table 4.1.2.

4.2 Modelling

For training **Machine Learning** models, the dataset was divided into training (80%) and test (20%). Support Vector Machine, DecisionTree, RandomForest, and XGBoost algorithms were applied, as they

Table 2: Variables in the final dataset.

Student ID
Gender
Age
School Year
Fail
Mother's Education Level
Father's Education Level
Parents' Marital Status
Education Goal
Tutoring
Risk of Failure (automated)
Student Emotional Scores

have shown good results such as in Faria et al. (2022), Magalhães et al. (2023b), Bouhlal et al. (2020) and Baragash et al. (2022).

Default settings were predominantly used these your machine learning models. While this is a common practice, especially in exploratory analyses or when the default parameters are known to perform well generally, it's important to acknowledge that tuning these parameters could potentially improve model performance. As this presents an exploratory analysis, default settings were chosen.

4.3 Evaluation Metrics

Independent Samples T-Test

For the hypothesis test, the independent samples t-test (also known as the two-sample t-test) was performed. This test is used to compare the means of two independent groups to determine if there is a statistically significant difference between them. This is a common statistical method for such comparisons when the data is approximately normally distributed and the sample sizes and variances are similar.

Mean Absolute Error and Mean Squared Error

The chosen evaluation metrics are **Mean Absolute Error (MAE)** and **Mean Squared Error (MSE)**. **MAE** measures the average error magnitude in a set of predictions, while **MSE** computes their variance, indicating how closely a line aligns with a group of points. These evaluation metrics are commonly used in regression tasks.

Cross-Validation

Cross-validation is a technique that shows how the model generalizes, i.e., how it behaves with data it has never seen before. This means that a portion of the training data is removed and used to test different models. This method suffers from too much variance because it is not certain which data end up in the training set, and for different sets, the model's performance can be entirely different. In other words, there is a risk of losing important patterns and trends, which can introduce bias.

K-Fold: The data is divided into k subsets, and the validation method is repeated k times. In each iteration, one of the k subsets is used as the test set, and the remaining subsets are combined to form a training set. This significantly reduces bias because most of the data is used.

The number of fold to use depends on the size of the data. If the value of K is too large, it will result in reduced variance across the training set and constrain the model's ability to vary across iterations. What is more, the number of folds is inversely proportional to the size of the dataset. In other words, if the dataset size is too small, the number of folds may increase.

The commonly used and optimized value for K is 10, applied with a sufficiently large dataset. For this reason, for all **ML** models used, cross-validation with 10 folds was performed.

Machine Learning Implementation

In summary, for the implementation of the machine learning approach, Python (version 3.8) was employed as the preferred programming language. Additionally, the following libraries were utilized:

- Numpy and Pandas for data exploration and processing.
- Scikit-learn, XGBoost, Support Vector Machine, Decision Tree, Random Forest with and without cross-validation, for constructing, training, and fine-tuning models.
- Matplotlib for evaluating model performance.
- Seaborn, Scikit-posthocs and Scipy for statistic analysis.

Chapter 5

Emotion Detection

Emotions play a crucial role in the human experience, impacting behavior, cognitive functions, decision-making, resilience, well-being, and interpersonal communication. In our daily lives, we see emotions like anger, sadness, and happiness as a inner state or feeling. Experiencing emotions is personal and can be confusing, especially when we feel more than one at a time. Trying to study emotions objectively is complex because people often keep their thoughts and feelings to themselves, and it is important to be careful about trusting what others say about their emotions. It's a special challenge to scientifically study something as complex and hidden as emotions.

On their part, psychoanalysts have pointed out that emotions can be held back, suppressed, or unconscious, making them hard to reflect on. Additionally, language itself introduces ambiguity, adding confusion and making it difficult to express mixed feelings clearly. The meanings of emotion words are often unclear; for instance, people may struggle to differentiate between fear and anxiety, guilt and shame, or envy and jealousy. This leads us to use metaphors, like ”blowing off steam,” ”hating someone's guts,” ”pain in the neck,” ”lump in the throat,” and ”a broken heart,” to describe emotions (Plutchik, 2001).

So, how can we study and understand emotions? It's crucial to develop a theoretical approach because emotions are a fundamental part of who we are and how we cope.

In an educational setting, emotions have an impact on students' cognitive processes, performance, as well as their psychological and physical well-being. School is a major theme in both children's, teenagers' and young adults' lives and academic learning and success are hugely important in society because they affect education, careers, relationships, and resource distribution. Therefore, it's reasonable to think that students feel a variety of emotions in school. This makes learning and achievement major triggers of human emotions, covering personal, task-related, and social feelings. Emotions are likely to influence students' thinking, performance, and overall well-being, especially those connected to school like enjoying learning, feeling proud of success, or being anxious about tests.

Prior research has indicated a consistent connection between emotion and learning. It is recognized that emotions can significantly impact students' understanding and overall objectives. Therefore, the importance of emotions in learning should not be overlooked (Binali et al., 2009).

That being the case, affective computing can be useful in identifying student emotion, allowing for the creation or development of actions to both prevent school failure and dropout and improve the learning-teaching experience.

However, comprehending a student's emotional response in a complex learning setting poses a considerable challenge. The most common practises of emotion classification either in text, audio, image or video involve the utilization of emotion lexicons, both manually annotated or generated by **ML** algorithms, which are in both cases available mostly in the English language and require genuine and spontaneous text created by the user (Jayalekshmi and Mathew, 2017).

In the context of the project in which this dissertation is inserted, the students did not produce a valid sample of text to use existing emotion detection or even **SA** techniques. Thus, a new proposal for emotion recognition emerged.

5.1 Understanding Emotion Detection

The following emotions were identified within the scope of the project: happiness (also known as joy), trust, optimism, interest, boredom, anxiety, distraction, and shame. The interpretation of each emotion should be as follows, based on Plutchik (2001)'s Wheel of Emotions:

- HAPPINESS. Sense of energy and possibility, feeling excited and pleased. This emotion helps by sparking creativity, connection and giving energy.
- TRUST. It is the customary experience of warmth, acceptance, and a sense of security. It fosters interpersonal connections, encourages openness, and facilitates the establishment of alliances between individuals.
- OPTIMISM. The feeling of being energized, hopeful and looking forward. It gives the idea that the future is better than the present. It generates options and motivates action.
- INTEREST. Interest is a mild sense of curiosity, feeling open and looking. It helps individuals pay attention and explore.
- BOREDOM. This emotion is characterized by a sense of fatigue and low energy, accompanied by feelings of weariness and disinterest. It signifies the need to redirect attention towards aspects

that can be controlled within a given situation. Additionally, it indicates that the full potential of the situation is not being realized.

- ANXIETY. Anxiety is a blend of expectation and apprehension, implying that it involves sensations of heightened alertness, stress, and fear.

- DISTRACTION. This emotion implies feeling scattered, uncertain and unfocused. It is the feeling of not knowing that to prioritize.

- SHAME. Shame is characterized by a blend of fear and disgust, typically involves sensations of bitterness and stress.

5.2 Proposed Methodology

Once the students were not given the chance to and did not produce enough textual data to use already existing emotion detection methods, the proposed methodology emerged. The potential to extract emotional insights from a questionnaire filled out by the students surfaced and was subsequently tested.

It derived from Likert-scale questionnaires, which use statements where respondents express their opinions on a scale. This scale can measure agreement (from totally disagree to totally agree), satisfaction (from very dissatisfied to very satisfied), likelihood (from very unlikely to very likely), quality (from very bad to very good), or frequency (from never to always). Typically, this scale is five points, with 1 indicating the lowest score, 3 representing neutrality, and 5 representing the highest score. (Shaikh and Doudpotta, 2019).

These questionnaires help measure people's responses to statements, capturing personal opinions often linked to emotions. Based on Plutchik's idea that emotion is a reaction to a stimulus, like a question in a Likert-scale questionnaire, it is possible understand that when someone responds to a question, they are articulating an emotion influenced by the intensity of their answer. By linking the emotion connected to the question with the actual response, the emotion the respondent might be conveying can be suggested. The proposed methodology is then based on the hypothesis that it can be possible to extract emotional insights from a questionnaire not originally intended for such a purpose.

First, each sentence is individually examined, leading to a proposed emotional inference for which a score ranging from 0 to 1 will be manually designated. To ensure consistency, a major challenge in manual annotation (Kiritchenko et al., 2014), scores denoting the absence (0), very weak presence (0.25), mild presence (0.50), significant presence (0.75), and strong presence (1) of emotion, respectively, were used.

A dataset is constructed, featuring distinct columns for each identified emotion in sentences, accompanied by their respective scores. Subsequently, the emotional weight will be computed by analyzing the respondent's answers, as can be seen in Table 3.

Table 3: Demonstration of the proposed approach.

response	fully agree	partially agree	not sure	partially disagree	fully disagree
emotion weight	1.0	0.5	0.0	-0.5	-1.0

That is, if the student fully agrees with the statement presented, it means that they are articulating the scores assigned for each specific emotion. If the they student partially agrees, they are still articulating a bit of that emotion, just not to its full intensity. If they student replies with "not sure", no action is taken. If the student fully disagrees, they are not articulating that emotion. Fully disagreeing, however, still refers to an emotional articulation to the sentences, and therefore the total score of for each emotion annotated for that sentence is inverted. If the student partially disagrees, half of the score is inverted.

Ultimately, emotional scores for each respondent and each emotion are computed and expressed within a range spanning from -1 to 1. To normalize the final score, the respondent's score for each emotion is divided by the maximum value achievable for that specific emotion.

5.3 Annotation of the questionnaire

Manual annotation serves as a prevalent method in sentiment analysis (Kiritchenko et al., 2014). This can involve leveraging the "knowledge of the crowds," utilizing platforms like Mechanical Turk to gather evaluations from numerous individuals (Mohammad and Turney, 2013b). Alternatively, it may rely on specialist knowledge, where one to three experts annotate the dataset (Mohammad and Turney, 2013a).

The manual annotation of emotions was conducted by a team comprising researchers from the fields of linguistics and psychology. Each sentence in the questionnaire was assigned a score from 0 to 1, as explained in the preceding section, reflecting the emotional expression it conveys. The overall scores are then computed based on the respondent's answers.

Take, for example, the statement "I think this is going to be a good week." By referring to Plutchik's wheel of emotions (Plutchik, 2001), it becomes evident that this phrase aligns with the emotion of optimism. In simpler terms, this statement expresses an optimistic emotion. Optimism, as according to Plutchik's dimensional model (Plutchik, 2001), is the feeling of being energized, hopeful and looking forward. It gives the idea that the future is better than the present. Additionally, the statement "I think this is

Table 4: Example of manual annotation of the questionnaire.

Sentence	Happiness	Trust	Optimism	Interest	Boredom	Anxiety	Distraction	Shame
1- It is important for me to learn many things new things in this school year.	0,25	0,25	0,75	0,75	0,00	0,25	0,00	0,00
2- I often forget important deadlines.	0	0	0	0	0.25	0	0.75	0.25
3- What I learn in class is important for my future.	0,25	0,50	0,75	0,50	0,00	0,00	0,00	0,00
4- I participate actively in class.	0,25	0,25	0,25	0,75	0,25	0,25	0,00	0,00
5- My parents are present when I need them.	0,75	0,75	0,50	0,25	0,00	0,00	0,00	0,00

Table 5: Demonstration of the annotation process.

Emotion Weight	Quantity of Answers
0	0-7
0.25	8-14
0.50	15-21
0.75	22-28
1	29-36

going to be the best week of my life" would receive a score of 1.0, while the previous one would be scored at 0.5.

Even though it is not possible to provide the entirety of the questionnaire for privacy reasons, its structure is shown in Appendix A. An example of this approach can be seen in Table 4.

Initially, the proposed idea would be to try and identify Plutchik's primary emotions from his dimensional model: anger, anticipation, joy, trust, fear, surprise, sadness and disgust. However, insufficient scores for anger, anticipation, sadness, fear, surprise and disgust were found.

By analyzing the questionnaire again, researchers and psychologists concluded that the statements inferred other emotion and thus, anxiety, boredom, distraction, shame, optimism and interest were included, replacing the six previously emotions from Plutchik's model.

This happened due to the nature of the questionnaire, which aimed to assess various aspects, such as personal motivation, engagement in school, task management, perseverance, beliefs and attitudes toward the educational environment, active participation, family support, interpersonal relationships, and students' perceptions of teachers and the school environment and therefore presented statements that allowed to extract happiness (also known as joy), trust, optimism, interest, boredom, anxiety, distraction, and shame.

These emotions are, in fact, within the project's scope, as they are emotions that are often present amongst young learners.

To validate this hypothesis, a crowdsourcing approach was employed. The questionnaire was distributed to a panel of 36 annotators with backgrounds in linguistics, humanities, computer science, and social sciences. Each annotator reviewed the questionnaire, identified the emotions expressed in each sentence, and assigned corresponding scores. Ultimately, the results obtained from this process aligned with those obtained by the the researchers.

Table 5 shows how the annotation process happened. Taking into consideration the annotators' answers, a certain weight was assigned to each emotion present on the questionnaire.

A dataset was constructed, encompassing all questionnaire statements with dedicated columns for each extracted emotion. For every statement, a score ranging from 0 to 1 was assigned to each emotion (see Table 6 considering the emotions conveyed when the respondent agrees with the statement. This annotation took into account the intensity of the specific emotion, as elaborated previously. Assuming full agreement with a statement, a particular emotion or combination of emotions would be inferred, constituting a specific emotional inference. Subsequently, the scores for each emotion based on the student's response were calculated according to the proposed method.

Table 6: Demonstration of the proposed approach.

(a) Emotion Classification for the Questionnaire.

question	happiness	trust	optimism	interest	boredom	anxiety	distraction	shame
Q1.1	0.25	0.25	0.75	0.75	0	0.25	0	0
Q5.9	0	0	0	0	0.25	0.50	0.75	0.50
Q13.3	0.50	0.50	0.50	0.25	0.25	0.25	0	0.25

(b) Emotion Classification results.

student	happiness	trust	optimism	interest	boredom	anxiety	distraction	shame
1	0.19	0.14	0.25	0.25	0.56	0.53	0.50	0.53
2	0.58	0.51	0.60	0.57	0.69	0.83	0.80	0.56
3	0.73	0.69	0.70	0.66	0.52	0.61	0.37	0.46

Chapter 6

Results and Discussion

This sections aims to analyse the results and present the main findings.

6.1 Emotion Classification

The initial analysis of emotion classification shows that *anxiety* is the emotion that presented the highest scores amongst the 8 detected emotions, as presented on Table 7. *Anxiety*, defined as a blend of anticipation and fear, manifests through heightened alertness, stress, and fear. Pekrun et al. (2002)'s had previously reported anxiety as the most present emotion, accounting for 15% to 25% of all emotions across studies and being associated not only with exams but also class attendance and home study. Achievement pressure and fear of failure were identified as significant contributors to emotional arousal, emphasizing the importance of promoting students' psychological well-being by addressing excessive demands and providing opportunities for success.

A surprising discovery emerged as average *anxiety* scores were slightly elevated among male students. Contrary to findings in previous studies, which generally indicate higher anxiety levels among female students Pekrun et al. (2011) Hosseini and Khazali (2013), these results revealed a small but noteworthy difference. This finding aligns with the trend observed in *shame*, where male students also reported slightly higher scores than their female counterparts. However, insights gained from the questionnaire responses suggest that, on the whole, male students express greater concerns about improving their academic performance, showcasing proficiency in schoolwork, and maintaining a positive image among peers. This

Table 7: Mean Values for all Emotions in Pilot Study.

Anxiety	**Shame**	**Boredom**	**Happiness**	**Distraction**	**Optimism**	**Trust**	**Interest**
0.60	0.55	0.55	0.52	0.52	0.50	0.48	0.46

sheds light on the higher anxiety scores among male students and lends support to the outcomes derived from the proposed emotion classification approach.

Furthermore, an examination of anxiety scores across grades revealed that male students consistently reported higher anxiety scores in grades 5, 7, and 8. In contrast, female students exhibited higher anxiety scores in grades 6 and 9. It is noteworthy, however, that the disparity in scores during grade 6 is minimal.

Shame and *boredom* also displayed elevated scores while positive emotions such as *optimism*, *trust*, and *interest* exhibited the lowest scores. In fact, *interest* seemed to show the lowest scores out of all emotions in all school years.

Table 8 shows the mean emotional scores throughout the school years, which reveals noteworthy patters, and in Figure 17 we can see that distribution in a line chart. There is a consistent decrease in positive emotions (Happiness, Trust, Optimism, and Interest) indicating a potential decline in overall positive well-being as students progress through the school years.

Fifth-grade students exhibited higher values for all emotions compared to students in other grades. Across the school years, there is a gradual decrease in overall happiness, trust, optimism, and interest, indicating a potential decline in emotional well-being as students progress from the fifth grade to the ninth grade.

The shift from sixth grade to seventh grade seems to mark the initial indications of a significant influence on emotional states. Happiness, trust, and optimism undergo a noticeable decline, while levels of boredom and distraction rise. This transitional phase could present difficulties for students' emotional well-being, given the increased responsibilities and workload associated with this period. In Portugal, the move to seventh grade is recognized as particularly challenging, with students encountering more subjects than in the preceding fifth and sixth grades.

While it is true that all (almost) 8 emotions decrease over the course of the fifth to the eighth grade, it is important to note that positive emotions (Happiness, Trust, Optimism, and Interest) decline significantly more than the negative emotions (Boredom, Anxiety, Distraction, Shame). Nonetheless, all emotions seem to reach their lowest scores on the eighth grade. In Portugal there is a conception that the eighth grade is "hard" and demanding. The results obtained go along that conception, showing that students seem to struggle emotionally at that point.

Conversely, *boredom* demonstrates stability, remaining relatively constant with minor fluctuations across the observed grades. This suggests that feelings of boredom persist consistently despite advancing school years. This is also the case of shame, which seems to remain relatively consistent. Exhibiting a slightly higher mean value among male students, the emotion of shame appears to display stability in

Table 8: Mean Values for all Emotions per School Year in Pilot Study.

	5	**6**	**7**	**8**	**9**
Happiness	0.57	0.57	0.50	0.44	0.49
Trust	0.54	0.53	0.47	0.40	0.44
Optimism	0.56	0.54	0.48	0.41	0.45
Interest	0.52	0.50	0.45	0.38	0.42
Boredom	0.57	0.56	0.54	0.52	0.53
Anxiety	0.62	0.60	0.60	0.56	0.59
Distraction	0.54	0.51	0.53	0.50	0.52
Shame	0.56	0.55	0.54	0.53	0.54

the fifth and sixth grades. It undergoes a slight decrease in the eighth grade but sees a modest increase by the ninth grade. Shame, characterized by a blend of fear and disgust, typically involves sensations of bitterness and stress. Notably, male students consistently report higher shame scores in each school year, except for the ninth grade.

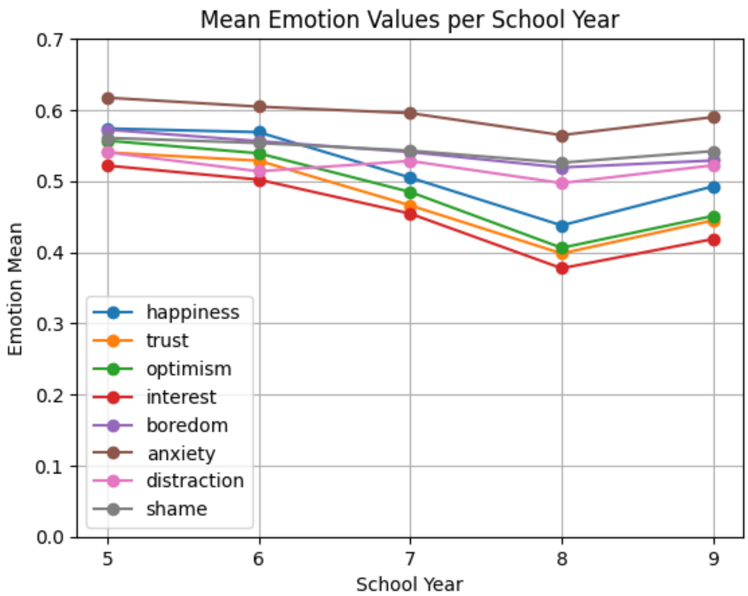
It is interesting to note that the scores for *distraction* increase slightly from the sixth to the seventh grade. Distraction, as according to Plutchik's model, refers to feeling scattered, uncertain and unfocused, not knowing that to prioritize. This might have to do with the fact that in the Portuguese school system, students take on more subjects in the seventh grade, implying more schoolwork, a longer schedule and new teachers. This can make students feel rather lost, hence explaining this score.

Emotions and Parents' Education Level

Mother's Education. As the level of the mother's education increases, we generally see an increase in the average scores for positive emotions like happiness, trust, and optimism. For example, happiness increases from about 0.500 for the lowest education level to 0.568 for the highest. Similarly, there is an increase in the average scores for negative emotions like anxiety, distraction, and shame. This suggests that higher maternal education correlates with higher emotional intensity in both positive and negative aspects. *Father's Education*. A similar trend is observed with the father's education. Higher levels of paternal education are associated with higher average scores for positive emotions. Likewise, the scores for negative emotions also increase with higher levels of paternal education.

These results indicate a general trend where higher parental education is associated with more pronounced emotional states, both positive and negative, in students.

Figure 17: Emotions per school year.



Emotions and Student Gender

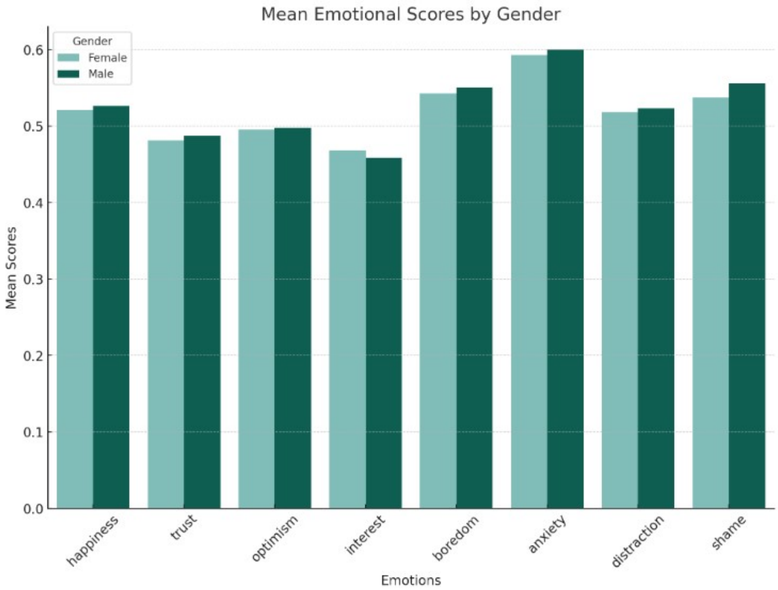
The differences in emotional scores between the genders are relatively small but consistently show that male students tends to have slightly higher scores in both positive and negative emotions. *Interest*, however, seems to report a higher average for female students than for male students, as represented in Figure 18.

Emotions and Parents' Marital Status

The analysis of students' emotions in relation to their parents' marital status reveals interesting trends, represented in Figure 19. Students whose parents are married or in a civil union typically exhibit higher levels of positive emotions such as happiness, trust, and optimism. This suggests a possible correlation between a stable family environment and the emotional well-being of students. This trend is closely followed by students with divorced and single parents, with students in the "other/unknown" category often showing the lowest levels of these emotions. Interestingly, anxiety presents a different pattern, being higher among students with single parents, indicating that factors other than parents' marital status might play a more significant role in this aspect. Emotions like distraction, shame, and boredom do not show considerable variations across different parental marital statuses, suggesting a lesser influence of the family environment on these feelings.

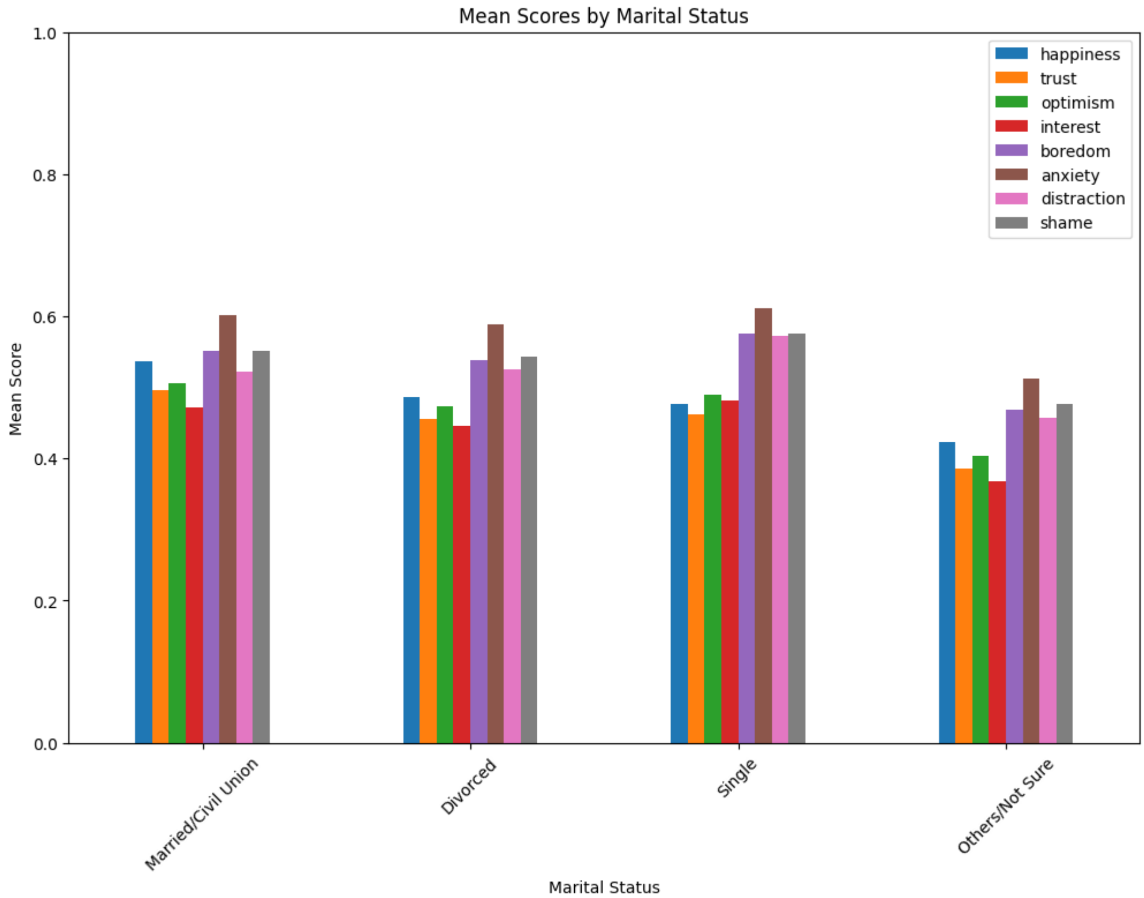
After the t-test, results indicate that, for the emotions of happiness, trust, interest, optimism, and anxiety, there is a significant difference between students whose parents are married/in a civil union and others, supporting the hypothesis that these positive emotions are higher in that group. For distraction,

Figure 18: Mean Emotions per Student Gender.



shame, and boredom, the results are not statistically significant, indicating that there is no significant difference between the groups for these emotions.

Figure 19: Mean Emotions per Parents' Marital Status.



Emotions and Failing

Table 9 shows the mean emotion scores for students who have previously failed and students who have not.

Students who have not failed exhibit higher average scores for both positive (happiness, trust, optimism, interest) and negative (boredom, anxiety, distraction, shame) emotions compared to those who have failed. The results suggest a more intense emotional experience in students who have not failed. This could reflect a higher engagement or investment in their academic performance.

The higher scores in negative emotions for students who have not failed may indicate a greater level of stress or concern about maintaining their performance. The higher positive emotional scores in the same group might reflect satisfaction or confidence in their academic abilities.

6.2 Correlation to Risk Failure

The correlation between the emotions and the risk of failure was calculated and is shown on Figure 20.

Table 9: Comparison of Mean Emotions per students who have failed and students who have not failed.

Emotion	Has Failed Before	Has Not Failed Before
Happiness	0.437	0.528
Trust	0.414	0.488
Optimism	0.421	0.500
Interest	0.374	0.468
Boredom	0.478	0.550
Anxiety	0.518	0.601
Distraction	0.458	0.524
Shame	0.478	0.550

Emotions as *interest*, *optimism* and *happiness* seem to have a moderate correlation with the risk of failure. For example, there seems to be a trend where higher levels of optimism seem to be associated with a lower risk of failure.

Emotions such as *distraction* and *shame* seem to have weaker correlations.

This distribution can be seen individually in Figure 21. Cluster 4 represents low risk of failure while cluster 1 represents high risk of failure.

It can be seen that students at a low risk of failing seem to present higher scores for all the emotions than students within the other risk clusters.

This distribution shows that as the risk of failure increases, positive emotions such as *happiness*, *trust*, *optimism* and *interest* decrease significantly. Whereas negative emotions, such as *boredom*, *anxiety*, *distraction* and *shame* also experience a decrease, it is not as sharp as it is for positive emotions.

The fact that students at a low risk of failure show higher scores for all emotions can be explained due to the fact that these students might be more engaged with the school setting than students at a high risk of failing.

6.3 Machine Learning Models

Both **Mean Absolute Error** and **Mean Squared Error** are commonly used for regression tasks, and lower values indicate better performance.

In general, the values of **MAE** and **MSE** decrease after cross-validation for all models. This suggests that cross-validation helps in obtaining more robust and reliable performance metrics by assessing the

Figure 20: Mean Emotions per Risk of Failing.

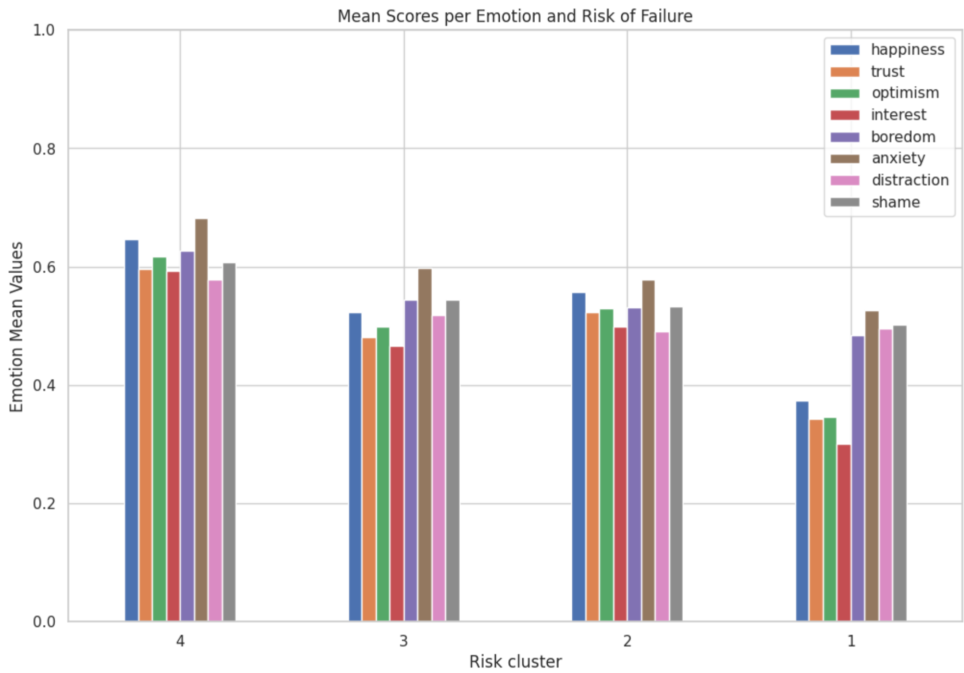
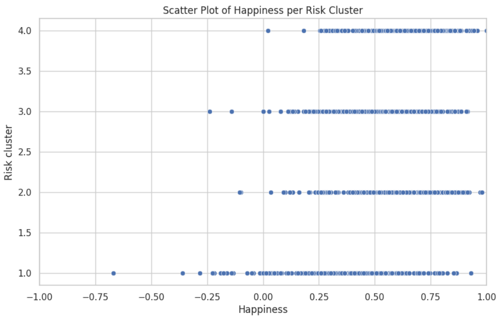


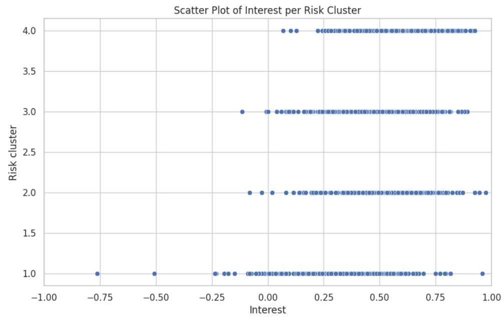
Table 10: **MAE** and **MSE** results before and after cross-validation.

	Before Cross Validation		**After Cross Validation**	
	MAE	**MSE**	**MAE**	**MSE**
XGBoost	.1819	.0546	.1763	.0517
SVM	**.1539**	**.0405**	**.1491**	**.0361**
Decision Tree	.2025	.0677	.1968	.0635
Random Forest	.1649	.0454	.1567	.0403

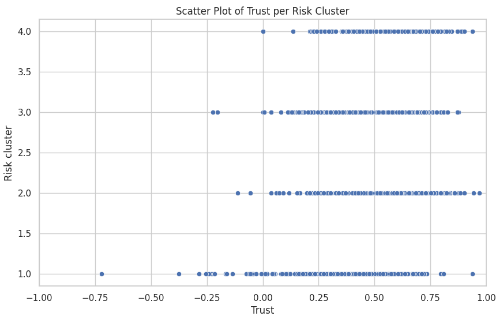
Figure 21: Scatter Plots of Emotions per Risk Clusters.



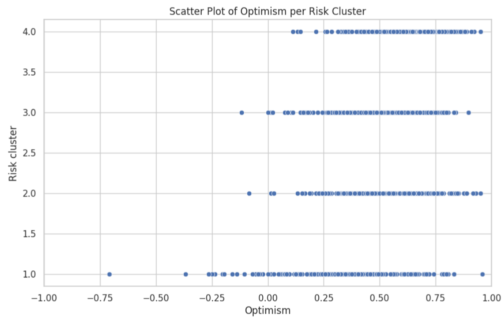
(a) Happiness.



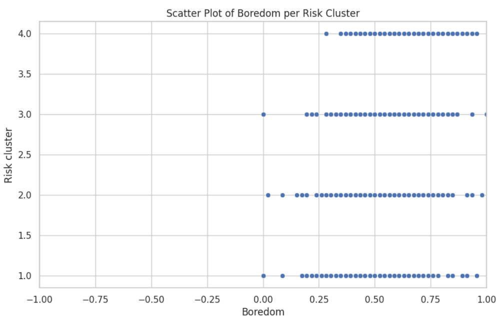
(c) Interest.



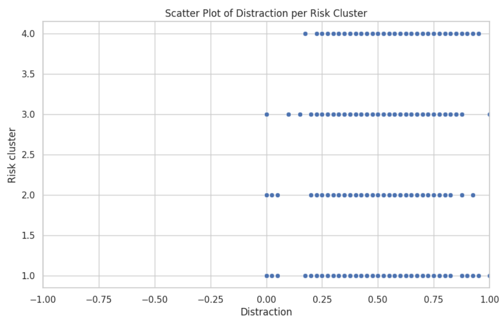
(b) Trust.



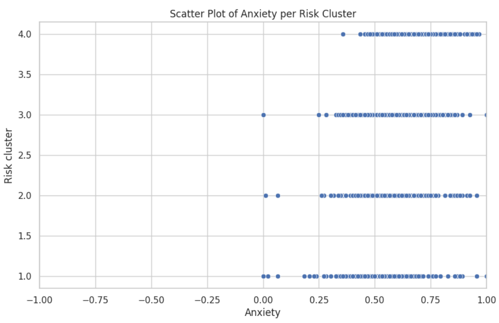
(d) Optimism.



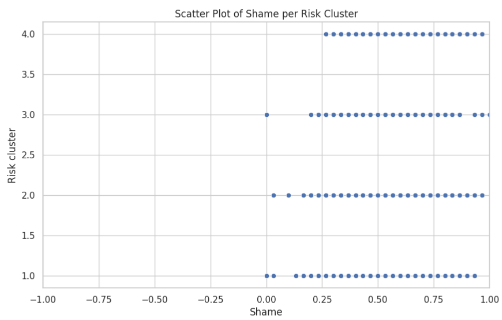
(e) Boredom.



(g) Distraction.



(f) Anxiety.



(h) Shame.

model on multiple subsets of the data. The improvement in **MAE** and **MSE** after cross-validation is relatively small, but it still suggests that cross-validation helps in refining the model and reducing potential overfitting (see Table 10).

Comparing the models, SVM seems to have the lowest **MAE** and **MSE** both before and after cross-validation, indicating better performance compared to XGBoost, Decision Tree, and Random Forest. Random Forest appears to be the second-best performer.

The fact that the model rankings (SVM > Random Forest > XGBoost/Decision Tree) are consistent before and after cross-validation is a positive sign. Consistency in performance indicates that the models are not highly sensitive to the specific data splits used in cross-validation.

This means that SVM is the model that showed the best performance in accurately predicting the targeted emotions.

6.4 Discussion

The proposed approach demonstrates the potential of extracting additional features from Likert-scale questionnaires. The case study conducted provides an example of how this questionnaire-based approach to emotion mining facilitates a comprehensive exploratory analysis of students' emotional states at the start of the school year. This contributes significantly to the existing literature in the fields of sentiment analysis and emotion classification. Traditional sentiment analysis often struggles with language ambiguity and the development of lexicons, particularly in languages other than English.

This research shows the importance of developing different ways to detect emotions, especially for different languages and areas. Sentiment analysis is growing beyond the English language, fueling the necessity to make good emotion lexicons for other languages too (Medhat et al., 2014). Current practices report that most lexicons are developed in English and they are often not complete, or they present literal translations of English words, as the NRC Emotion Lexicon (EmoLex), for example. This represents an issue in performing sentiment analysis in other languages. Thus, this research presents itself as a solution to this problem because it looks at each statement in the questionnaire in the language it is written in. Another challenge that Affuso et al. (2023) state is predicting and creating advantages for the educational system at a low cost.

It's well-known that support from teachers and parents is important for keeping students interested and motivated in school (Affuso et al., 2023). The questionnaire used in this study covers all the key points that matter for having academic success – from what the students themselves think about their school goals to how much support they get from friends and family. Additionally, this emotion detection approach is not invasive.

What is more, the significance of this advancement extends beyond the academic setting. While this approach was applied in an educational context, its potential applications are vast and diverse. Industries such as marketing, healthcare, and customer service could benefit from this refined emotion detection technique, enabling better understanding and responsiveness to client and patient emotions.

Chapter 7

Conclusions and Future work

7.1 Conclusions

This dissertation presents the development and application of an emotion detection approach through a Likert-Scale questionnaire, applied to an educational setting. This allowed for gaining insights into emotions such as happiness, trust, interest, optimism, boredom, distraction, anxiety, and shame, contributing, therefore, to the understanding of the psychological well-being of students.

The main findings lie on anxiety presenting the highest scores amongst all emotions, particularly among male students.

Additionally, the transition to the seventh grade seems to affect students' happiness, trust, and optimism, most likely due to the increase of workload and class schedule.

Correlations with the risk of failure unveiled crucial associations, indicating that higher levels of optimism, interest, and happiness were linked to lower risks.

The evaluation of machine learning models, including XGBoost, Decision Tree, Random Forest, and SVM, demonstrated that SVM outperformed others in terms of **Mean Absolute Error** and **Mean Squared Error**, both before and after cross-validation. This highlights the potential of advanced machine learning techniques in accurately predicting and understanding emotional states among students.

In summary, this research sheds light on the intricate relationship between emotions and academic experiences among students aged 10-17. The findings not only contribute to the existing literature on emotional well-being but also provide practical implications for educators, parents, and policymakers. As the emotional well-being of students appears to decline with academic progression, addressing these issues becomes imperative for creating a supportive and nurturing educational environment.

7.2 Research Questions

The following research questions were presented initially and are now revisited analysed:

- **Is it possible to detect students' emotional traits through tools that were not intended for that purpose?** Yes. Students' emotional traits were extracted through a questionnaire of 50 statements, aimed at evaluating multiple aspects including personal motivation, engagement in school, task management, perseverance, beliefs and attitudes towards the educational environment, active participation, family support, interpersonal relationships, and students' perceptions of teachers and the school environment.

- **Is it possible to relate school failure and emotions?** Yes. Research carried out in this dissertation showed that students who have not failed show higher average scores for both positive (happiness, trust, optimism, interest) and negative (boredom, anxiety, distraction, shame) emotions compared to those who have failed. This suggests that students who have not failed experience more intense emotions, potentially reflecting higher engagement or investment in their academic performance. Interestingly, students who have not failed also score higher in negative emotions, which might indicate a greater level of stress or concern about maintaining their performance. This could be a reflection of the pressure to continue succeeding. What is more, emotions like interest, optimism, and happiness show a moderate correlation with the risk of failure, where higher levels of these emotions are associated with a lower risk of failure. It was found that emotions such as distraction and shame have weaker correlations with the risk of failure. Students at a low risk of failing present higher scores for all emotions compared to those in higher risk clusters. As the risk of failure increases, positive emotions decrease significantly, while negative emotions also decrease, but not as sharply. Finally, the fact that students at low risk of failure show higher scores for all emotions could be due to their higher engagement with the school setting compared to students at a high risk of failing.

- **How correlated is school dropout and/or school failure to emotions?** Previous studies have shown that high levels of emotional engagement were linked to lower risk of school failure and dropout and the results presented in this disseration also confirm that students at a low risk of failure prevent higher emotional scores. Therefore, these seems to be a significant correlation between school dropout and or/failure and students' emotions. What is more, understanding a student's emotional state allows for more personalized intervention strategies. This gives teachers and schools an opportunity to work on tailored strategies to promote school engagement and success, as well as predicting cases of school failure and act in a timely manner.

7.3 Contributions

The work developed in this dissertation resulted in the following publications:

- Magalhães, R., Veloso, B., Marcondes, F. S., Lima, H., Durães, D., & Novais, P. (2023, June). A Conversational Agent for Smart Schooling A Case Study on K-12 Dropout Risk Assessment. In Sustainable Smart Cities and Territories International Conference (pp. 124-134). Cham: Springer Nature Switzerland. Magalhães et al. (2023b)
- Magalhães, R., Marcondes, F. S., Durães, D., & Novais, P. (2023, November). Emotion Extraction from Likert-Scale Questionnaires: –An Additional Dimension to Psychology Instruments–. In International Conference on Intelligent Data Engineering and Automated Learning (pp. 166-176). Cham: Springer Nature Switzerland. Magalhães et al. (2023a)

It is important to note that specific sections of this dissertation, specifically the introduction and the emotion detection methodology were drawn from the content presented in these papers.

7.4 Prospect for future work

While the current research employed an emotion detection approach without relying on spontaneous textual data, there remains a valuable opportunity for further refinement.

Administering the questionnaire at both the beginning and end of the school year could provide a longitudinal perspective on students' emotional states, allowing for a nuanced analysis of emotional trends over time. This approach would facilitate a comprehensive comparison of emotional dynamics, offering insights into possible fluctuations or oscillations.

Additionally, the integration of a frontend solution could enhance the visualization of students' emotional data, providing a user-friendly interface for educators and administrators to interpret and respond to the findings effectively.

Machine Learning (ML) models such as Support Vector Machines could also successfully be implemented in order to accurately predict students' emotions, allowing for timely monitoring and intervention.

It would also be quite interest to expand the proposed methodology to detect students' personality traits. This unexplored area has the potential to give us a more complete understanding of students' psychological profiles, providing a wider basis for creating personalized interventions and support strategies.

Bibliography

Adamopoulou, E. and Moussiades, L. (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications*, 2:100006.

Affuso, G., Zannone, A., Esposito, C., Pannone, M., Miranda, M. C., De Angelis, G., Aquilar, S., Dragone, M., and Bacchini, D. (2023). The effects of teacher support, parental monitoring, motivation and self-efficacy on academic performance over time. *European Journal of Psychology of Education*, 38(1):1–23.

Algarni, A. (2016). Data mining in education. *International Journal of Advanced Computer Science and Applications*, 7(6).

Aljarrah, A., Ababneh, M., Karagozlu, D., and Ozdamli, F. (2021). Artificial intelligence techniques for distance education: A systematic literature review. *TEM Journal*, 10(4).

Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., and Christopherson, R. (2009). Emotion sensors go to school. In *International Conference on Artificial Intelligence in Education*.

Baragash, R. S., Aldowah, H., and Umar, I. N. (2022). Students' perceptions of e-learning in malaysian universities: Sentiment analysis based machine learning approach. *J. Inf. Technol. Educ. Res.*, 21:439–463.

Bibi, M., Abbasi, W. A., Aziz, W., Khalil, S., Uddin, M., Iwendi, C., and Gadekallu, T. R. (2022). A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for twitter sentiment analysis. *Pattern Recognition Letters*, 158:80–86.

Binali, H. H., Wu, C., and Potdar, V. (2009). A new significant area: Emotion detection in e-learning using opinion mining techniques. In *2009 3rd IEEE international conference on digital ecosystems and technologies*, pages 259–264. IEEE.

Bóbó, M. L., Campos, F., Stroele, V., David, J. M. N., Braga, R., and Torrent, T. T. (2022). Using sentiment analysis to identify student emotional state to avoid dropout in e-learning. *International Journal of Distance Education Technologies (IJDET)*, 20(1):1–24.

Bouhlal, M., Aarika, K., Abdelouahid, R. A., Elfilali, S., and Benlahmar, E. (2020). Emotions recognition as innovative tool for improving students' performance and learning approaches. *Procedia Computer Science*, 175:597–602.

Bustos-López, M., Cruz-Ramírez, N., Guerra-Hernández, A., Sánchez-Morales, L. N., and Alor-Hernández, G. (2021). Emotion detection from text in learning environments: A review. *Studies in Computational Intelligence*.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., et al. (2000). Crisp-dm 1.0: Step-by-step data mining guide. *SPSS inc*, 9(13):1–73.

Conselho Nacional de Educação, C. (2023). Estado da educação 2021. https://www.cnedu.pt/content/EE2021/EE2021-Web$_{s}ite.pdf$.

Cordeiro, P., Paixão, M., Silva, J. T., de Castro, M., and Morais, M. (2020). O sistema alerta precoce para o insucesso escolar – sapie. *Psychologica*, 63:139–152.

Cui, J., Wang, Z., Ho, S., and Cambria, E. (2023). Survey on sentiment analysis: evolution of research methods and topics. *Artif. Intell. Rev.*, 56(8):8469–8510.

Dake, D. K. and Gyimah, E. (2023). Using sentiment analysis to evaluate qualitative students' responses. *Education and Information Technologies*, 28(4):4629–4647.

Das, S., Dey, A., Pal, A., and Roy, N. (2015). Applications of artificial intelligence in machine learning: review and prospect. *International Journal of Computer Applications*, 115(9).

Devi, C. P. et al. (2021). Banking chatbot (b-bot. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(10):5795–5804.

Direção Geral de Estatística da Educação e da Ciência (2018). https://www.dgeec.medu.pt/np4/96/.

Dorfman, M., Grondona, A., Mazza, N., and Mazza, P. (2001). Asistentes virtuales de clase en la educación universitaria. *UNDÉCIMAS JORNADAS DE TECNOLOGÍA APLICADA A LA EDUCACIÓN MATEMÁTICA UNIVERSITARIA*, page 41.

ENCAMINA (2019). https://www.encamina.com/en/samebullying/.

Erdem, C. and Kaya, M. (2020). A meta-analysis of the effect of parental involvement on students' academic achievement. *Journal of Learning for Development*, 7(3):367–383.

European Commission (2019). *Assessment of the implementation of the 2011 Council recommendation on policies to reduce early school leaving – Final report*. Publications Office.

Faria, H., Barbosa, M. A., Veloso, B., Marcondes, F. S., Lima, C., Durães, D., and Novais, P. (2022). Edubot: A proof-of-concept for a high school motivational agent. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 223–232. Springer.

Freeman, J., Simonsen, B., Mccoach, D. B., Sugai, G., Lombardi, A., and Horner, R. (2015). An analysis of the relationship between implementation of school-wide positive behavior interventions and support and high school dropout rates. *The High School Journal*, 98:290–315.

Gorde, M. A., Shilimkar, M. A., Kulkarni, M. S., Chavan, M. T., and Ali, N. R. Q. (2023). A survey on emotion detection using machine learning techniques.

Gubareva, R. and Lopes, R. P. (2020). Virtual assistants for learning: A systematic literature review. *CSEDU (1)*, pages 97–103.

Gupta, S., Kumar, P., and Tekchandani, R. K. (2023). Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. *Multimedia Tools and Applications*, 82(8):11365–11394.

Gupta, V. (2014). A survey of natural language processing techniques. *International Journal of Computer Science & Engineering Technology*, 5(1):14–16.

Han, J. (1996). Data mining techniques. In *SIGMOD Conference*, page 545. ACM Press.

Hosseini, L. and Khazali, H. (2013). Comparing the level of anxiety in male & female school students. *Procedia-social and behavioral sciences*, 84:41–46.

Hu, T., Zhang, X., and Li, N. (2022). Design of intelligent evaluation system for college students' mental health based on big data. *Computational Intelligence and Neuroscience*, 2022.

Iqbal, A., Amin, R., Iqbal, J., Alroobaea, R., Binmahfoudh, A., and Hussain, M. (2022). Sentiment analysis of consumer reviews using deep learning. *Sustainability*, 14(17):10844.

Jayalekshmi, J. and Mathew, T. (2017). Facial expression recognition and emotion classification system for sentiment analysis. In *2017 International Conference on Networks & Advances in Computational Technologies (NetACT)*, pages 1–8. IEEE.

Joshi, A. K. (1977). Natural language processing. *ACM SIGART Bulletin*, (61):44–46.

Kaur, P., Kumar, H., and Kaushal, S. (2021). Affective state and learning environment based analysis of students' performance in online assessment. *International Journal of Cognitive Computing in Engineering*, 2:12–20.

Keele, S. et al. (2007). Guidelines for performing systematic literature reviews in software engineering.

Kelly, S., Kaye, S.-A., and Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? a systematic review. *Telematics and Informatics*, 77:101925.

Kiritchenko, S., Zhu, X., and Mohammad, S. M. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50:723–762.

Kort, B., Reilly, R., and Picard, R. W. (2001). External representation of learning process and domain knowledge: Affective state as a determinate of its structure and function. In *Workshop on Artificial Intelligence in Education (AI-ED 2001), San Antonio,(May 2001)*, pages 64–69.

Li, W. and Xu, H. (2014). Text-based emotion classification using emotion cause extraction. *Expert Systems with Applications*, 41(4):1742–1749.

Magalhães, R., Marcondes, F. S., Durães, D., and Novais, P. (2023a). Emotion extraction from likert-scale questionnaires: –an additional dimension to psychology instruments–. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 166–176. Springer.

Magalhães, R., Veloso, B., Marcondes, F. S., Lima, H., Durães, D., and Novais, P. (2023b). A conversational agent for smart schooling a case study on k-12 dropout risk assessment. In *Sustainable Smart Cities and Territories International Conference*, pages 124–134. Springer.

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9(1):381–386.

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M. J., and Flach, P. (2019). Crisp-dm twenty years later: From data mining processes to data science trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8):3048–3061.

Medhat, W., Hassan, A., and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113.

Mendoza, S., Sánchez-Adame, L. M., Urquiza-Yllescas, J. F., González-Beltrán, B. A., and Decouchant, D. (2022). A model to develop chatbots for assisting the teaching and learning process. *Sensors*, 22(15):5532.

Miguel, R. R., Rijo, D., and Lima, L. N. (2012). Fatores de risco para o insucesso escolar: a relevância das variáveis psicológicas e comportamentais do aluno. *Revista portuguesa de pedagogia*, pages 127–143.

Milne-Ives, M., de Cock, C., Lim, E., Shehadeh, M. H., de Pennington, N., Mole, G., Normando, E., and Meinert, E. (2020). The effectiveness of artificial intelligence conversational agents in health care: systematic review. *Journal of medical Internet research*, 22(10):e20346.

Mohammad, S. M. and Turney, P. D. (2013a). Crowdsourcing a word–emotion association lexicon. *Computational intelligence*, 29(3):436–465.

Mohammad, S. M. and Turney, P. D. (2013b). Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Mondal, B. (2020). Artificial intelligence: state of the art. *Recent Trends and Advances in Artificial Intelligence and Internet of Things*, pages 389–425.

Muñoz, S., Sánchez, E., and Iglesias, C. A. (2020). An emotion-aware learning analytics system based on semantic task automation. *Electronics*, 9(8):1194.

N, G., C, K., and Shetty, H. (2023). Human emotion detection using machine learning. *International Journal for Research in Applied Science and Engineering Technology*, 11:3065–3070.

Norvig, P. R. and Intelligence, S. A. (2002). A modern approach. *Prentice Hall Upper Saddle River, NJ, USA: Rani, M., Nayak, R., & Vyas, OP (2015). An ontology-based adaptive personalized e-learning system, assisted by software agents on cloud storage. Knowledge-Based Systems*, 90:33–48.

Peguero, A. (2011). Violence, schools, and dropping out: Racial and ethnic disparities in the educational consequence of student victimization. *Journal of interpersonal violence*, 26:3753–72.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., and Perry, R. P. (2011). Measuring emotions in students' learning and performance: The achievement emotions questionnaire (aeq). *Contemporary educational psychology*, 36(1):36–48.

Pekrun, R., Goetz, T., Titz, W., and Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist*, 37(2):91–105.

Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89(4):344–350.

Poria, S., Cambria, E., Bajpai, R., and Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information fusion*, 37:98–125.

Ranjan, S. and Mishra, S. (2022). Perceiving university students' opinions from google app reviews. *Concurrency and Computation: Practice and Experience*, 34(10):e6800.

Rasa (2023). Introduction to rasa open source & rasa pro. https://rasa.com/docs/rasa/.

Ray, S. (2019). A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*, pages 35–39. IEEE.

Rebala, G., Ravi, A., and Churiwala, S. (2019). Natural language processing. *An Introduction to Machine Learning*.

Rodriguez, P., Ortigosa, A., and Carro, R. M. (2012). Extracting emotions from texts in e-learning environments. In *2012 Sixth International Conference on Complex, Intelligent, and Software Intensive Systems*, pages 887–892. IEEE.

Ruiz, S., Urretavizcaya, M., Lafuente, C. R., and de Castro, I. F. (2018). Predicting students' outcomes from emotional response in the classroom and attendance. *Interactive Learning Environments*, 28:107 – 129.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.

Samuel, R. and Burger, K. (2020). Negative life events, self-efficacy, and social support: Risk and protective factors for school dropout intentions and dropout. *Journal of educational psychology*, 112(5):973.

Shaikh, S. and Doudpotta, S. M. (2019). Aspects based opinion mining for teacher and course evaluation. *Sukkur IBA Journal of Computing and Mathematical Sciences*, 3(1):34–43.

Shawar, B. A. and Atwell, E. (2015). ALICE chatbot: Trials and outputs. *Computación y Sistemas*, 19(4).

Shinde, P. P. and Shah, S. (2018). A review of machine learning and deep learning applications. In *2018 Fourth international conference on computing communication control and automation (ICCUBEA)*, pages 1–6. IEEE.

Swathi, B., Babu, S. S., and Ayyavaraiah, M. (2019). Artificial intelligence: Characteristics, subfields, techniquesand future predictions.

Tabassum, A. and Patil, R. R. (2020). A survey on text pre-processing & feature extraction techniques in natural language processing. *International Research Journal of Engineering and Technology (IRJET)*, 7(06):4864–4867.

Talaei Khoei, T. and Kaabouch, N. (2023). Machine learning: Models, challenges, and research directions. *Future Internet*, 15(10):332.

Taurah, S. P., Bhoyedhur, J., and Sungkur, R. K. (2020). Emotion-based adaptive learning systems. In *Machine Learning for Networking: Second IFIP TC 6 International Conference, MLN 2019, Paris, France, December 3–5, 2019, Revised Selected Papers 2*, pages 273–286. Springer.

UNESCO (2018). Out-of-school children and youth. https://uis.unesco.org/en/topic/out-school-children-and-youth.

Vijayarani, S., Ilamathi, M. J., Nithya, M., et al. (2015). Preprocessing techniques for text mining-an overview. *International Journal of Computer Science & Communication Networks*, 5(1):7–16.

Wang, R. and Shi, Z. (2021). Personalized online education learning strategies based on transfer learning emotion classification model. *Security and Communication Networks*, 2021:1–11.

Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., Gao, S., Sun, Y., Ge, W., Zhang, W., et al. (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion*, 83:19–52.

Wankhade, M., Rao, A. C. S., and Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7):5731–5780.

WatsomApp (2019). https://watsomapp.com/. "accessed on 30th October 2023".

Wirth, R. and Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, volume 1, pages 29–39. Manchester.

Xing, W., Tang, H., and Pei, B. (2019). Beyond positive and negative emotions: Looking into the role of achievement emotions in discussion forums of moocs. *The Internet and Higher Education*, 43:100690.

Zablocki, M. (2009). Predicting school dropout among youth with disabilities: The roles of youth characteristics, academic experiences and emotional engagement factors.

Appendix A
Questionnaire

The questionnaire administered to the students contains a total of 50 statements and is divided into seven parts:

- Q1. Personal motivation, school involvement and validation. The importance of learning and academic performance.
- Q5. Task management, perseverance and self-regulation. Distraction and commitment.
- Q14. Beliefs and attitudes related to the educational environment and its significance.
- Q7. Attitudes and behaviors related to active participation and engagement in school.
- Q11. Family support and involvement in school-related matters.
- Q12. Friendships and peer relationships within the school context.
- Q13. Perceptions of students regarding their teachers and the overall school environment.