INTERNATIONAL MULTI-AWARD WINNING INSTITUTION FOR SUSTAINABILITY

Advancement in ICT: Exploring Innovative Solutions (AdICT)

Series 1/2024

KICT Publishing

ADVANCEMENT IN ICT: EXPLORING INNOVATIVE SOLUTIONS (AdICT) SERIES 1/2024

Editors
Noor Azura Zakaria
Dini Oktarina Dwi Handayani
Elin Eliana Abdul Rahim
Ahmad Fatzilah Misman

KICT Publishing

ADVANCEMENT IN ICT: EXPLORING INNOVATIVE SOLUTIONS (AdICT)
SERIES 1/2024

First Publication 2024

Published by

KICT Publishing
International Islamic University Malaysia
53100 Kuala Lumpur, Selangor, Malaysia

e ISBN 978-629-99388-0-4

9 786299 938804

KICT Publishing

(Online)

Preface

Advancement in ICT: Exploring Innovative Solutions (AdICT) Series 1/2024 is an e-book showcases the collective achievements of Final Year Project (FYP) in Kulliyyah of Information and Communication Technology (KICT). This compilation represents evidence to the technical passion and academic skills of our students before they venture into the professional realm.

FYP is a journey that demands creativity, critical thinking, and perseverance. This book encapsulates the diverse range of projects undertaken by our students, each a unique exploration into the vast landscape of Information and Communication Technology (ICT). From cutting-edge software applications to groundbreaking research, these projects not only demonstrate technical proficiency but also the ability to address real-world challenges.

In this comprehensive collection, the topics covered span a spectrum from cutting-edge software development, cybersecurity, artificial intelligence and multimedia technologies reflecting the breadth and depth of our academic program. This offers a curated journey through the diverse landscape of final year ICT projects to the readers while appreciating the impact these projects can have on the wider community.

This e-book carries significant benefits and impact whereby it serves as a valuable knowledge repository, offering a diverse audience—from students and educators to industry professionals—a comprehensive view of the latest innovations and technological solutions in ICT. Moreover, the book fosters a culture of knowledge sharing and collaboration, as each project represents a unique contribution to the broader technological landscape.

“When the human being dies, his deeds end except for three: ongoing charity, beneficial knowledge, or a righteous child who prays for him” – *Sahih Muslim*

Editors

Noor Azura Zakaria
Dini Oktarina Dwi Handayani
Elin Eliana Abdul Rahim
Ahmad Fatzilah Misman

TABLE OF CONTENTS

No.	Content	Page No.
17	**Unveiling Volunteer4U: Mobilizing Opportunities for Volunteering** *Nur Hanani Ab Hannan, Nurul ‘Aqilah Zakaria, Akram M Zeki*	121
18	**Journey of Hajj: The Simulation** *Muhammad Asyraf Azman, Nur Khaliesah Muhamad Radzali, Suhaila Samsuri*	129
19	**Cyber Security Awareness Training (SecurityGuts)** *Fatin Najwa Ramli, Nur Alya Nadirah Mohd Fauzi, Shuhaili Talib*	134

Exploring Students' Performance in Mathematics in Portugal Using Data Analytics Techniques: A Data Science Use-Case

Raini Hassan
Department of Computer Science
Kulliyyah of ICT, International Islamic University Malaysia
Kuala Lumpur, Malaysia
hrai@iium.edu.my

Nur Zulfah Insyirah Fadzleey
Department of Computer Science
Kulliyyah of ICT, International Islamic University Malaysia
Kuala Lumpur, Malaysia
syiraleey@gmail.com

Annesa Maisarah Ab Hamid
Department of Computer Science
Kulliyyah of ICT, International Islamic University Malaysia
Kuala Lumpur, Malaysia
annesamaisarahabhamid@gmail.com

Rabiatul Adawiyah Abd Aziz
Department of Computer Science
Kulliyyah of ICT, International Islamic University Malaysia
Kuala Lumpur, Malaysia
qdina01@gmail.com

Afiefah Jamalullain
Department of Computer Science
Kulliyyah of ICT, International Islamic University Malaysia
Kuala Lumpur, Malaysia
afiefahjamalullain123@gmail.com

Fatin Syafiqah Syaiful ‘Adli
Department of Computer Science
Kulliyyah of ICT, International Islamic University Malaysia
Kuala Lumpur, Malaysia
ftnsyfqh193@gmail.com

***Abstract*—This research project investigates the relationship between family background and student performance in Mathematics in Portugal. The analysis is based on an open-source dataset from Kaggle Datasets, comprising 395 rows and 33 columns, with 24 key features used for predictive analysis. The purpose is to identify the key factors influencing academic performance, providing insights for targeted interventions and support systems. Machine learning algorithms, specifically Random Forest Regression and Decision Trees, are utilized to analyze the dataset and determine the most significant factor impacting student performance. The study employs descriptive and predictive analytics techniques to understand student performance patterns and forecast future outcomes based on family background factors. The practical application of this research lies in developing predictive models that inform data-driven decisions by educators and policymakers. The results, as shown in Table III, indicate that the Random Forest Regression model outperforms the Decision Tree model, achieving lower Mean Squared Error (9.6212), Root Mean Squared Error (3.0842), and Mean Absolute Error (2.4060). The findings highlight the importance of parental education levels and positive family relationships in influencing academic performance in Mathematics. Future research endeavours should explore the applicability of these findings to other nations, such as Malaysia, to gain a broader understanding of the factors influencing student academic success and adapt data-driven interventions accordingly.**

Keywords—Student Performance, Data Analysis, Mathematics, Portugal, Random Forest Regression, Decision Trees, Machine Learning

I. Introduction

A. Research Overview

In today's data-driven world, applying machine learning algorithms and analytics has become increasingly important across various domains. In the field of education, these technologies have the potential to revolutionize the learning environment and improve student outcomes. By leveraging data analysis techniques, educators can gain valuable insights into factors influencing student performance, enabling them to develop targeted interventions and support systems.

This research project explores the relationship between family background and student performance in Mathematics in Portugal. The choice of student performance in Mathematics in Portugal as our case study stems from the subject's universal significance as a fundamental pillar of education and its role in shaping students' academic trajectories and future opportunities. Additionally, the unique characteristics of the Portuguese educational system, including curriculum structures and grading practices, make it an intriguing case for analysis. Additionally, the findings from this study can serve as a valuable reference for policymakers in other nations seeking to enhance their education systems.

To guide through this exploration, we will first delve into a comprehensive background review, followed by an overview of the methodology and data used. Subsequently, we will present the results, discuss their implications, and outline avenues for future research in this domain. This structured approach ensures that readers gain a clear understanding of our research journey and its contributions to the field of education and predictive analytics.

Academic performance is a crucial aspect of a student's educational journey and has long-term implications for their future success [6]. By identifying the factors that significantly impact academic performance, educators and policymakers can implement strategies to enhance learning experiences and support students in achieving their full potential.

The research problem at hand revolves around understanding the key factors that contribute to a student's academic performance in Mathematics. This problem falls within the education domain, specifically in the context of improving mathematics proficiency among school students in Portugal. The goal is to analyze the dataset and utilize machine learning algorithms, such as Random Forest Regression and

Decision Trees, to identify the most significant factor impacting student performance.

We can employ descriptive and predictive analytics techniques through the application of analytics. Descriptive analytics will help us understand student performance in Mathematics in Portugal and identify any existing patterns or trends. Predictive analytics will enable us to forecast academic outcomes based on family background factors, providing insights into future performance and the potential for targeted interventions.

The practical application of this research lies in its ability to develop predictive models that can assist educators and policymakers in making data-driven decisions. By understanding the relationship between family background and academic performance, stakeholders can design interventions to support students who may be at risk of underperforming. Additionally, parental involvement can be emphasized, fostering collaborative efforts between parents, guardians, and educators to optimize student outcomes.

In summary, this research project utilizes machine learning algorithms and analytics to investigate the relationship between family background and student performance in Mathematics in Portugal. By identifying the primary factor influencing academic performance, we can provide valuable insights and practical implications for educators, policymakers, and other stakeholders in the education system. Through the development of predictive models, this research can improve educational outcomes and foster personalized learning approaches for students.

B. Background

Portugal's unique educational landscape, characterized by a diverse socio-economic fabric and distinct regional variations, makes it an intriguing case study for exploring the intricate interplay between family background factors and student performance in Mathematics [9]. By delving into this specific context, we aim to offer insights that are not only valuable to Portugal's education system but can also inform similar studies in other nations facing educational challenges rooted in family-related dynamics.

The invention of the internet and the growth of more potent computer technology took place in the 21st century. Machine learning and artificial intelligence are two of these new technologies. The application of this technology in Educational Data Mining is one of the scenarios. Contemporary methods, tactics, and implementations derived from educational data mining are pivotal in enhancing the educational ecosystem [4]. Early performance evaluation of students can help identify their strengths and weaknesses and improve their performance. To determine what features correspond with high student academic performance, our study proposed using these machine learning technologies to perform data analysis on the Student Performance Dataset. Machine learning technologies have emerged as powerful tools for analyzing educational data because they can uncover complex patterns, handle large datasets, and make data-driven predictions. By employing machine learning algorithms, this study will be able to uncover intricate relationships between various student-related features and academic performance, providing valuable insights for educators and policymakers.

In predictive analytics in education, a comprehensive body of research has explored the multifaceted relationship between various factors and academic performance. For instance, studies such as that by Tanujaya et al. [13] have delved into the connection between higher-order thinking skills and students' academic performance in mathematics, revealing that higher-order thinking skills correlate positively with GPA in mathematics instruction. This indicates the significance of considering various factors, including cognitive skills, in understanding and predicting academic outcomes.

According to this study, a number of various features, including the nature of the parent-child relationship and the distance between home and school, among others, were positively correlated with students' academic performance. Educators can pay attention to these features of their students and take proactive steps to ensure that students who need more help will be found and assisted by discovering the target variable most strongly associated with high student grade performance.

C. Problem Statement

Education is an essential aspect of a country's long-term economic growth and success. Proficiency in core subjects, particularly in Mathematics, should be emphasized as it is an essential prerequisite for success, specifically in modern society. The problem to be addressed here is the low academic performance in Mathematics by students in the range of 15 to 22 years old in Portugal. Even though the Portuguese educational level has improved, Portugal continues to rank last in Europe because of its high rates of student failure and dropout, as stated by Cortez P. and Silva A. [4]. Study shows that in 2006, 40% of 18- to 24-year-olds in Portugal dropped out of school, while the average for the European Union was only 15% [4]. This reflects the low academic performance in foundational classes, specifically in core subjects such as mathematics, which is very serious and concerning. Furthermore, as stated in the research study by Mata et al. [9], the result of a comparative international evaluation revealed that even though Portuguese students show an improvement from the year 2003 to 2009, on a scale of six levels, almost 25% of Portuguese students still performed at level 2 or below in 2009. The consequences of neglecting academic performance in Mathematics include the lack of development in reasoning ability, problem-solving ability and thinking capacity, which will contribute to one's preparation for the outside world and the welfare of the individual and society [15].

Performing well academically in mathematics is essential in achieving the Sustainable Development Goal (SDG) of SDG 4. Mathematics, which is known as one of the core subjects, is essential in the growth of critical thinking, logical reasoning, and mental rigour. These traits are required for individuals to become informed, engaged citizens who can contribute to sustainable development. It is also worth mentioning that the learning objectives defined by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) have also clearly stated the importance of mathematics in reaching SDG 4 of Quality Education.

D. Research Questions

Here are some relevant research questions that we have come up with in the process of organizing this project:

Research Question 1: "Is there a correlation between family background and student performance in Mathematics in Portugal?"

Research Question 2: "What are the effects of family-related factors on students' academic performance?"

Research Question 3: "What is the primary factor that has the largest impact on students' academic performance in Mathematics in Portugal?"

E. Hypotheses

Having prior insights into students' academic performance in individual courses is a necessity. This information is vital in aiding students facing potential academic challenges, facilitating their educational progress, and supporting their journey toward excellence. A student's scholastic achievements, especially during higher secondary education, significantly influence various aspects, such as obtaining scholarships for university, making informed choices about college subjects, and establishing a foundation for future career stability and financial independence. Nevertheless, every student encounters phases in life where external factors contribute to declining academic performance. Therefore, we assume that there is a direct relationship between a family background that contributes to students' academic performance in Mathematics. In this project, we will utilize the Random Forest Regression and Decision Tree algorithms to explore the relationships and predict students' academic performance in Mathematics in Portugal. Using these algorithms, we aim to gain insights into the factors affecting student performance and develop models for predicting educational outcomes.

F. Research Objectives

Therefore, this study aims to examine the factors affecting students' academic performance in Mathematics in Portugal. The objective for easy analysis of these factors is a further breakdown.

1. To identify whether there is a direct relationship between family background and students' academic performance.
2. To assess the effects of family-related factors on students' academic performance.
3. To identify and determine the primary factors that impact students' academic performance in Mathematics in Portugal, considering their family background.

G. Research Significances

The significance of conducting this research on the relationship between family background and student performance in Mathematics in Portugal lies in its potential to provide valuable insights and practical implications for educators, policymakers, and other stakeholders in the education system.

a) **Development of Predictive Models:** One of the critical contributions of this research is the development of predictive models that utilize data science techniques to forecast students' academic performance in Mathematics based on their family background factors. By leveraging advanced algorithms and machine learning, these models can provide valuable insights into the relationship between family background and academic outcomes. The application of data science in education enables educators and policymakers to make data-driven decisions, identify students who are at risk of underperforming, and implement targeted interventions to address their specific learning needs. Integrating data science and education has immense potential to improve educational outcomes and foster personalized learning approaches.

b) **Identifying Factors Influencing Academic Performance:** By examining the correlation between family background and student performance, this research can help identify the key factors that significantly impact academic achievement in Mathematics. Understanding these factors is crucial for developing targeted interventions and support systems to improve student outcomes.

c) **Enhancing Parental Involvement:** Recognizing the impact of family factors on student performance can emphasize the importance of parental involvement in education. Parents and guardians can be encouraged to participate actively in their child's learning process, provide support at home, and engage in collaborative efforts with educators to optimize student outcomes.

II. RELATED WORKS

Numerous studies and research projects have been conducted to examine how well students perform in mathematics. We chose the one that seemed to be closest to our project and shared some of our objectives. The table below provides a summary of them.

TABLE I. LITERATURE REVIEWS

No.	Year	Authors	Research Problem	Main Techniques Applied	Results
1	2010	Yara, P. O., & Otieno, K. O. [16]	Teaching/learning resources and academic performance in mathematics in secondary schools in Bondo District of Kenya.	Ridge regression	23.6% of the total variance in academic performance in mathematics could be accounted for by the eight independent variables when taken together
2	2018	Macdonald, K., Milne, N., Orr, R., & Pope, R. [8]	Relationships between motor proficiency and academic performance in	Critical Appraisal of Methodological Quality, CASP (tool)	The effect of motor skill interventions on academic performanc

No.	Year	Authors	Research Problem	Main Techniques Applied	Results
			mathematics and reading in school-aged children and adolescents: A systematic review.		e warrants further investigation. Future works: Should consider consistently using valid and reliable, standardized instruments to assess both fine and gross motor proficiency and academic performance variables.
3	2017	Tanujaya, B., Mumu, J., & Margono, G. [13]	The Relationship between Higher Order Thinking Skills and Academic Performance of Student in Mathematics Instruction.	Correlation Coefficient	Students who have high higher order thinking skills (HOTS) tend to get high GPA in mathematics instruction, whereas those with low HOTS tend to have low GPA.
4	2018	Adamma, O. N., Ekwutosim, O. P., & Unamba, E. C. [1]	Influence of Extrinsic and Intrinsic Motivation on Pupils Academic Performance in Mathematics.	Correlation Coefficient	Female college students are more likely to have higher academic ethics than male students, which is characterized by higher academic attainment.
5	2019	Ozkal, N. [10]	Relationships between self-efficacy beliefs, engagement and academic performance in math lessons.	Multiple linear regression	Secondary school sixth, seventh and eighth graders' self-efficacy beliefs for Math learning and performance were found to predict Math achievement positively and significantly.
6	2017	Hartanto, A., Toh, W. X., & Yang, H. [7]	The different implications of weekday and weekend video gaming for academic performance in mathematics, reading, and science.	Markov Chain	Negative relationship between weekday video gaming and academic performance.
7	2015	Bravo-Sanzana, M. V., Salvo-Garrido, S., & Muñoz, C. [3]	Profiles of Chilean students according to academic performance in mathematics.	Classification and regression tree (CART) and random forest.	The type of school and the index of mathematical abilities influence good performance.

No.	Year	Authors	Research Problem	Main Techniques Applied	Results
8	2021	Aksan, J. A. [2]	Effect of modular distance learning approach to academic performance in mathematics of students in Mindanao State University-Sulu Senior High School amidst COVID-19 pandemic.	ANOVA (tool) and correlation coefficient.	The effectiveness of modular distance learning approach in learning Math despite of its challenges amidst COVID-19 pandemic
9	2020	Trigueros, R., Aguilar-Parra, J. M., Lopez-Liria, R., Cangas, A. J., González, J. J., & Álvarez, J. F. [14]	The role of perception of support in the classroom on the student's motivation and emotions: The impact on metacognition strategies and academic performance in math and English classes.	Structural equation model (SEM)	The metacognition strategy positively predicted academic achievement.
10	2021	Spitzer, M. W. H., & Musslick, S. [12]	Academic performance of K-12 students in an online learning environment for mathematics increased during the shutdown of schools in the wake of the COVID-19 pandemic.	Linear mixed model.	The shutdown of schools in 2020 had a positive impact on the performance of students in an online learning environment for mathematics relative to the year before.

The ten previous works make it evident that several techniques were applied. Each study or research has a unique collection of chosen models or analytical tools. The methods employed in the various research papers also offer distinct strengths and limitations. Yara's [16] use of Ridge regression showcases its robustness in handling multicollinearity in complex educational datasets, yet it relies on the assumption of linear relationships among variables. In contrast, Macdonald et al.'s [8] critical appraisal tool ensures the inclusion of high-quality studies in systematic reviews, but it doesn't generate new empirical data.

Meanwhile, correlation coefficients, as used by Tanujaya et al. [13] and Adamma et al. [1], quantify relationships but cannot establish causation and may be subject to bias due to self-reported data. Multiple linear regression, adopted by Özkal [10], accounts for multiple predictor variables but assumes linearity and may not capture nuanced nonlinear relationships. Hartanto et al.’s [7] use of Markov Chain analysis is apt for modelling sequential data but may not generalize well across all types of video gaming, while Bravo-Sanzana et al. [3] Classification and Regression Trees and Random Forests capture complex interactions but risk overfitting. Aksan's [2] use of ANOVA and correlation coefficients offers group comparisons but may not control for all variables during a pandemic. Lastly, Structural Equation Modeling, as employed by Trigueros et al. [14], captures complex relationships among latent variables but necessitates a substantial sample size and may be sensitive to model misspecification. Linear Mixed Models, used by Spitzer and Musslick [12], handle repeated measures and temporal changes but require statistical expertise.

However, there is currently not much data showing that family background affects student performance. Mostly, they were about teaching and learning techniques. The main differentiator between our project and theirs is that. The algorithms for this project that our team has chosen to use are Random Forest Regression and Decision Trees. Additionally, we added the training and testing processes utilizing the developed model, using 80% and 20% of the input data, respectively.

III. METHODOLOGY

A. Data Description

The dataset approaches secondary school student achievement in two Portuguese schools. The dataset was compiled using school reports and surveys. We used open-source data from Kaggle Datasets. The dataset contains 395 rows and 33 columns. Data attributes include student grades and demographic, social, and school-related features, where the G3 (final grade) is the output target. Below is the sample of the dataset:

school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob
GP	F	18	U	GT3	A	4	4	at_home	teacher
GP	F	17	U	GT3	T	1	1	at_home	other
GP	F	15	U	LE3	T	1	1	at_home	other
GP	F	15	U	GT3	T	4	2	health	services
GP	F	16	U	GT3	T	3	3	other	other
GP	M	16	U	LE3	T	4	3	services	other
GP	M	16	U	LE3	T	2	2	other	other
GP	F	17	U	GT3	A	4	4	other	teacher
GP	M	15	U	LE3	A	3	2	services	other
GP	M	15	U	GT3	T	3	4	other	other
GP	F	15	U	GT3	T	4	4	teacher	health
GP	F	15	U	GT3	T	2	1	services	other
GP	M	15	U	LE3	T	4	4	health	services
GP	M	15	U	GT3	T	4	3	teacher	other

Fig. 1. Sample dataset (Part 1)

reason	guardian	traveltime	studytime	failures	schoolsup	famsup	paid	activities	nursery
course	mother	2	2	0	yes	no	no	no	yes
course	father	1	2	0	no	yes	no	no	no
other	mother	1	2	3	yes	no	yes	no	yes
home	mother	1	3	0	no	yes	yes	yes	yes
home	father	1	2	0	no	yes	yes	no	yes
reputatio	mother	1	2	0	no	yes	yes	yes	yes
home	mother	1	2	0	no	no	no	no	yes
home	mother	2	2	0	yes	yes	no	no	yes
home	mother	1	2	0	no	yes	yes	no	yes
home	mother	1	2	0	no	yes	yes	yes	yes
reputatio	mother	1	2	0	no	yes	yes	no	yes
reputatio	father	3	3	0	no	yes	no	yes	yes
course	father	1	1	0	no	yes	yes	yes	yes
course	mother	2	2	0	no	yes	yes	no	yes

Fig. 2. Sample dataset (Part 2)

higher	internet	romantic	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
yes	no	no	4	3	4	1	1	3	6	5	6	6
yes	yes	no	5	3	3	1	1	3	4	5	5	6
yes	yes	no	4	3	2	2	3	3	10	7	8	10
yes	yes	yes	3	2	2	1	1	5	2	15	14	15
yes	no	no	4	3	2	1	2	5	4	6	10	10
yes	yes	no	5	4	2	1	2	5	10	15	15	15
yes	yes	no	4	4	4	1	1	3	0	12	12	11
yes	no	no	4	1	4	1	1	1	6	6	5	6
yes	yes	no	4	2	2	1	1	1	0	16	18	19
yes	yes	no	5	5	1	1	1	5	0	14	15	15
yes	yes	no	3	3	3	1	2	2	0	10	8	9
yes	yes	no	5	2	2	1	1	4	4	10	12	12
yes	yes	no	4	3	3	1	3	5	2	14	14	14
yes	yes	no	5	4	3	1	2	3	2	10	10	11

Fig. 3. Sample dataset (Part 3)

B. *Tools Used*

a) Google Colab: Colab, alternatively recognized as Google Colaboratory, is a complimentary web-based utility provided by Google. This platform empowers individuals to generate and execute Python code within the framework of a Jupyter Notebook environment. Its purpose is to simplify collaborative efforts on data analysis and research. This is achieved by facilitating concurrent contributions from multiple team members on a shared endeavour. Access to Colab's powerful computing resources, such as GPUs and TPUs, can be used to speed up computations like data processing or machine learning training. It also comes with pre-installed libraries and data science tools like TensorFlow, NumPy, and Pandas.

b) TensorFlow: TensorFlow stands as an open-source machine learning framework by Google. Google is responsible for developing the open-source machine learning framework recognized as TensorFlow. TensorFlow is fundamentally a software framework for numerical computing that enables programmers to construct and tune machine learning models. It is made to deal with vast amounts of data and complex mathematical operations, such as gradients, convolutions, and matrix multiplications, which are frequently employed in deep learning algorithms. Utilizing TensorFlow, developers can construct models catering to a diverse array of applications, spanning from speech and image identification to natural language processing and anticipatory data analysis.

c) Numpy: NumPy is a Python library for numerical computing that supports large, multi-dimensional arrays and matrices and a collection of mathematical functions for performing operations on these arrays. Applications for scientific computing, data analysis, and machine learning frequently use NumPy. NumPy arrays offer a practical way to carry out mathematical operations on large datasets and are tuned for quick computations.

d) Pandas: Pandas is a free, open-source Python library for analyzing and manipulating data. It includes tools for reading and writing data to and from various sources, including CSV files, Excel spreadsheets, and SQL databases. It also offers data structures for effectively storing and manipulating large datasets. The DataFrame, a two-dimensional table-like structure with named rows and columns, is the fundamental data structure in Pandas. Moreover, Pandas provides extensive data analysis and manipulation tools, including filtering, sorting, grouping, joining, and reshaping. These operations may be easily performed on large datasets because they are created to be fast and efficient.

e) Scikit-Learn: Scikit-Learn is an open-source Python library dedicated to machine learning, offering various resources for constructing and implementing machine learning models. Its repertoire encompasses a variety of techniques, including decision trees, random forests, support vector machines, k-nearest neighbours, and neural networks, among others. It also supports model selection, dimension reduction, feature extraction, and evaluation. The emphasis on usability and consistency by Scikit-Learn is one of its benefits. Even for users unfamiliar with the field, it provides a clear and intelligible API that makes it simple to start with machine learning.

C. *Algorithms Implemented*

For this research project on the relationship between family background and student performance in Mathematics in Portugal, we will implement the following machine learning algorithms:

a) Random Forest Regression: Random Forest is a robust ensemble learning algorithm that combines multiple decision trees to make predictions [5]. In this project, we will utilize the Random Forest Regression algorithm to analyze the dataset and identify the primary factors that significantly impact students' academic performance in Mathematics. This algorithm can handle complex relationships and provide accurate predictions by leveraging the collective knowledge of multiple decision trees.

b) Decision Trees: Decision Trees are a simple yet effective machine learning algorithm that uses a tree-like model of decisions and their possible consequences. Decision Trees can be used for classification and regression tasks [11]. In this project, we will employ Decision Trees to analyze the dataset and explore the relationships between family background and student performance in Mathematics. By visualizing and interpreting the decision tree model, we can gain insights into the factors influencing academic performance.

These selected algorithms offer different strengths and approaches to analyzing the dataset and predicting academic outcomes. Random Forest Regression provides robust predictions by leveraging the collective knowledge of multiple decision trees. Decision Trees offer interpretability, allowing us to understand the factors driving student performance. By combining these algorithms, we can gain comprehensive insights into the factors affecting academic performance in Mathematics in Portugal based on family background.

D. *Training and Testing Dataset*

The dataset for this research project will be divided into training and testing sets using an 80:20 ratio. The training set, comprising 80% of the dataset, will be used to train the machine learning algorithms. This set contains the "seen" data, allowing the algorithms to learn the patterns and relationships between family background and student performance. The remaining 20% of the dataset will be allocated to the testing set containing the "unseen" data. This set will be used to evaluate the performance and generalization capability of the trained algorithms by assessing their ability to make accurate predictions on new, unseen instances.

E. *Model Diagram*

The model flow can be summarized in Figure 4 below:

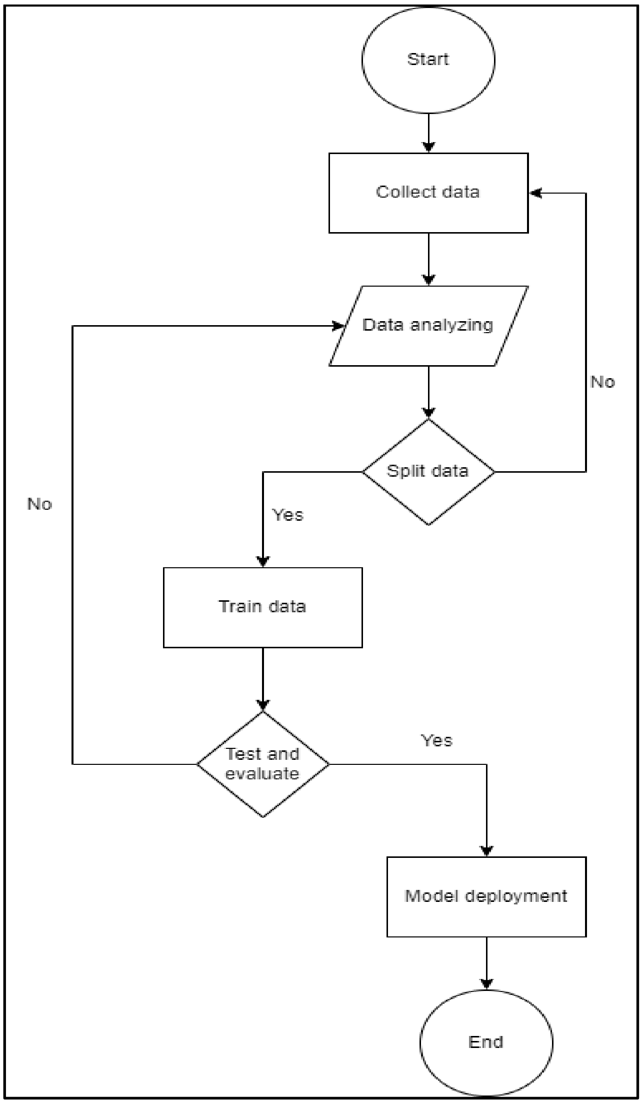

Fig. 4. Model flowchart

In our model flow, we start by obtaining the dataset from Kaggle. We then perform data analysis to gain insights and understand the dataset. After that, we split the data into 80% training and 20% testing sets. Once the data is split, we proceed to train our model using the training data and evaluate its performance using evaluation metrics such as MSE, RMSE, and MAE. If the model performs well, we proceed with model deployment. However, if the performance is not satisfactory, we go back to the data analysis stage to investigate further and potentially improve our approach. This iterative process helps us develop an effective and accurate model for our task.

F. *Data Cleaning*

Data cleaning is a crucial step in this research project to ensure the accuracy and reliability of the input datasets. Various techniques address potential issues such as missing or duplicate entries, incorrect values, and improperly formatted data. By carefully examining the data, we identify and handle missing values, remove duplicate entries, and correct any erroneous or inconsistent ones. A feature selection process is also conducted to determine the most relevant and informative variables before proceeding with Exploratory Data Analysis (EDA). This data-cleaning process aims to enhance the dataset's quality, ensuring it is suitable for subsequent analysis and modelling.

a) Column Re-labeling: In this project, column re-labeling is conducted to improve the clarity and interpretability of the dataset. By re-labelling the columns, the variables become more descriptive and aligned with the research context, making it easier for researchers and stakeholders to understand the data and its corresponding features. This process facilitates practical data analysis and enhances the overall comprehension of the dataset, ensuring that the variables are appropriately labelled and understandable to all parties involved in the research project. The original columns' names are as follows:

```
Index(['school', 'sex', 'age', 'address', 'famsize', 'Pstatus', 'Medu', 'Fedu',
       'Mjob', 'Fjob', 'reason', 'guardian', 'traveltime', 'studytime',
       'failures', 'schoolsup', 'famsup', 'paid', 'activities', 'nursery',
       'higher', 'internet', 'romantic', 'famrel', 'freetime', 'goout', 'Dalc',
       'Walc', 'health', 'absences', 'G1', 'G2', 'G3'],
      dtype='object')
```

Fig. 5. Original columns' name

After renaming the columns, the dataset reflects the updated column names that represent the variables. The following are the updated column names:

```
Index(['School', 'Gender', 'Age', 'LivingArea', 'FamilySize', 'ParentStatus',
       'MotherEdu', 'FatherEdu', 'MotherJob', 'FatherJob', 'Reason',
       'Guardian', 'SchoolTravelTime', 'StudyTime', 'Failures',
       'SchoolSupport', 'FamilySupport', 'ExtraPaidClasses', 'CocuActivities',
       'NurserySchool', 'DesireHigher', 'InternetAccess', 'RomanticRelay',
       'FamilyRelay', 'FreeTime', 'HangOut', 'DayAlcohol', 'WeekAlcohol',
       'HealthStatus', 'SchoolAbsences', 'G1', 'G2', 'G3'],
      dtype='object')
```

Fig. 6. Updated columns' name

b) Data Standardization: This process ensures consistency and uniformity in representing categorical variables within the dataset. Replacing values with appropriate words and capitalizing the first letter makes interpreting and analyzing the data easier.

```
# Replace values in some columns
clean_data['LivingArea'] = clean_data['LivingArea'].replace({'U': 'Urban', 'R': 'Rural'})
clean_data['FamilySize'] = clean_data['FamilySize'].replace({'GT3': '>3', 'LE3': '<=3'})
clean_data['ParentStatus'] = clean_data['ParentStatus'].replace({'A': 'Apart', 'T': 'Together'})

# Capitalize the initial letter in all columns
clean_data = clean_data.applymap(lambda x: x.capitalize() if isinstance(x, str) else x)
```

Fig. 7. Data standardization

c) Feature Selection: In this project, feature selection is performed to identify the most relevant and informative variables for predictive research while excluding unnecessary features. Based on the dataset, the total number of columns before selection is 33. The figure below shows the details of 33 columns in our datasets:

#	Column	Non-Null Count	Dtype
0	School	395 non-null	object
1	Gender	395 non-null	object
2	Age	395 non-null	int64
3	LivingArea	395 non-null	object
4	FamilySize	395 non-null	object
5	ParentStatus	395 non-null	object
6	MotherEdu	395 non-null	int64
7	FatherEdu	395 non-null	int64
8	MotherJob	395 non-null	object
9	FatherJob	395 non-null	object
10	Reason	395 non-null	object
11	Guardian	395 non-null	object
12	SchoolTravelTime	395 non-null	int64
13	StudyTime	395 non-null	int64
14	Failures	395 non-null	int64
15	SchoolSupport	395 non-null	object
16	FamilySupport	395 non-null	object
17	ExtraPaidClasses	395 non-null	object
18	CocuActivities	395 non-null	object
19	NurserySchool	395 non-null	object
20	DesireHigher	395 non-null	object
21	InternetAccess	395 non-null	object
22	RomanticRelay	395 non-null	object
23	FamilyRelay	395 non-null	int64
24	FreeTime	395 non-null	int64
25	HangOut	395 non-null	int64
26	DayAlcohol	395 non-null	int64
27	WeekAlcohol	395 non-null	int64
28	HealthStatus	395 non-null	int64
29	SchoolAbsences	395 non-null	int64
30	G1	395 non-null	int64
31	G2	395 non-null	int64
32	G3	395 non-null	int64

dtypes: int64(16), object(17)

Fig. 8. Original columns

After careful consideration, the following features, School, Gender, Age, Reason, Failures, and DesireHigher, were dropped from the analysis as they were deemed less influential for the research objectives.

After dropping unnecessary columns, the total number of columns left is 27. The features selected for analysis are as follows:

#	Column	Non-Null Count	Dtype
0	LivingArea	395 non-null	object
1	FamilySize	395 non-null	object
2	ParentStatus	395 non-null	object
3	MotherEdu	395 non-null	int64
4	FatherEdu	395 non-null	int64
5	MotherJob	395 non-null	object
6	FatherJob	395 non-null	object
7	Guardian	395 non-null	object
8	SchoolTravelTime	395 non-null	int64
9	StudyTime	395 non-null	int64
10	SchoolSupport	395 non-null	object
11	FamilySupport	395 non-null	object
12	ExtraPaidClasses	395 non-null	object
13	CocuActivities	395 non-null	object
14	NurserySchool	395 non-null	object
15	InternetAccess	395 non-null	object
16	RomanticRelay	395 non-null	object
17	FamilyRelay	395 non-null	int64
18	FreeTime	395 non-null	int64
19	HangOut	395 non-null	int64
20	DayAlcohol	395 non-null	int64
21	WeekAlcohol	395 non-null	int64
22	HealthStatus	395 non-null	int64
23	SchoolAbsences	395 non-null	int64
24	G1	395 non-null	int64
25	G2	395 non-null	int64
26	G3	395 non-null	int64

dtypes: int64(14), object(13)

Fig. 9. Columns after feature selection

By focusing on the selected features, we aim to better understand the factors concerning family background that impact students' academic performance in Mathematics in Portugal.

d) Exploratory Data Analysis (EDA): In this project, Exploratory Data Analysis (EDA) is explicitly performed for outliers treatment and data visualization to detect and handle outliers and to visually explore and understand the dataset's characteristics, patterns, and relationships.

1) Outliers Treatment: In this project, box plots were utilized to visually examine the presence of outliers in the dataset. Identifying and treating outliers are essential for ensuring the integrity and reliability of subsequent analyses and modelling processes.

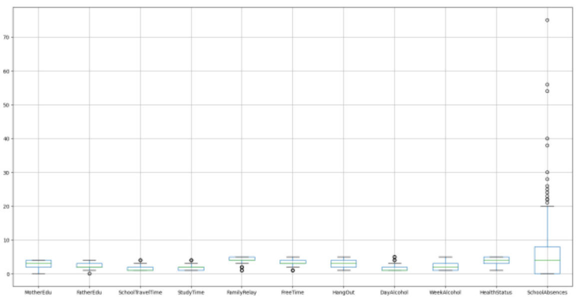

Fig. 10. Outliers

Outliers were found in the following columns: FatherEdu, SchoolTravelTime, StudyTime, FamilyRelay, FreeTime, DayAlcohol, and SchoolAbsences.

To address the outliers identified in the variables FatherEdu, SchoolTravelTime, StudyTime, FamilyRelay, FreeTime, DayAlcohol, and SchoolAbsences, the Interquartile Range (IQR) approach was employed in this project to remove these extreme values. The IQR approach is a commonly used method for outlier treatment in data analysis. It involves calculating the range between a variable's distribution's first quartile (Q1) and the third quartile (Q3). An outlier is defined as any data point that lies below Q1 minus 1.5 times the interquartile range (IQR) or exceeds Q3 plus 1.5 times the IQR, and these values can be excluded from the dataset.

After applying the IQR approach to remove outliers, the updated box plots reflect a distribution free from the influence of extreme values, providing a more precise visualization of each variable's central tendencies and variability.

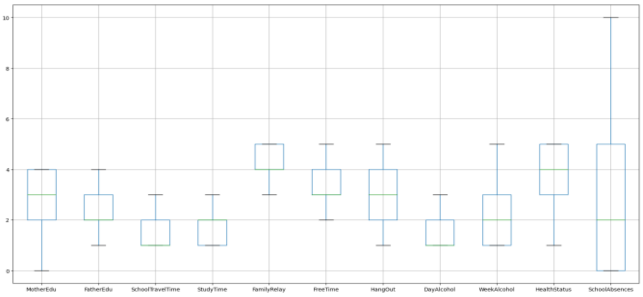

Fig. 11. Outliers treatment

2) Data Visualization: In this project, box plots were utilized to visually examine the presence of outliers in the dataset. Identifying and treating outliers are essential for ensuring the integrity and reliability of subsequent analyses and modelling processes.

a) Grades Distribution in Dataset: As discussed earlier, students' performance in Mathematics is graded based on three sets of examinations, and their grades in those examinations are represented as G1, G2, and G3. From the grades, we have calculated the average of the grades and visualized them using a bar plot. Our objective is to make the average of the grades suitable for further analysis, including using machine learning modelling as the target variable to measure the students' performance in mathematics. Below are the bar plot visualizations of the distribution of grades in these three examinations and grades' averages.

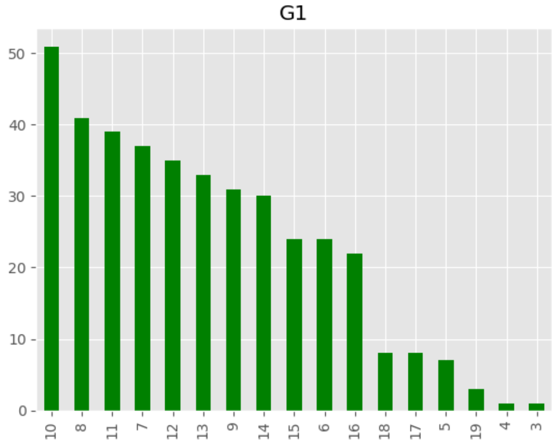

Fig. 12. G1 distribution

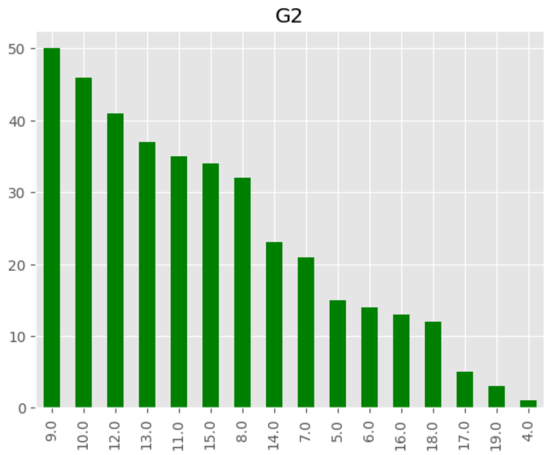

Fig. 13. G2 distribution

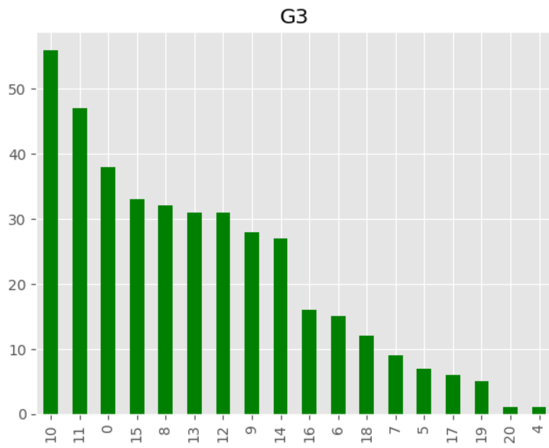

Fig. 14. G3 distribution

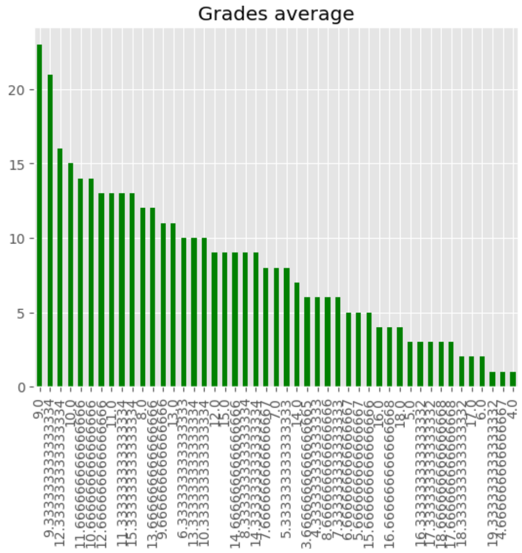

Fig. 15. Grades average distribution

b) Family Background-Related Features Distribution in Dataset: To align with our research objectives, the features of interest are highly associated with family background-related features. The chosen set of five features encompasses parental status (united, separated), residential setting (urban, rural), maternal educational attainment (ranging from 0 to 4), paternal educational level (ranging from 1 to 4), and finally, the intensity of family bonds (graded between 3 and 5).

- *Parent Status:* The parent status distribution is shown as a bar plot that distinguishes between parents who are together and those who are not. The y-axis shows the frequency of students falling into each parent status category, and the x-axis lists the parent status categories. The plotted graph provides evidence that our dataset is primarily characterized by the prevalence of students belonging to family units with parents who remain in a conjugal relationship.

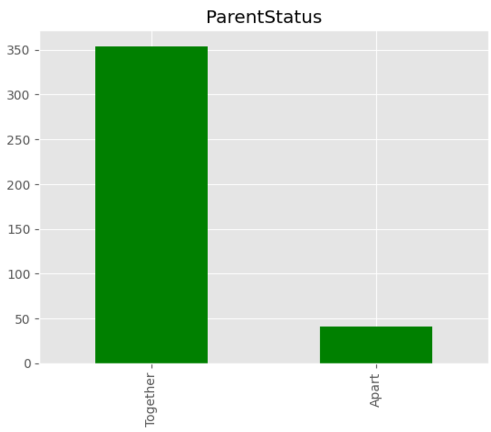

Fig. 16. Parental status distribution

- *Living Area:* The bar plot below illustrates the distribution of family residential areas, highlighting rural and urban environments. Each bar represents a living area category and shows the number of people who live in an urban or rural area. Based on the bar plot below, most students in our dataset reside in urban areas. This insight emphasizes how urban-centric the dataset is and how prevalent students from urban backgrounds are.

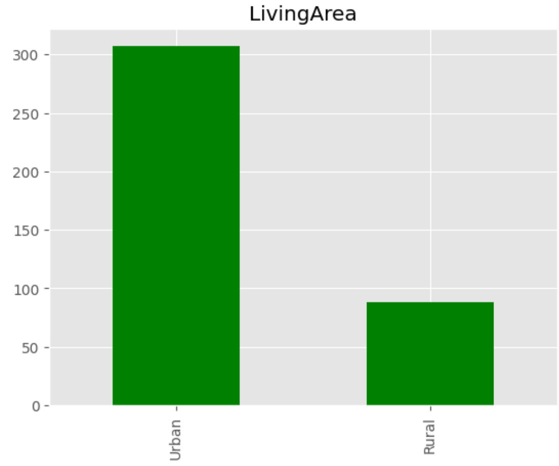

Fig. 17. Family living area distribution

- *Parent Education Level:* In the bar chart, every education level category is depicted as a bar, illustrating how educational levels are distributed among mothers and fathers of students separately. By comparing the educational backgrounds of mothers and fathers, this visualization illustrates how prevalent different educational levels are in the dataset.

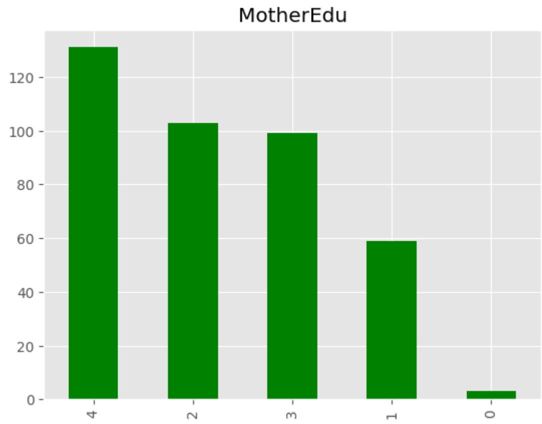

Fig. 18. Mother educational level distribution

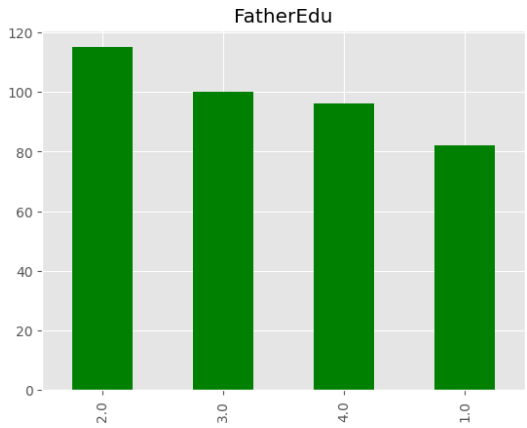

Fig. 19. Father educational level distribution

- *Family Relationship:* An intriguing pattern may be seen by carefully examining the bar plot below, showing our dataset's distribution of family relationship scores. It becomes clear that students with a level 4 (moderate) family relationship score have a noticeable dominance. The fact that these students make up a substantial portion of the dataset suggests that our dataset population has a high prevalence of moderate family relationships.

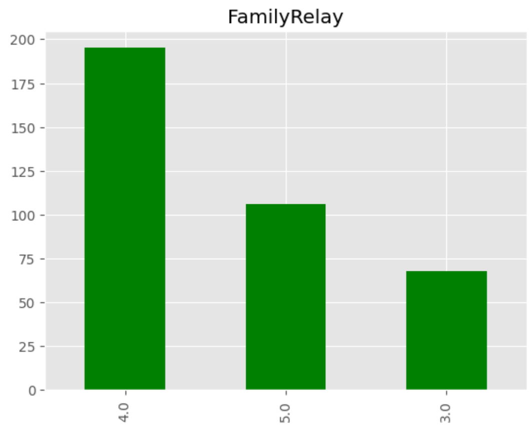

Fig. 20. Family relationship strength distribution

IV. EXPERIMENTS

A. Model Development

Two machine learning algorithms, Random Forest and Decision Tree models, were employed to analyze the dataset. We conducted a series of rigorous experiments to assess the effectiveness of our proposed model. Our dataset was divided into two sets: the training set, which comprised 80% of the data, and the testing set, which contained the remaining 20%. This division allowed us to train our machine learning algorithms on a substantial portion of the data, ensuring they learned the underlying patterns and relationships within the dataset. Subsequently, we evaluated the performance of our model on the testing set, which represented "unseen" data, providing a robust assessment of its generalization capabilities.

B. Evaluation Metrics

Based on our model's performance measurements, we will evaluate the models using three metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These metrics provide a quantitative assessment of the model's accuracy in predicting the target variable. We can assess the models' performance by analyzing these metrics and determine which one provides better predictions [11].

a) Mean Squared Error (MSE): The Mean Squared Error (MSE) stands as a prevalent metric employed for assessing regression models, quantifying the average magnitude of squared variances between the predicted values and the ground truth. It gives higher weights to more significant errors, making it helpful in understanding the overall model performance. A lower MSE indicates better model performance, representing more minor prediction errors. Below is how the MSE is being calculated:

$$(1/n) * \Sigma(y\_true - y\_pred)^2 \quad (1)$$

Where:
- n is the number of data points
- Σ denotes the summation symbol

b) Root Mean Squared Error (RMSE)): The square root of the Mean Squared Error (MSE) is represented by RMSE, which offers a comprehensible metric for gauging the typical magnitude of forecast inaccuracies. It is in the same unit as the target variable and is commonly used to assess the quality of regression models. Like MSE, a lower RMSE signifies better model performance, indicating more minor average prediction errors. The formula for RMSE is:

$$\sqrt{} (1/n) * \Sigma(y\_true - y\_pred)^2 \quad (2)$$

c) Mean Absolute Error (MAE): The Mean Absolute Error (MAE) quantifies the average absolute variance between the predicted and actual values. This metric offers a more precise and intuitive comprehension of prediction inaccuracies, indicating the average extent of errors irrespective of their orientation. Similar to MSE and RMSE, a lower MAE indicates better model performance, signifying more minor average prediction errors. Below is the formula of MAE:

$$(1/n) * \Sigma|y\_true - y\_pred| \quad (3)$$

Where:
- n is the number of data points.
- Σ denotes the summation symbol.
- | | denotes the absolute value.

These metrics allow us to quantify and compare the performance of different regression models. By examining the values of MSE, RMSE, and MAE, we can determine which model provides more accurate predictions and better overall performance.

V. RESULT AND DISCUSSION

A. Correlation between family background and student performance in Mathematics in Portugal

A correlation map is plotted to satisfy this objective or research question. From the correlation map plotted, we will focus on the top three family family-background-related correlated features with the highest correlation values. Below is the correlation map.

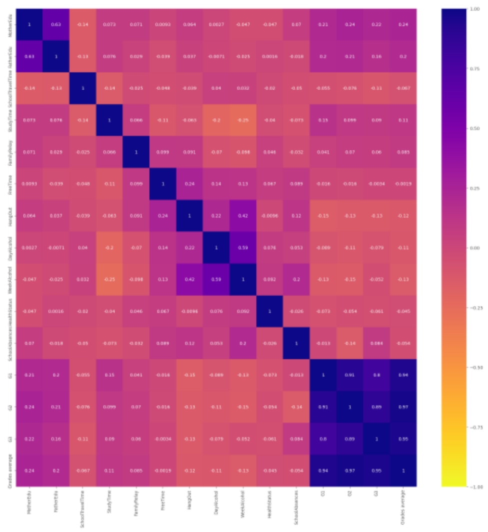

Fig. 21. Correlation map

TABLE II. TOP-THREE FAMILY-BACKGROUND RELATED CORRELATED FEATURES

Feature Names	Description	Correlation Values
MotherEdu	Mother educational level	0.240
FatherEdu	Father educational level	0.200
FamilyRelay	Family relationship strength	0.085

The features above can be further visualized and analyzed for a better comprehensive understanding of their impact on the grade average of the mathematics examinations as the target variable.

- Parent Educational Level versus Grades Average

Two box plots are being constructed to visualize the relationship between parents' educational level and students' grade averages. Below is the visualization plotted:

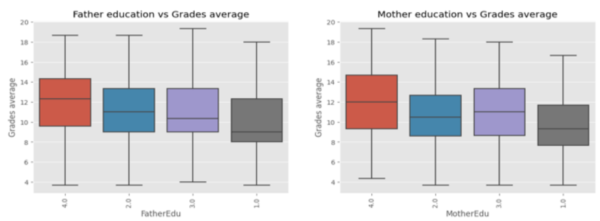

Fig. 22. Box plots of parents' educational level versus grade average

As these box plots are examined, a clear pattern emerges. Students with parents who possess a level 4 education have the highest average marks across all educational levels. This finding suggests a positive correlation between higher academic performance as represented in the average grades and higher levels of parental education, particularly level 4.

- Family Relationship Strength versus Grades Average

Below is the bar plot plotted to exhibit the relationship between family relationship strength (scale: 3 - 5) and students' grades average:

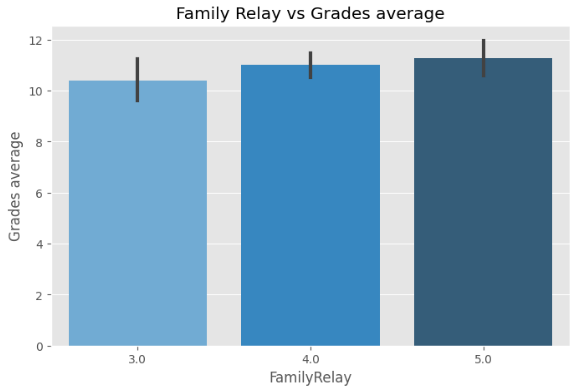

Fig. 23. Bar plot of family relationship strength versus grades average

The notable pattern can be seen when the bar plot is further examined. Notably, although the distribution chart depicting family relationships (FamilyRelay) portrays a prevalence of students with intermediate family connections (level 4), it is evident that those students who possess a family relationship level of 5 – indicating the most robust and flourishing family bonds – display markedly superior academic performance, as reflected by their elevated average grades. This finding strongly implies a positive association between the strength of family relationships and academic achievement among students.

B. The Effects of Family-related Factors on Student's Academic Performance

The outcomes of our analysis provided insight into the effects of family-related factors on students' academic performance in Mathematics. To begin with, there is a significant correlation between parents' educational level and students' grade point averages. Greater levels of parental education exhibit a favourable correlation with improved academic performance, exemplified by the observation that students whose parents possess an education level of 4 attain the highest average grades across all subjects. Second, examining the relationship between family relationship strength and grade average highlights the significance of strong familial bonds [11]. On average, students who continuously obtain noticeably high grades have family relationship levels of 5, which signifies the healthiest and strongest family relationship. This shows a connection between higher performance in school and healthier familial relationships. These findings highlight the critical role that family-related factors play in determining students' academic performances, particularly in Mathematics, highlighting the importance of parental education levels and the quality of family relationships as significant factors in students' academic performance.

C. The Primary Factor that Has the Largest Impact on Student's Academic Performance in Mathematics in Portugal

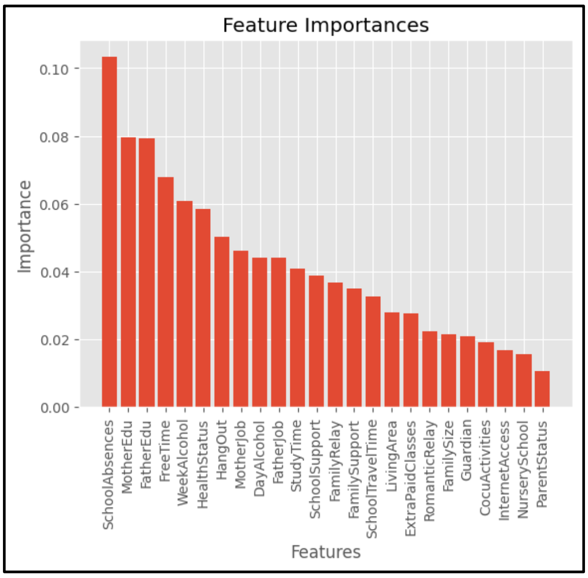

Fig. 24. Feature importances by Random Forest

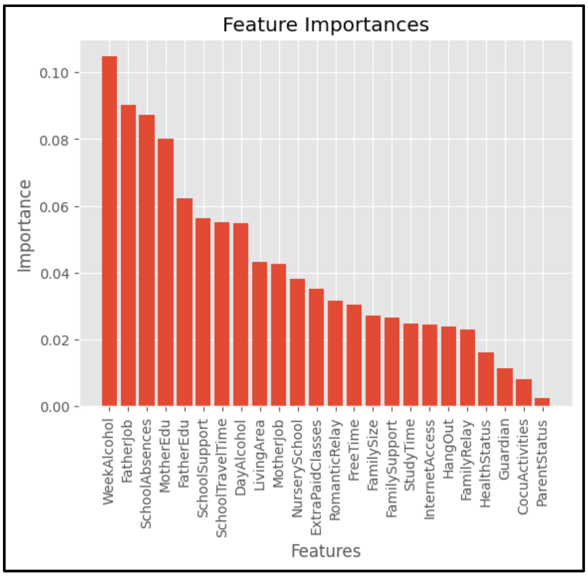

Fig. 25. Feature importances by Decision Trees

To determine the primary factor that has the most significant impact on students' academic performance in Mathematics in Portugal, we can analyze the feature importance values obtained from the Random Forest and Decision Tree models. Comparing the feature importance values between the two models can provide more robust insights.

Based on the feature importance values from both the Random Forest and Decision Tree models, it can be concluded that "SchoolAbsences" is the primary factor that has the largest impact on students' academic performance in Mathematics in Portugal. Both models consistently highlight the importance of this factor.

"SchoolAbsences" refers to the number of absences a student has from school. It indicates how often a student is absent from their classes. High absenteeism can have a negative impact on student's academic performance as it can lead to missed learning opportunities, difficulty in keeping up with the curriculum, and gaps in knowledge. Therefore, the importance of the feature suggests that the frequency of school absences plays a significant role in determining students' academic performance in mathematics in Portugal.

TABLE III. MODEL'S PERFORMANCE

Model	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)
Random Forest Regression	9.6212	3.0842	2.4060
Decision Tree	17.6782	4.1872	3.1839

Based on these results, we can observe that the Random Forest model outperforms the Decision Tree model in all three evaluation metrics (MSE, RMSE, and MAE) for the project's analysis. The Random Forest model achieves a lower MSE, RMSE, and MAE than the Decision Tree model. This indicates that the Random Forest model has better predictive performance and can make more accurate predictions than the Decision Tree model.

VI. CONCLUSION AND FUTURE WORKS

In conclusion, the analysis's findings shed important light on the relationships between student arithmetic achievement and family background. The results emphasize the importance of parental education, with greater parental education levels being linked to better mathematics grades. This shows that parents with more education may have the skills and resources necessary to assist their kids' learning successfully. Positive familial ties also play an influence on a student's mathematical achievement. Nurturing and supportive family settings can help students be more motivated and engaged in their studies, improving their academic performance.

The study also emphasizes the significance of regular attendance and engaged involvement in class. The main factor affecting students' academic performance in mathematics in Portugal is absences from school. Regular attendance is essential to guarantee children receive continuous education and learning opportunities. These findings have implications for educators, policymakers, and parents, highlighting the importance of parental involvement, creating supportive family environments, and promoting regular attendance to enhance students' mathematics performance.

More research is needed to build on the findings of this study. Future studies should include different nations in the research, such as Malaysia, to extend an understanding of the factors impacting student academic success. Comparing different educational systems and cultural situations will reveal if the found relationships are universal or context-specific. Furthermore, studying courses other than mathematics will offer a more complete picture of the influence of familial background on students' overall educational performance. Researchers can acquire a more detailed knowledge of the precise aspects contributing to student achievement by considering many disciplines.

Furthermore, future research should identify successful approaches for improving student academic performance. Finding evidence-based techniques, such as parental education programs or innovative instructional approaches, will allow educators, policymakers, and parents to undertake focused interventions to improve student learning outcomes. Researchers can help improve educational results and ensure students attain their full potential in many academic courses by addressing the identified issues with practical solutions.

REFERENCES

[1] Adamma, O. N., Ekwutosim, O. P., & Unamba, E. C. (2018, July). Influence of Extrinsic and Intrinsic Motivation on Pupils Academic Performance in Mathematics. SJME (Supremum Journal of Mathematics Education), 2(2), 52-59. https://doi.org/10.35706/sjme.v2i2.1322.

[2] Aksan, J. A. (2021). Effect of Modular Distance Learning Approach to Academic Performance in Mathematics of Students in Mindanao State University-Sulu Senior High School Amidst COVID-19 Pandemic. Open Access Indonesia Journal of Social Sciences, 4(4). https://doi.org/10.37275/oaijss.v4i2.64.

[3] Bravo-Sanzana, M. V., Salvo-Garrido, S., & Muñoz, C. (2015, March). Profiles of Chilean students according to academic performance in mathematics: An exploratory study using classification trees and random forests. Studies in Educational Evaluation, 44(1). https://doi.org/10.1016/j.stueduc.2015.01.002.

[4] Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. http://www3.dsi.uminho.pt/pcortez/student.pdf.

[5] Cutler, Adele & Cutler, David & Stevens, John. (2011). Random Forests. 10.1007/978-1-4419-9326-7_5. https://www.researchgate.net/publication/236952762_Random_Forests.

[6] Dr. Heena Rachh. (2023). 10 Factors Affecting Academic Performance Among Higher Secondary School Students. Global Indian International School. Retrieved March 23, 2023, from https://abudhabi.globalindianschool.org/blog-details/factors-affecting-academic-performance.

[7] Hartanto, A., Toh, W. X., & Yang, H. (2017, December). Context counts: The different implications of weekday and weekend video gaming for academic performance in mathematics, reading, and science. Computers & Education, 120(3). https://doi.org/10.1016/j.compedu.2017.12.007.

[8] Macdonald, K., Milne, N., Orr, R., & Pope, R. (2018). Relationships Between Motor Proficiency and Academic Performance in Mathematics and Reading in School-Aged Children and Adolescents: A Systematic Review. International Journal of Environmental Research and Public Health, 15(8), 1603. https://doi.org/10.3390/ijerph15081603.

[9] Mata, M. D. L., Monteiro, V., & Peixoto, F. (2012). Attitudes towards mathematics: Effects of individual, motivational, and social support factors. Child development research, 2012. https://downloads.hindawi.com/archive/2012/876028.pdf.

[10] Özkal, N. (2019, June). Relationships between self-efficacy beliefs, engagement, and academic performance in math lessons. Cypriot Journal of Educational Sciences, 14(2), 190-200. https://doi.org/10.18844/cjes.v14i2.3766.

[11] Plevris, Vagelis & Solorzano, German & Bakas, Nikolaos & Ben Seghier, Mohamed. (2022). Investigation of performance metrics in regression analysis and machine learning-based prediction models. 10.23967/eccomas.2022.155. https://www.researchgate.net/publication/362719708_Investigation_of_performance_metrics_in_regression_analysis_and_machine_learning-based_prediction_models.

[12] Spitzer, M. W. H., & Musslick, S. (2021, August 3). Academic performance of K-12 students in an online-learning environment for mathematics increased during the shutdown of schools in the wake of the COVID-19 pandemic. PLoS One, 16(8), e0255629. https://doi.org/10.1371/journal.pone.0255629.

[13] Tanujaya, B., Mumu, J., & Margono, G. (2017, October). The Relationship between Higher Order Thinking Skills and Academic Performance of Students in Mathematics Instruction. International Education Studies, 10(11), 78. https://doi.org/10.5539/ies.v10n11p78.

[14] Trigueros, R., Aguilar-Parra, J. M., Lopez-Liria, R., Cangas, A. J., González, J. J., & Álvarez, J. F. (2020). The role of perception of support in the classroom on the students' motivation and emotions: The impact on metacognition strategies and academic performance in math and English classes. Frontiers in Psychology, 10, Article 2794. https://doi.org/10.3389/fpsyg.2019.02794.

[15] Valero, P. (2017). Mathematics for all, economic growth, and the making of the citizen-worker. In A political sociology of educational knowledge (pp. 117-132). Routledge.

[16] Yara, P. O., & Otieno, K. O. (2010). Teaching/Learning Resources and Academic Performance in Mathematics in Secondary Schools in Bondo District of Kenya. Asian Social Science, 6(12), 126. https://doi.org/10.5539/ass.v6n12p126.