Solent University

Faculty of Business, Law and Digital Technologies

MSc APPLIED AI AND DATA SCIENCE

2021/22

Mezie Brighton Chukwuemeka

"Enhancing student performance with AI Techniques"

Supervisor: Dr. Shakeel Ahmad

Date of submission: **September 2022**

ACKNOWLEDGMENTS

I want to thank Dr. Shakeel Ahmad, my supervisor, for helping me stay focused on my work and keeps pushing ahead with the project. My dissertation was much aided by his supervision and clarification of my thinking. Finally, I'd want to express my gratitude to everyone who contributed to this study, as well as the members of my family and friends who have encouraged me spiritually.

ABSTRACT

The proposed study creates a framework to help students and teachers in secondary schools achieve greater success via the use of cutting-edge data analysis and AI, specifically machine learning techniques., Existing works have attempted to solve separate classroom challenges related to measuring the attention of the students and marking attendance of the students using all necessarily related variables to determine and enhance student academic performance, but these efforts have not yet combined the crucial aspects into a single system. Therefore, studies have been conducted on the specified topics which include variables such as grades, parental education, parental profession, and study time etc. The system implemented a 33 column dataset with 1044 entries which was further run through Linear regression and Random forest algorithms in the modelling phase to predict future scores of select variables. Testing and implementation of the suggested approach led to the discovery that it significantly outperformed alternatives in terms of student involvement and grade recognition (86.6% vs. 63.4%). The suggested system is able to analyze the impact of these factors on a student's grades and provide suggestions for how they might improve if necessary or maintained at a satisfactory level. This study finds that both students and teachers benefit from the introduction of technological solutions to facilitate routine classroom tasks. Additional benefits include improved future productivity for secondary education institutions thanks to the system's useful outcomes.

LIST OF ABBREVIATION

AR – At Risk

NAR-NOT AT RISK

AI- ARTTIFICIAL INTELLIGENCE

ML- MACHINE LEARNING

SMS- SCHOOL MANAGEMENT SYSTEM

CAI-COMPUTER ASSISTED INSTRUCTION

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENT

CHAPTER ONE

INTRODUCTION AND BACKGROUND

1.0 Introduction

Education is the most important factor in breaking out of poverty and allowing for social advancement. Despite this, there are still millions of children who do not attend school, and even among those who do, not all of them are really learning anything. Over half of the world's children and adolescents do not satisfy the minimal competence criteria in reading or mathematics. This number is staggering. There are significant gaps in educational opportunities and results across different areas of the world, with sub-Saharan Africa and some portions of Central and Southern Asia being particularly disadvantaged. As a direct consequence of this, a significant number of students are not adequately equipped to take part in the very complicated global economy.

Even though there has been significant improvement in educational access and engagement, there were still 262 million children and adolescents between the ages of 6 and 17 who were not in school in the year 2017. That accounted for close to one fifth of the total population in the world within that age bracket. Out of that total, 64 million were children who were of primary school age (about ages 6 to 11), 61 million were adolescents who were of lower secondary school age (ages 12 to 14), and 138 million were young people who were of upper secondary school age (15 to 17 years old).

The COVID-19 epidemic has been wreaking havoc on education systems all around the globe over the last two years, especially harming students who are at a greater risk of academic failure. The epidemic has made existing inequities worse and has

compounded the educational catastrophe that was already there. The extent of the disruptions has varied widely from country to country, with some nations seeing no school closures at all while others experiencing closures lasting for more than a year. At least one third of students were unable to pursue online education because they lacked the necessary expertise.

There is a vast array of untapped potential that could be achieved by integrating and fully accepting artificial intelligence (AI) into the secondary educational systems. AI is here to stay and quite frankly we haven't scratched the surface of the possibilities it provides for educational enhancement. The applicability of AI in the educational sector like in every other sector is endless. Consequently, the subfields of artificial intelligence and machine learning are continually increasing, and researchers are now investigating how their applications might be utilised in the broad number of subject areas, including the disciplines of medicine and even technology that seem to have birthed artificial intelligence itself.

The applicability of AI in the educational sector like in every other sector is endless. Consequently, the subfields of artificial intelligence and machine learning are continually increasing, and researchers are now investigating how their applications might be utilised in the broad number of subject areas, including the disciplines of medicine and even technology that seem to have birthed artificial intelligence itself.

AI has been proven to be beneficial to all parties in the education sector. One of the major benefits for tutors is that AI can help teachers teach their students effectively because teachers now have the opportunity to use AI to create lesson plans that are customized to each student's level of understanding. The students

on the other hand can then learn at their own pace and receive personalized feedback on their progress. Also in the education sector, AI has helped teachers identify gaps in knowledge and provide targeted instruction to keep a steady pace of learning amongst the students, and as a result of that, teachers have been able to monitor students' progress and evaluate whether or not they are meeting the course objectives.

The best way out of poverty and into better economic circumstances is via education. Over the last decade, there has been significant progress achieved in expanding educational opportunities and boosting overall enrollment rates, especially amongst females. However, in 2018, there were still around 260 million children who were not enrolled in school, representing roughly one fifth of the world's population of children aged 0 to 14. Furthermore, more than half of the world's children and adolescents are not performing at the level expected of them in reading and mathematics. More than 91% of the world's pupils would be affected by the temporary shutdown of schools in 2020 as the COVID-19 virus swept over the globe. A projected 1.6 billion children and teenagers would no longer be enrolled in school by April 2020. Even more concerning is the fact that approximately 369 million school-aged children cannot depend on school lunches alone for their daily nutritional needs. There has never been such a widespread absence of children from school at once, threatening to derail the futures of the most vulnerable and marginalised among them. The worldwide pandemic will have far-reaching effects and might threaten recent progress in international education.

The academic performance of students is not only important for the students themselves, but also for the secondary education board that manages the affairs of the institutions. This is because the board uses the academic performance of students as a basis for measuring the success of their educational programs. Many different measures may be used to evaluate the performance of students. The attrition rates stand out among these indicators as an important factor. In other words, a decrease in the incidence of students dropping out of school might be seen as an indication of an increase in students' academic performance and proper feedback from the overall analysis to make students be able to manage their time properly.

The standard of education is directly impacted by factors such as the availability of necessary resources and the level of training received by educators. The provision of fundamental resources to educational institutions is the most difficult to achieve in sub-Saharan Africa compared to other locations. The situation is most dire at the elementary and lower secondary levels, where less than half of the schools in sub-Saharan Africa have access to potable water, electrical power, computers, and the internet. This is where the problem is at its worst. Increasing the number of educated teachers working in schools is an additional significant step toward the realisation of the vision of high-quality education for everyone. Again, sub-Saharan Africa falls behind other regions of the world. This area has the lowest percentages of trained teachers in pre-primary education (48%), primary In 2017, primary education (64%), and secondary education (50%), respectively.

1.1 Problem Statement

The academic performance of students is not only important for the students themselves but also for the secondary education board that manages the affairs of the institution. This is because the board uses the academic performance of students as a basis for measuring the success of their educational programmes. There are many different measures that may be used to evaluate the performance rate as well as the performance indicators of students. Further relating to how a decrease in the incidence of students dropping out of school might be seen as an indicator of increase in students' academic performance.

The secondary purpose of this activity is to determine which students are "at risk" (AR) and which are "not at risk" (NAR) in terms of their grades. The first category of students consists of individuals who are at risk of failing at least one of their classes and so not graduating from the institution. The second group of students is comprised of those who will be successful in making use of the elements that contribute considerably to their academic achievement during the course of the session. For the goal of conducting data analysis, this research work utilized both an artificial neural network and a decision tree algorithm. Subsequently, a comparison analysis was carried out to determine which method provides the most accurate prediction and send back proper feedback to the student in order to curb the subsequent academic sessions soon in order to improve their performance.

1.1.1 Value Proposition

Education is crucial. It's crucial to future achievement and life prospects. Education is beneficial. It brightens the intellect and reasoning. It prepares people

for jobs or college. Education teaches people how to think, feel, and act in a manner that contributes to their success and enhances the amount of personal and social enjoyment they experience. Education builds personality, ideas, and social abilities. It prepares for life. It raises people's standing in their own culture and beyond. Everyone deserves a "cradle-to-grave" education. Education benefits include a solid career, social standing, and self-confidence. Education gives us a strong professional start. We can work anywhere we choose. It improves job prospects. Education improves life opportunities. Education polishes our intellect and enhances our thoughts, character, and conduct toward others. It gives us knowledge in general and in our specialty, specifically, what we need to know for our job. Without education, we can't survive or have a good job.

The main motivation for this research is to improve students' academic performance based on the analysis of curriculum and student academic performance in different courses. Identifying prerequisites and time management for a course is key to enhancing the student experience, improving curriculum design, and increasing graduation rates.

1.2 Objectives

The study's primary goals are to use AI in practical settings and tools for assessing and implementing alterations that might improve students' academic results.

1.3. Research Hypotheses

- There will be an increase in productivity for the classes that utilized the SMS in: delivery efficiency, timetable management, and Test and grading proficiency.

- The students in classes that adopted the use of an SMS as an AI tool for academic enhancement will notice a positive spike in their overall performance.

1.4 Report Outlines

The second section provides context for the discussion of textual analysis and performance prediction in the third. The research methods, including a general approach and a systematic literature evaluation, are discussed in Section 3. The report's experiments, findings, and discussion can be found in Section 4, while recommendations for moving forward can be found in Section 5.

CHAPTER TWO

BACKGROUNDS AND RELATED WORKS

2.0 Introduction

This systematic review was performed according to the standards set out by Kitchenham and Charters [BA and Charters, 2007]. The procedure has three primary phases: preparation, execution, analysis, and interpretation. In the planning phase, you may find related studies and formulate your research topic. Using the PICOC approach [BA and Charters, 2007], we formulated our research topic and its associated keywords and synonyms in order to construct our search string for locating pertinent linked publications. The final query is shown below

(Computer Science, Computer Engineering, or Informatics Course Outline, Learning Objectives, Curriculum, or Learning Objects) .This systematic review's inclusion criteria required that papers meet the following conditions: In order to be considered, papers must meet the following criteria: (1) they must have been written in English; (2) they must be directly related to semantic technologies, computer science, and curriculum; (3) they must be 4 pages or longer (i.e., complete research papers); (4) they must be available online; and (5) they must be

scientifically sound, present a clear methodology, and conduct a proper evaluation of the proposed method, tool, or model.

The steps involved in selecting the final papers are shown in Figure 2.1. At first, we gathered a total of 4,510 documents. The ACM Digital Library1, the IEEE Xplore Digital Library2, and the Google Scholar. In this study, we use classification algorithms, ensemble classification techniques, and the ensemble filtering method to analyse the academic performance of students as well as the influence of various student characteristics on that performance.

Therefore, springer, Science Direct, and the Web of Knowledge as electronic repositories and the search of the Springer online library produced 58 results. We created a basic crawling tool to aid us in the process of including/rejecting publications in our systematic review due to the overwhelming number of papers returned by the Springer search engine.

Everything that was sent back by Springer was compiled into a relational database. After that, we were able to choose the appropriate publications from the Springer database for our systematic review with precision.

The forward and backward snowballing strategy was also used in this systematic review to find relevant publications that were not returned by our first query. Phase 1's final paper list's references were mined for further relevant articles using the backward snowballing approach [Wohlin, 2014], while the papers citing these papers were gathered using the forward snowballing method. The forward snowball technique used Google Scholar7. The total number of eligible studies was 37. Timely; the majority of the research on this issue has just appeared in the previous several years, demonstrating the growing importance of the subject. In

the following chapters, we outline the primary sources that served as the impetus for this thesis's recommendations and methods.

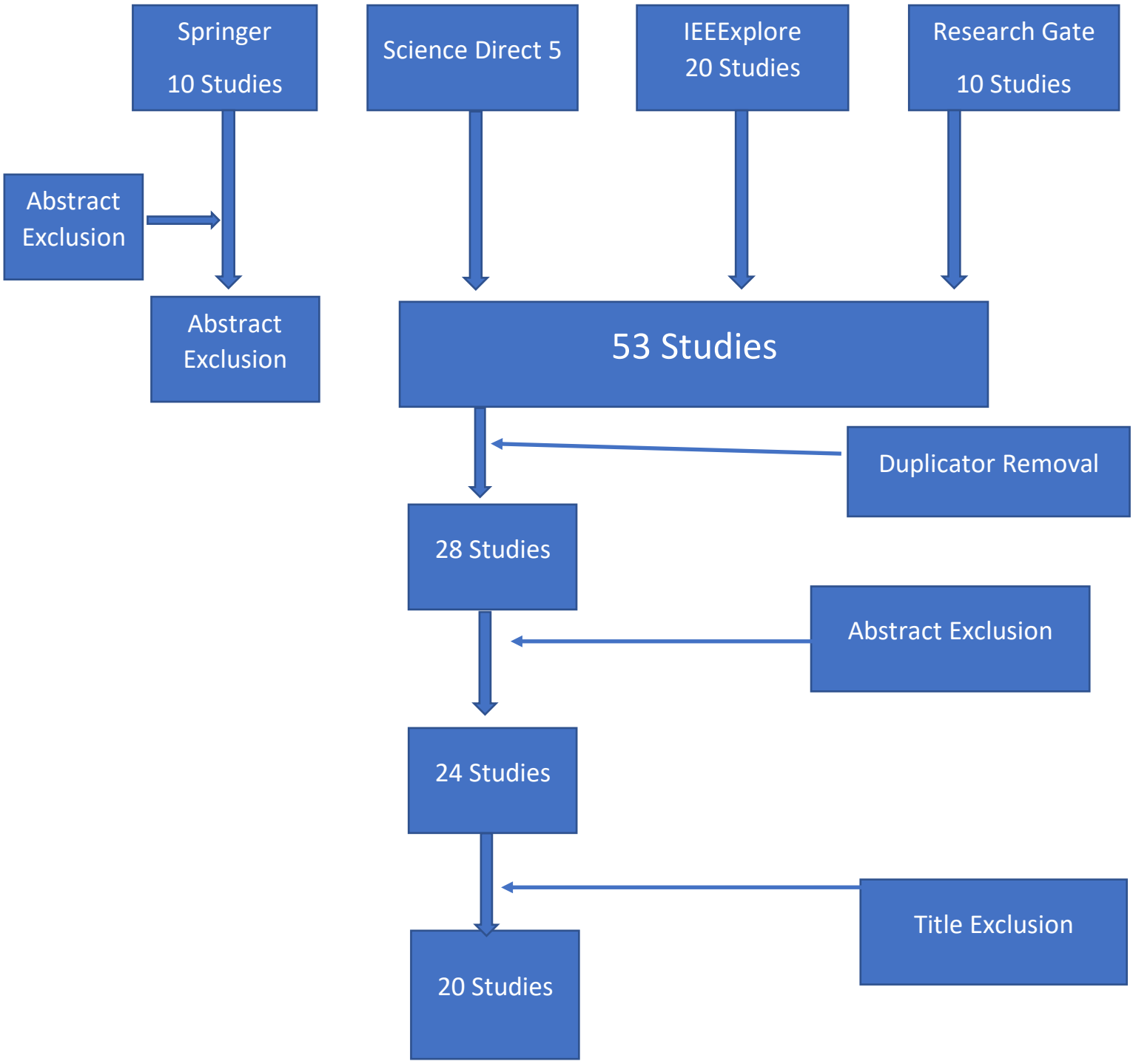


Figure 1. Related works

CHAPTER THREE

LITERATURE REVIEW

3.0 Introduction

A handful of decent researchers like Leonard et al. (2007), and Ricer et al. (2005) have sought out research and empirical data to support the role of some form of technological enhancement that could be achieved by implementing technology into Education. Artificial Intelligence (AI) from its most basic form of instructional technology like Microsoft PowerPoint to more sophisticated implementation in School Management systems to the predictive analysis system, at every point, has been a fundamental tool to enhance productivity as a whole. The span of AI in the past years has truly been limitless. Researchers of great educational standards Wiley et al. (2009) and Pucel and Stert (2005) have viewed web-based learning as an educational enhancement tool and have deemed the importance of the field to be exploited to get its full potential from Web 1.0 technologies to the current Web 3.0 we are starting to experience in 2022. As Ward et al. (2009) perceived, the study has not stayed up with the quickly changing automated developments of the intelligent Web 2.0 instruments. Cramer, Collins, Snider, and Fawcett's (2007) investigated the virtual amphitheatre climate. These instructive analysts have

analyzed the utilization of conventional techniques for guidance to the innovative strategies for guidance. Understudy instructive PC use in friendly and recreational conditions has been contemplated. Bartsch and Cobern (2003), Hansen and Williams (2008), D'Angelo and Wooley (2007), and Cramer et al. (2007) looked to distinguish understudies' inclinations and views of instructive techniques for guidance. Theoretical framework Artificial intelligence as a field has experienced significant development and changes over the years. It is a broad spectrum of how much AI has impacted the world from the early 2000s to the current year 2022. However, our focus in this research has been anchored on the evolution of artificial intelligence in education. To be able to align a positive and constructive future where the implementation of artificial intelligence both at the level of administration and at the level of students, we must delve into past studies and research to be able to grasp some degree of positive empirical values. Identifying strengths and opportunities that might lie ahead might breathe room for more research that could benefit the implementation of AI in the future. Two parallel strands of research that might bring an overhaul to the implementation or utilization of AI in the educational sector were identified by Roll and Wylie (2016); one of them is an evolutionary process that focuses on current classroom practices cooperating with teachers and varying technologies and related domain. The second one is an innovative process where an argument for embedding artificial intelligence technology within students' everyday lives, supports their cultures, practices, goals, and communities. Evolution of technology in education in Africa Technology is a labyrinth on its own and to comprehend it, we must be able to associate it with the various aspects of functionality in this instance we would look at educational technology. Educational technology is defined as "the use of

information and communication technologies (ICT) to improve learning outcomes". Educational technology encompasses a wide range of activities including teaching, learning, assessment, research, and evaluation. Educators have been using technology to enhance education since the early 1900s. However, the advent of the internet has led to a revolution in how people learn. Today, technology is being used to provide students with opportunities to engage with knowledge in ways that were not possible before. Although educational technology provides numerous benefits, it also presents some challenges. One challenge is ensuring that all students have access to the same level of technology. Another challenge is ensuring that all teachers and administrators are trained in how to effectively use educational technology. Africa like most African countries has picked up the mantle of embracing technology in the educational sector this was seen more vividly when President Paul Biya (2016) initiated the "e-national higher education" convention, which was signed on Feb. 10, 2016, between the ministry of higher education and Sichuan Telecom Construction Engineering Co. Ltd, in line with President Biya's "Plan Special Jeunes" (Special Youth Plan) which aimed to provide a laptop to every student registered in a public or private college in Africa. The initiative was entitled "un étudiant un ordinateur" (A student, a laptop). This was a bold step by the country's head of state to enforce the implementation of technology into education. In the late 1990s and early 2000s, only the most elite and prestigious secondary schools in the country had the opportunity to have some form of technological equipment such as computers running Microsoft Windows 3.1 and the Disk operating system (DOS). These computers were not being used for educational enhancement but more succinctly to teach a particular subject called Information Technology (IT). Knowledge about technology in this era was gotten

from textbooks. The overall administrative process of most institutions was been handled manually. This was reality right up to 2010 when there was a shift in the accessibility of computers and circulation of laptops and internet technology.

3.1 Artificial Intelligence Implementation in Education

In 1955, while working at Dartmouth College, John McCarthy was the first person to create the phrase "artificial intelligence." . AI is defined as “the science and engineering of making intelligent machines”. In essence, AI is the study of how computers think and learn. AI is often associated with machine learning, where systems learn without being explicitly programmed. AI is becoming increasingly relevant in our daily lives and is now applied to many different fields, including healthcare, finance, manufacturing, and even education. Artificial Intelligence in education can be broken down into 3 facets;

Educational Technology: Educational technology is the application of information and communication technologies to educational settings. Educational technology is a broad field that includes computer hardware, software, and web applications. Educational technology encompasses a wide range of topics, including distance learning, digital textbooks, online tutoring, virtual classrooms, and eLearning.

Educational Management Information System (EMIS) or School Management System (SMS): Education management information systems (EMIS) are a system of managing the administrative aspects of schools and colleges. EMIS is designed to help teachers, administrators, and students manage their school's finances, curriculum, and student records.

Computer Assisted Instruction (CAI): Computer-assisted instruction (CAI) is a method of teaching that uses computers to assist instructors in delivering lessons. CAI is commonly used in higher education institutions, especially those that offer distance learning programs.

SDG 4 aims to ensure that all people have access to quality education by promoting lifelong learning opportunities and reducing inequalities in access. Artificial intelligence (AI) has the potential to address some of the biggest challenges in education today, innovate teaching and learning practises (Sustainable Development Goal 4 is about quality education and is among the 17 sustainable development goals established by the United Nations in September 2015). However, there are many hazards and concerns associated with these rapid technical advancements, and thus far, policy discussions and regulatory frameworks have not kept up. The United Nations Educational, Scientific, and Cultural Organization (UNESCO) is dedicated to assisting its Member States in realising the potential of artificial intelligence technologies in advancing the Education 2030 Agenda, while also ensuring that the use of AI in educational settings is guided by the principles of equity and inclusion. To wit: UNESCO's "Artificial Intelligence in Education" (2016). Education institutions, in particular, have been early adopters and major users of AI in a variety of forms for pedagogical purposes. AI has evolved from its original form in computers and related technologies to web-based and online intelligent education systems, and then to humanoid robots and web-based chatbots that can perform instructors' duties and functions independently or in tandem with instructors using embedded computer systems and other technologies. "Artificial Intelligence in Education: A Review," by L. Chen, P. Chen, and Z. Lin (2020).

3.2 Impact of Artificial Intelligence on Education

Artificial Intelligence has the capacity to learn from experience and improve its performance over time. In education, AI is being implemented in various forms including automated grading, tutoring, a personalized learning environment, and enhancing teaching methods and assessment.

3.1.1Automated Grading

Automated grading refers to the use of software programs to grade student assignments. A program would read each assignment and then assign a score based on certain criteria. The program could be programmed to look at things like grammar, spelling, punctuation, sentence structure, and word choice. If the program detects any errors, it may give the assignment a low score. However, if the program finds no errors, it might give the assignment a high score.

3.1.2 Tutoring

Tutoring involves having a teacher or tutor help students' complete assignments; Students who receive tutoring often do much better than those who don't. Tutors can provide feedback on assignments and answer questions about concepts

covered in class. Tutors can also help students prepare for tests and quizzes with the aid of tutor-bots.

3.1.3 Personalized learning environment (PLE)

Personalized learning environment refers to the use of technology to provide individualized instruction based on student performance. Students can receive feedback on their progress and teachers can adjust their lessons accordingly. Teachers can also use data analytics to identify patterns in student behavior and make adjustments to lesson plans.

3.1.4 Enhancing Teaching Method

Enhancing teaching methods involves using AI to improve the effectiveness of classroom activities. For example, AI can help teachers analyse student responses and determine whether they need additional practice before moving on to the next step. AI can also help teacher’s grade assignments automatically and track student progress over time.

3.1.5 Education Administration

AI application in education, in its various forms and serving different functions, has had a major impact on the performance of administrative and management functions in education. It has enabled instructors or teachers to perform their administrative functions, such s as grading and providing feedback to students more effectively S. Pokrivcakova, (2019) "Preparing teachers for the application of AI-powered technologies in foreign language education". The era of the School Management system has gone a long way to expedite processes in secondary

schools which was somewhat cumbersome in the ear without Artificial technology. Grading and providing students with a result slip used to be a manual process where all the teachers of a particular class had to do their calculation manually and then take it to their respective class masters for cross-checking and validation. Before the class master could aggregate and calculate the average of each student and submit it to the dean of studies to prepare the final script beforehand report cards could be issued to the students on closing day. It has always been a rigorous and tedious process. With the number of students increasing every day, one will be tempted to say the task is becoming impossible without the aid of Artificial Intelligence. In this research, it was covered; there was evidence of attainment of improvements in the administrative processes and tasks quality, as well as effectiveness and efficiency of the instructors or educators in the performance of various administrative tasks.

3.1.6 Student Learning Experience

Another area that is greatly impacted by the benefits of Artificial Intelligence in education is the aspect of learning. As the administrators and instructors gain more time from automating the process that were manually been handled, the additional time has been used to transfer more knowledge to the students and also to break down the already transferred knowledge for better understanding. V. Rus, S. D'Mello, X. Hu, and A. Graesser (2013) other studies illustratively discuss the numerous benefits of AI in students' learning experiences in different ways. AI enables the tracking of learning progression, including knowledge and understanding and uses the findings to enhance the capabilities of the system to

customize content to the student's needs and capabilities, which motivates students and leverages personal capabilities to enhance uptake and retention.

3.1.7 Performance of Instructors and Students

It would be important to view how artificial intelligence as a technological enhancement tool, will affect the performance of students and instructors. Each and every year the number of admissions increases in secondary school leaving a huge number of students to be managed by a handful of administrators, it is in this light that an Artificial intelligence system will shine its light on easing the workload. As it can help in analyzing and drafting syllabi and setting up timetables for more customized content M. Chassignol, A. Khoroshavin, A. Klimova and A. Bilyatdinova, (2018) "Artificial intelligence trends in education: A narrative overview". AI systems have the ability to make futuristic predictions based on students' performance and strength in particular subjects. This provides administrators like the Dean of student lives with data that could enable them to give decent and fair advice to both the students and parents with respect to a desire or a preferred career path. Based on the above discussion Artificial intelligence hold great potential in both the administrative aspect of education and the student-centric aspect of learning. As it automates manual processes leaving the administrators with ample time to focus on the academic welfare of the students and indirectly increases students' performance as a result of the additional time. AI will soon be able to work as a full-fledged assistant and adapt to a wide variety of learning styles to help instructors and students. In specific, it helps instructors and students with their educational needs in just about any area of need

CHAPTER FOUR

METHODOLOGY

4.0 Classification Approach A

4.1 Data Set

The dataset used examines the academic performance of students attending two secondary schools in Portugal. The information was obtained via the use of school reports and questionnaires, and the characteristics include educational factors, demographic factors, social factors, and educational elements. Two separate datasets are presented here, one focusing on performance in the Portuguese language and the other on performance in mathematics (mat) (por). The two datasets were modelled in [Cortez and Silva, 2008] by employing binary/five-level classification and regression tasks. The fact that the target characteristic G3 has a strong connection with the qualities G2 and G1 indicates that the link is considerable. This is because the grade for the last period of the year, G3, is given out during the third period, whilst the grades for the first and second periods, G1 and G2, respectively, are given out during the first and second periods. Without G2 and G1, it is far more difficult to make a prediction about G3, despite the fact that such a prediction is noticeably more accurate.

school	student's school	binary: 'GP' - Gabriel Pereira	'MS' - Mousinho da Silveira)	
sex	student's sex	binary	'F' - female	'M' - male

age	student's age	numeric	15 to 22	
address	student's home address	binary	'U' - urban	'R' - rural
famsize	family size	binary	'LE3' - less	'GT3' - greater than 3)
pstatus	parent's cohabitation status	binary	'T' - living together	'A' - apart
medu	mother's education	numeric	0 - none, 1 - primary education (4th grade),)	2-5th to 9th grade, 3:secondary education or 4 :higher education
fedu	father's education	numeric	0 - none, 1 - primary education (4th grade),)	2-5th to 9th grade, 3:secondary education or 4 :higher education
mjob	nominal	'teacher', 'health' care related	Civil ‘services' (e.g. administrative or police)	'at_home' or 'other')

fjob	nominal	'teacher', 'health' care related	Civil ‘services' (e.g. administrative or police)	'at home' or 'other')
reason	nominal	close to 'home', school	'reputation', 'course'	preference or 'other')
gaurdian	nominal	'mother'	'Father'	'other'
traveltime	numeric	1 - <15 min., 2 - 15 to 30	min., 3 - 30 min. to 1 hour	4 - >1 hour)
Study time	numeric	1 - <2 hours, 2 - 2 to 5 hours	3 - 5 to 10 hours	4 - >10 hours)
failures	numeric	n if 1<=n<3, else 4)	n if 1<=n<3, else 4)	1<=n<3, else 4)
schoolsup	binary	extra educational support	yes	no
famsup	binary	family educational support	yes	no
paid	binary	extra paid classes within the course subject	extra paid classes within the course subject (Math or Portuguese)	extra paid classes within the course subject (Math or Portuguese)

			(binary: yes or no)	(binary: yes or no)
activities	binary	extra-curricular activities	yes	No
nursery	binary	- attended nursery school	yes	no
higher	binary	higher - wants to take higher education	yes	no
internet	binary	Internet access at home	yes	no
romantic	binary	romantic - with a romantic relationship	yes	no
famrel	numeric	quality of family relationships	1 - very bad	5 - excellent)
freetime	numeric	free time after school	1 - very low to	5 - very high
Go out	going out with friends (numeric: from 1 - very low to 5 - very high)	going out with friends (numeric: from 1 - very low to 5 - very high)	going out with friends (numeric: from 1 - very low to 5 - very high)	going out with friends (numeric: from 1 - very low to 5 - very high)

Dalc	- workday alcohol consumption (numeric: from 1 - very low to 5 - very high)	- workday alcohol consumption (numeric: from 1 - very low to 5 - very high)	- workday alcohol consumption (numeric: from 1 - very low to 5 - very high)	- workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
Walc	Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)	Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)	Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)	Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
Health	- current health status (numeric: from 1 - very bad to 5 - very good)	- current health status (numeric: from 1 - very bad to 5 - very good)	- current health status (numeric: from 1 - very bad to 5 - very good)	- current health status (numeric: from 1 - very bad to 5 - very good)
Absences	number of school absences (numeric: from 0 to 93)	number of school absences (numeric: from 0 to 93)	number of school absences (numeric: from 0 to 93)	number of school absences (numeric: from 0 to 93)

G1	first period grade (numeric: from 0 to 20)	first period grade (numeric: from 0 to 20)	first period grade (numeric: from 0 to 20)	first period grade (numeric: from 0 to 20)
G2	second period grade (numeric: from 0 to 20)	second period grade (numeric: from 0 to 20)	second period grade (numeric: from 0 to 20)	second period grade (numeric: from 0 to 20)
G3	final grade (numeric: from 0 to 20, output target)	final grade (numeric: from 0 to 20, output target)	final grade (numeric: from 0 to 20, output target)	final grade (numeric: from 0 to 20, output target)

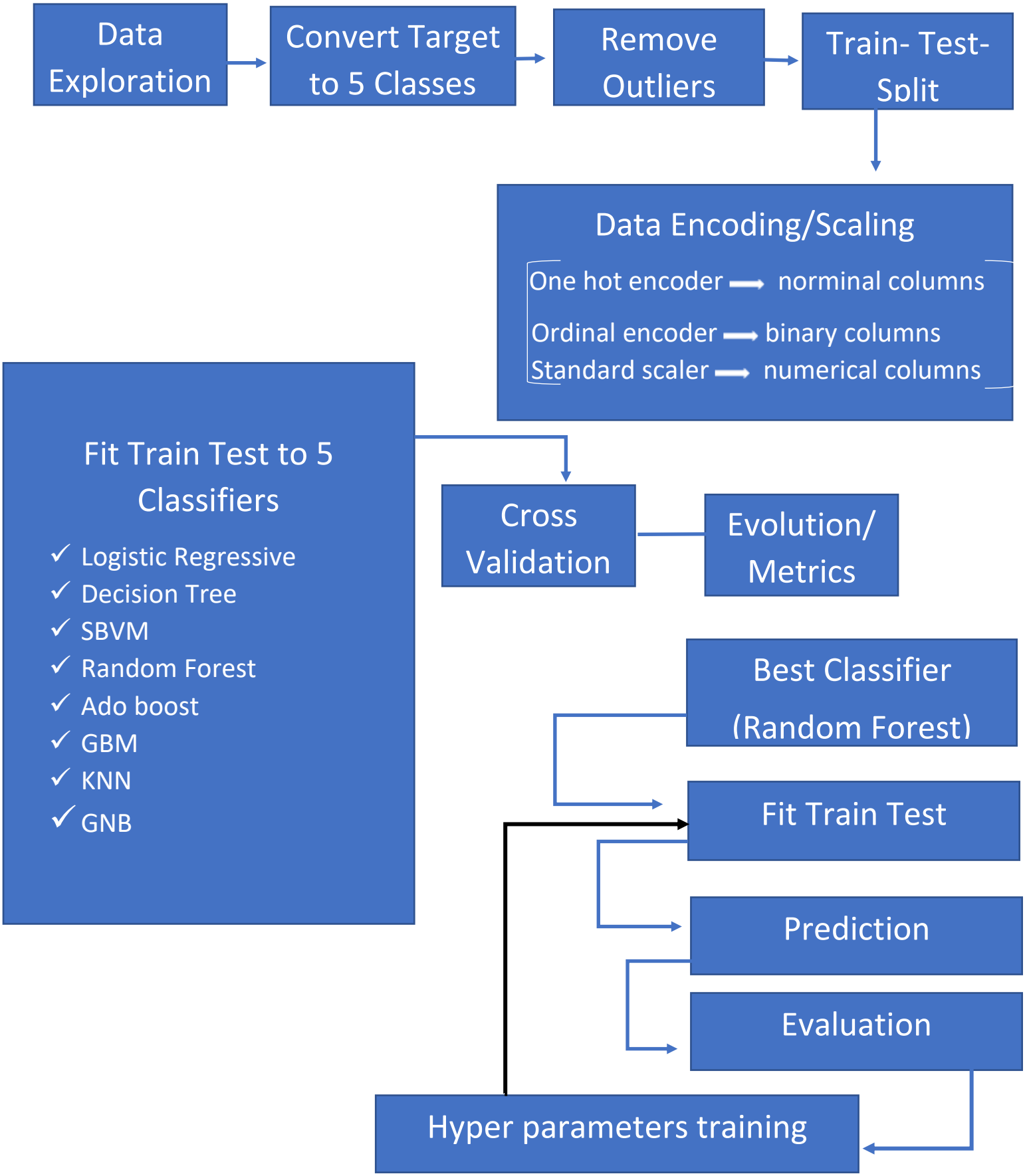
Table 1. Attributes for both student-mat.csv (Math course) and student-por.csv (Portuguese language course) datasets:

4.1.1 Data Validation

In the context of this project, the purpose of data validation is to verify the accuracy and quality of the datasets that were described in the previous section. This is accomplished by comparing the patterns that are present in the datasets to some of the patterns that were discussed in the literature review.

Data Pre-processing and feature engineering

Figure 2: Data pre-processing and modelling



The whole process for improving student performance is broken down into four stages, as is evident from the strategy that was used. The following sections will provide an explanation of the specifics of each stage. In a nutshell, the first phase in the process of improving student performance is responsible for the pre-processing of data, which includes gathering datasets. Techniques related to data cleaning and machine learning (ML) are used in this stage. The selection of the features is within the purview of the second stage. It makes use of the data that was pre-processed in the first step's output. The steps 3 and 4 have been combined into one, and the process now begins with training and testing the random forest.

Below is more detailed outline of the process followed during the approached:

- A new column ‘avg’ was formed from averaging the scores of G1, G2 and G3
- A new column ‘grades’ was then formed by classifying the average scores into 5 classes of grading thus: (0, 9.5) – Fail, (9.5, 11.5) -- Sufficient/ Fair, (11.5, 13.5) Satisfactory, (13.5, 15.5) – good, (15.5, and 20) – Excellent.
- Outliers were detected and deleted.
- Data was divided into numerical and categorical datasets.
- Insignificant and useless columns were dropped: 'G1', 'G2', 'G3', 'avg', 'school'
- Dataset was split into test and train sets. The target column was set to be ‘grades’ as that is what we will be predicting.
- Data Scaling was done Data was divided into binary, nominal, and numerical columns.
- A column transformer was then done to fit in encoders: binary columns encoded with Ordinal Encoder, nominal columns with one hot encoder, and

numerical columns scaled with a standard scaler. All from Scikit learn library.

- Cross-validation was done.
- The following functions—Logistic Regression, Decision Tree Classifier, Random Forest Classifier, AdaBoost Classifier, K Neighbors Classifier, Gaussian Naive Bayes, Support Vector Machine, and Gradient Boosting Classifier—were built to integrate several classification techniques to compare metrics. Random forest classifier seemed to perform better.

4.3 Data Collection

The facts presented here pertain to the academic accomplishments of students in two Portuguese secondary schools. The data was gathered through the use of school reports and questionnaires, and the properties of the data include student grades as well as demographic, social, and school-related characteristics. There are offered two datasets, one for the performance in the topic of mathematics (mat), and the other for the performance in the language of Portuguese (por). The two datasets were modelled in the research conducted by Cortez and Silva (2008) using binary/five-level classification and regression tasks

4.3.1 Data Set

Read in from csv format into the notebook using Python's Pandas library. It contains results of Maths and Portuguese of students for 3 trimesters with other different columns depicting other features contributing to the performances (results).

The merged dataset contains 1044 entries and 33 columns.

4.3.2 Data Cleaning

Technique	Operation
Data Cleaning	outliers/duplicates correction
Data Validation	inappropriate data identification
Data Enrichment	Data enhancement
Data Imputation	missing values filling
Data Normalisation	data re-scale

Table.2. Data cleaning process

Characters from the dataset that won't be used to train the model are to be expected. Characters like (),!, @, #, $, percent, & *, +, =,, [],?, and others. The data set excludes these characters and those that are similar. The removal of blank values from the data collection is another aspect of cleaning. As real-world data are being used, this is a crucial stage of the project's execution. is rarely clean, making effective cleaning methods crucial. Outlier detection is one of the other cleaning methods. There should be three categories in the data set's class column. It is crucial to verify that the data collection has just three categories. Any other category that isn't among the anticipated categories is eliminated. Characters from the dataset that won't be used to train the model are to be expected. Characters like (), ! , @, #, $, percent, & *, +, =,, [],?, and others. The data set excludes these characters and those that are similar. The removal of blank values from the data collection is another aspect of cleaning. As real-world data are being used, this is a crucial stage of the project's execution. is rarely clean, making effective cleaning methods crucial. Outlier detection is one of the other cleaning methods. There should be three categories in the data set's class

column. It is crucial to verify that the data collection has just three categories. Any other category that isn't among the anticipated categories is eliminated.

```
In [101]: #Check for null values
          df.isna().sum()

Out[101]: school        0
          sex           0
          age           0
          address       0
          famsize       0
          Pstatus       0
          Medu          0
          Fedu          0
          Mjob          0
          Fjob          0
          reason        0
          guardian      0
          traveltime    0
          studytime     0
          failures      0
          schoolsup     0
          famsup        0
          paid          0
          activities    0
          nursery       0
          higher        0
          internet      0
          romantic      0
          famrel        0
          freetime      0
          goout         0
          Dalc          0
          Walc          0
          health        0
          absences      0
          G1            0
          G2            0
          G3            0
          dtype: int64
```

Figure.3 checking of null values

Checking of null values: To prevent erroneous conclusions and misunderstandings, data must be cleaned before any analysis can be performed. The purpose is to get rid of rows that are mostly blank, have too many outliers, or contain duplicate information. Since the other two datasets have no blanks or duplicates, this one need to be cleaned up. In order to methodically sanitize data sets. The procedure consisted of the following two stages: - empty cell was found; hence nothing was dropped initially.

4.4 Exploratory Data Analysis

An essential component of this study is the visual representation of the goal variable. The target variable's visualization reveals the data set's overall balance. Unbalanced data hurts the model's performance. As a result, the model becomes biased, meaning that it tends to favour the dominating category in the data. Consequently, data set visualization illustrates the distribution of the output variable. We can address the imbalance in our output variable using a variety of methods of oversampling and undersampling. The oversampling method, as the name suggests, uses an iterative process to generate data in order to produce fake strings within the segment between two supplied samples of the training set (Castellanos et al., 2018). The number of samples in the data set rises as a result of oversampling. The rise depends on the dataset's size and the dataset's degree of imbalance. This method was used to create this report. The dominating class's size is decreased while using the under-sampling technique to balance the data. This strategy can result in the loss of vital information for the categorization model.

Libraries used:

- Numpy
- Pandas
- Matplotlib
- Seaborn.
- Matplotlib.
- Seaborn
- Sklearn

4.5 Univariate Analysis

Feature distribution was carried out on the categorical and numerical data analysis.

```
# Change categorical columns to datatype "category"

for col in['school', 'sex', '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']:
    df[col] = df[col].astype('category')
```

Figure: 4. Categorical columns to data type

- Categorical data: A bar chart was plotted on the categorical columns.

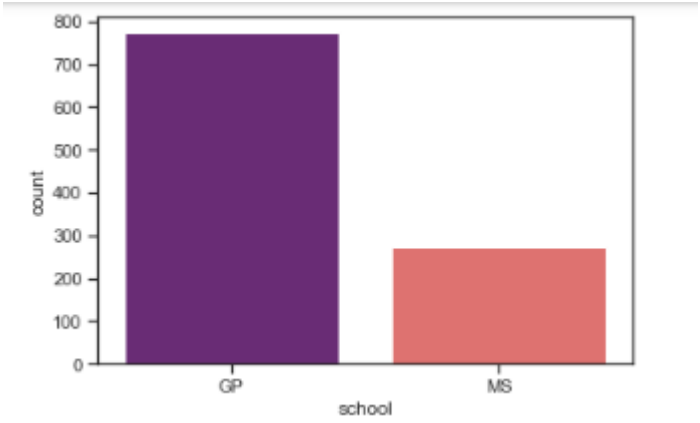
The categorical columns are: 'school', 'sex', 'address', 'famsize', 'Pstatus', 'Medu', 'Fedu', 'Mjob', 'Fjob', 'reason', 'guardian', 'traveltime', 'studytime', 'failures', 'schoolsup', 'famsup', 'paid', 'activities', 'nursery', 'higher', 'internet', 'romantic', 'famrel', 'freetime', 'goout', 'Dalc', 'Walc', 'health

Numerical Data: count plot was don on these columns: age, absences, G1, G2, G3

Findings from the data entries:

- Gabriel Pereira school are more than Mousinho da Silveira school.

Fig 5: Gabriel Pereira School are more than Mousinho da Silveira school



- Females are more than males.

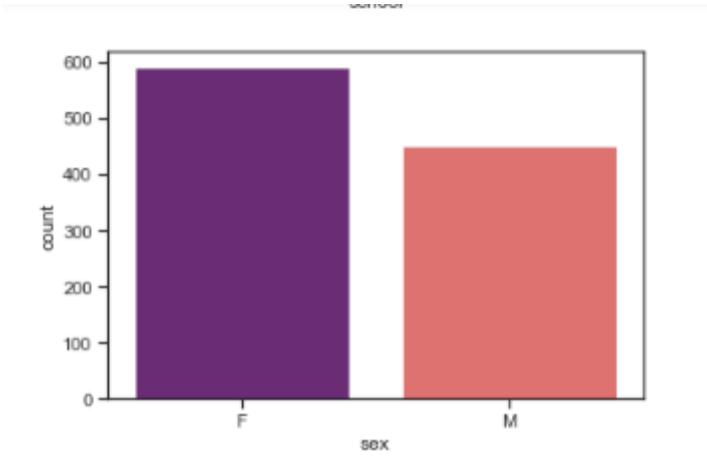


Fig 6: Females are more than males school

- Students from urban areas are more than those from rural areas.

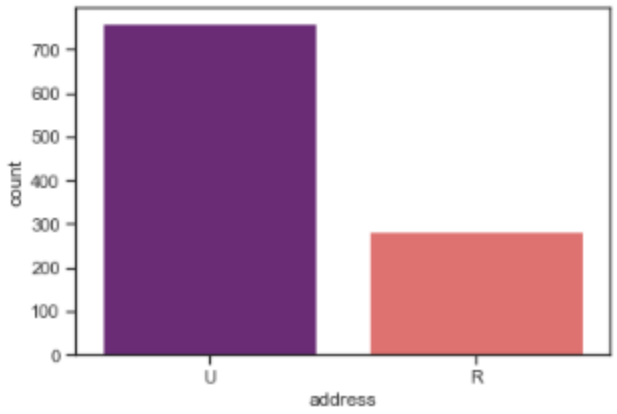


Fig 7: Students from urban areas are more than those from rural areas.

- Students from family sizes greater than three are more than those of family of less than three.

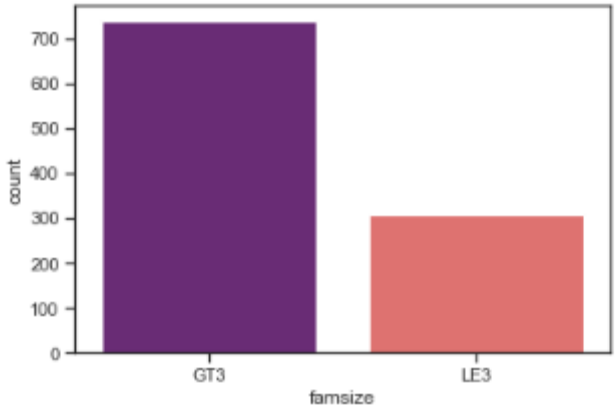


Fig 8: Students from family sizes greater than three are more than those of family of less than three.

- Students whose parents are together are more than those whose parents are not.

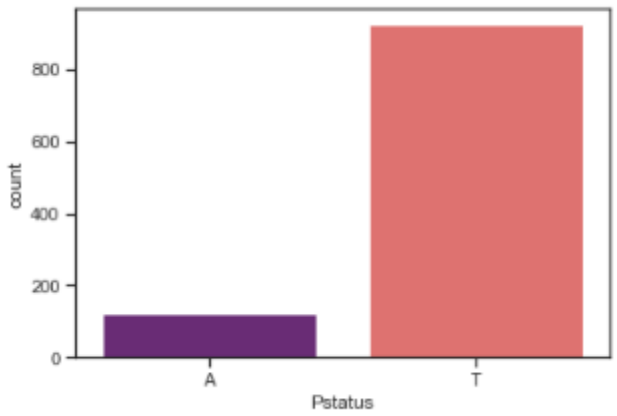


Fig 9: Students whose parents are together are more than those whose parents are not.

- Students with school support are lesser and those with family supports are more.

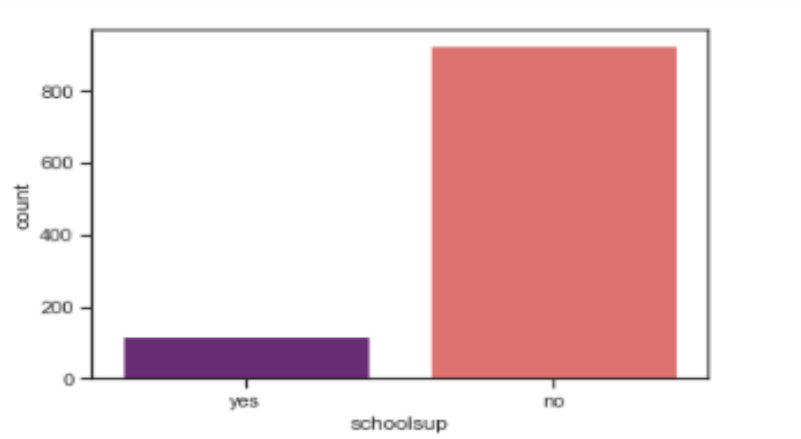


Fig 10: Students with school support are lesser and those with family supports are more.

- Students that attend paid extra classes are lesser and those that attended nursery classes are more.

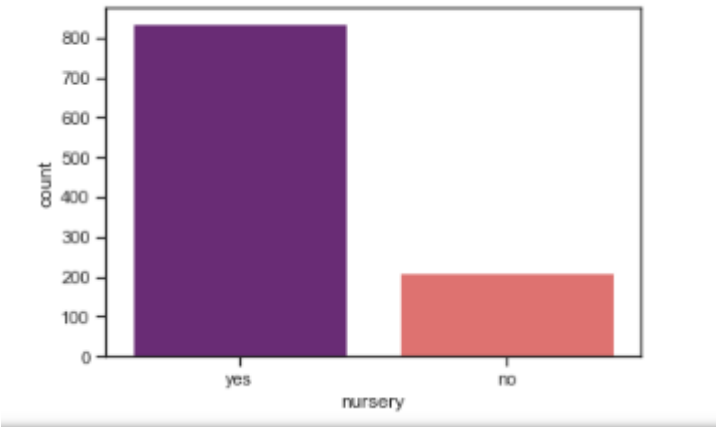


Fig 11: Students that attend paid extra classes are lesser and those that attended nursery classes are more.

4.6 Descriptive Analysis

The act of analysing data is known as descriptive analysis, and it is this process that enables the determination of information such as the mean, mode, median, standard deviation, variance, and skewness. The descriptive analysis does not present any opportunities for predictions to be made, nor does it make it possible to draw any conclusions; rather, it only provides a synopsis of the data (Kaliyadan and Kulkarni, 2019).This section is broken up into two sections, the first of which focuses on the category's columns, while the second looks at the numerical columns. Categorical columns are investigated using a contingency table, which displays the count for each distinct combination of values in the specified columns.

```
In [99]: # inspecting

         display(df_mat.head(),
                 df_por.head(),
                 df_mat.describe(),
                 df_por.describe(),
                 df_mat.info(),
                 df_por.info())
```

	age	Medu	Fedu	traveltime	studytime	failures	famrel	freetime	goout	Dalc	Walc	health	absences	
count	649.000000	649.000000	649.000000	649.000000	649.000000	649.000000	649.000000	649.000000	649.000000	649.000000	649.000000	649.000000	649.000000	64
mean	16.744222	2.514638	2.306626	1.568567	1.930663	0.221880	3.930663	3.180277	3.184900	1.502311	2.280431	3.536210	3.659476	1
std	1.218138	1.134552	1.099931	0.748660	0.829510	0.593235	0.955717	1.051093	1.175766	0.924834	1.284380	1.446259	4.640759	2
min	15.000000	0.000000	0.000000	1.000000	1.000000	0.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	(
25%	16.000000	2.000000	1.000000	1.000000	1.000000	0.000000	4.000000	3.000000	2.000000	1.000000	1.000000	2.000000	0.000000	1(
50%	17.000000	2.000000	2.000000	1.000000	2.000000	0.000000	4.000000	3.000000	3.000000	1.000000	2.000000	4.000000	2.000000	1
75%	18.000000	4.000000	3.000000	2.000000	2.000000	0.000000	5.000000	4.000000	4.000000	2.000000	3.000000	5.000000	6.000000	1
max	22.000000	4.000000	4.000000	4.000000	4.000000	3.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	32.000000	1(

```
None
None
```

Figure:12 Counts on the dataset

4.7 Bivariate and Multivariate Analysis

Barplot was done to compare the relationship between some columns.

The correlation was done using the heatmap and in the data frame table.

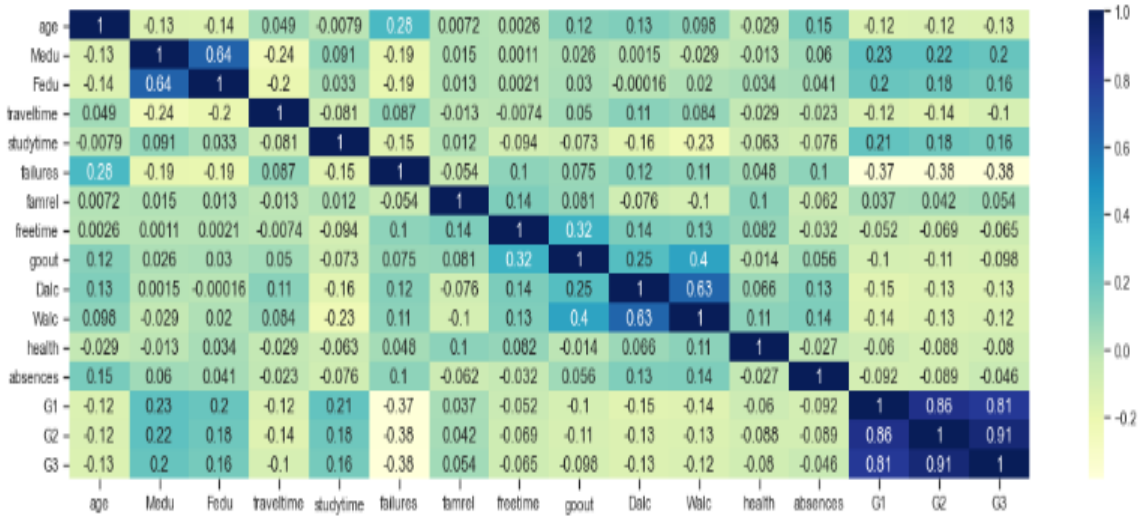


Fig 13: Bivariate and Multivariate Analysis

There seems to be an element of hypercorrection (G1, G2, G3) which will eventually be corrected to make the machine learning model.

4.7.1 Modelling

The problem is a regression problem, requiring us to predict the student's final scores, G3. Hence, we made use of regression algorithms (Linear regression and Random Forest).

However, good metrics and scores could not be gotten so another approach was used – the problem was converted to a classification problem and good classification algorithms were used which yielded better results

Libraries: Major library used was Scikit learn. Different sub libraries were imported as the needs arise.

Modeling

In [135]:
```
from sklearn.model_selection import train_test_split
from sklearn.base import BaseEstimator, TransformerMixin
from sklearn.preprocessing import OneHotEncoder, MinMaxScaler
from sklearn.impute import SimpleImputer
from sklearn.pipeline import Pipeline, FeatureUnion
from sklearn.compose import ColumnTransformer
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import r2_score
from sklearn.ensemble import RandomForestRegressor
```

Fig 14: Feature engineering:

- Outliers were removed.

Feature Engenering

In [137]:
```
# Outlier removal

result = df.select_dtypes(include='number')#selecting dtypes in dataset

for i in result.columns:
    percentile25 = df[i].quantile(0.25)
    percentile75 = df[i].quantile(0.75)

    iqr = percentile75-percentile25

    upper_limit = percentile75 + 1.5 * iqr
    lower_limit = percentile25 - 1.5 * iqr

    df[df[i] > upper_limit]
    df[df[i] < lower_limit]

    df_new = df[df[i] < upper_limit ]
    df_new = df[df[i] > lower_limit ]
```

Fig 15: Outliers were removed

4.7.2 Linear Regression Model

- Data was splitted into train and test set, setting G3 as the target column.
- Train set was divided into categorical and numerical sets.

1. Linner Regression Model

In [139]:
```
#  Spliting Column in Train and Test set

X_train, X_test, y_train, y_test = train_test_split(df_new.drop(columns=['G1','G3','G2','absences']),
                                                    df_new['G3'],
                                                    test_size=.2,
                                                    random_state=10)
```

In [140]:
```
#  Further Divide dataframe into numerical and categorical type of train dataset

# Define categorical columns
categorical = list(X_train.select_dtypes('category').columns)
print(f"Categorical columns are: {categorical}")

# Define numerical columns
numerical = list(X_train.select_dtypes('number').columns)
print(f"Numerical columns are: {numerical}")
```

```
Categorical columns are: []
Numerical columns are: ['age', 'Medu', 'Fedu', 'traveltime', 'studytime', 'failures', 'famrel', 'freetime', 'goout', 'Dal
c', 'Walc', 'health']
```

- Categorical sets were encoded using one-hot encoder and numerical datasets were normalized using normalizer, both were fitted in a pipeline. All functions from Scikit library.

In [141]:
```
#  One Hot enconding of cat dataset

cat_pipe = Pipeline([
    ('encoder', OneHotEncoder(handle_unknown='ignore', sparse=False))
])
```

In [142]:
```
#  Scaling of numerical dataset

from sklearn.preprocessing import Normalizer
num_pipe = Pipeline([('scaler', Normalizer())])
```

In [143]:
```
#  Combining of both categorical and numerical pipeline

# Combine categorical and numerical pipelines
preprocessor = ColumnTransformer([
    ('cat', cat_pipe, categorical),
    ('num', num_pipe, numerical)
])
```

In [144]:
```
# Fixing Pipeline on Linner regression

pipe = Pipeline([
    ('preprocessor', preprocessor),
    ('model', LinearRegression())
])

pipe.fit(X_train, y_train)
```

- G1 and G2 columns were not considered as the was associated hyper correlation.

- Pipeline was fixed on a linear regression model.

- The final pipeline was the fitted on the training set, predictions done and evaluation through the test sets.
- Evaluation metrics inform of r2 score was done and the score was 18.6 percent. This was quite low, so we had to a tree-based regression algorithm, the random forest regressor.

4.7.3 Random Forest Repressor

- Data was splitted into train and test set, setting G3 as the target column.
- Train set was divided into categorical and numerical sets.
- Categorical sets were encoded using one-hot encoder and numerical datasets were normalized using normlaiser, both were fitted in a pipeline. All functions from Scikit library.
- G1 and G2 columns were not considered as the was associat4ed hypercorrection
- Pipeline was fixed on a random forest regression model.
- The final pipeline was then fitted on the training set, predictions done and evaluation through the test sets.
- Evaluation metrics inform of r2 score was done and the score was 39 percent. This is not too bad. However, the problem is not practically addressed. It will be nice to consider the g1 and g2 as they were previous term results, and it would make more sense to use grading system classifications instead of scores (as the targets). Hence, we'll be converting

the problem to a classification problem, and classification algorithms will be used.

Libraries used: NumPy, Pandas, Matplotlib, Seaborn.

Findings from the data entries:

- Gabriel Pereira School are more than Mousinho da Silveira school.
- Females are more than males.
- Students from urban areas are more than those from rural areas.
- Students from family sizes greater than three are more than those of family of less than three.
- Students whose parents are together are more than those whose parents are not.
- Students with school support are lesser and those with family supports are more.
- Students that attend paid extra classes are lesser and those that attended nursery classes are more.

Libraries: Major library used was Scikit learn. Different sub libraries were imported as the needs arise.

4.7.4 Random Forest Repressor

- Data was splitted into train and test set, setting G3 as the target column.
- Train set was divided into categorical and numerical sets.

- Categorical sets were encoded using a one-hot encoder and numerical datasets were normalized using a normalizer. Both were fitted in a pipeline. All functions from Scikit library.
- G1 and G2 columns were not considered as the was associated hyper correlation
- Pipeline was fixed on a random forest regression model.
- The final pipeline was the fitted on the training set, predictions done and evaluation through the test sets.

• The evaluation metrics informed of r2 score were completed, and the score was 39%. This is not such a terrible thing. However, there is no concrete action taken to solve the situation. It would be wonderful to take into consideration the results of the G1 and G2 as they were from the previous term, and it would make more sense to apply the categories of the grading system rather of the scores (as the targets). Because of this, the issue will be recast as a classification issue, and classification techniques will be used in the resolution process.

4.8 Classification Approach B

4.8.1 Pre-Processing and Feature Engineering

Any prediction model requires first-rate data preparation. Any out-of-the-ordinary information may be discovered, as well as huge variables, false positives, and missing data. Cleaning up the data set improves the model's accuracy, which in turn improves the scenario. All data categories, nominal values, and data integrity are validated and cleaned for this research. Normalization and input variables Input variables have been constructed out of the aforementioned components by normalizing them to the appropriate ranges. All the input variables and their standard deviations.

- A new column ‘avg’ was formed from averaging the scores of G1, G2 and G3
- A new column ‘grades’ was then formed by classifying the average scores into 5 classes of grading thus: (0, 9.5) – Fail, (9.5, 11.5) -- Sufficient/ Fair, (11.5, 13.5) – Satisfactory, (13.5, 15.5) – good, (15.5, 20) – Excellent.
- Outliers were detected and deleted.
- Data was divided into numerical and categorical datasets.
- Insignificant and useless columns were dropped: 'G1', 'G2', 'G3', 'avg', 'school'
- Dataset was splitted into test and train sets. Target column was set to be ‘grades’ as that is what we will be predicting.
- Data Scaling was done
- Data was divided into binary, nominal and numerical columns.

- A column transformer was then done to fit in encoders: binary columns encoded with Ordinal Encoder, nominal columns with one hot encoder and numerical columns scaled with standard scaler. All from Scikit learn library.
- Cross validation was done.
- Functions were written to combine different classification algorithms to compare metrics, all fitted in a pipeline: 'Logistic Regression', 'Decision Tree Classifier', 'Support Vector Machine', 'Random Forest Classifier', 'AdaBoost Classifier', 'Gradient Boosting Classifier', 'K Neighbors Classifier', 'Gaussian Naive Bayes'.
- Random forest classifier seemed to perform better.

CHAPTER FIVE

RESULT AND DISCUSSION

4.1 Result

Artificial intelligence is no doubt here to stay and to say we have scratched the surface of its very own applicability would be an offense to such a great tool. Now the applicability of artificial intelligence in education, on the other hand, would be if not already one of the key drivers of drastic educational growth and transition. It will overhaul the way education has been handled from knowledge transfer, methods of delivery, predictive approach for futuristic correctness and implementation, Career orientation from data gotten from predictive analysis and so much more.

While AI-fueled arrangements have been in the EdTech space for quite a while, the business has been delayed to take on them. In any case, the pandemic definitely moved the scene, compelling teachers to depend on innovation for virtual learning. Presently, 86% of teachers say innovation ought to be a centrepiece of training. Computer-based intelligence has the ability to upgrade both l earning and instructing, assisting the schooling area with advancing to more readily help learners and educators the same Karandish, D. (2021, June 23). With all the possibility and fancy tricks that arises with Artificial intelligence and the benefits it p possesses in education, there are however some technological laggards who stand on the side of artificial intelligence being a bad influence on education. Some valid points have been raised over the years.

Loss of Skill

When people do certain activities multiple times, they ultimately become better at doing them for that, when teachers are spending years teaching students. They get better at their job day by day and gather much experience. But AI-based tools like Edu-bots scheme ahead by the algorithms which are not influenced by the repetition of activities. So, their experience does enhance the teacher.

Stimulating Technological Addiction

The classroom or school ground has been for eons a place where smartphones and tablets were kept aside for massive concentration by reading from physical textbooks and taking notes with pens and pencils. With the implementation of technology, the day-to-day traditional routine has erupted leaving technological junkies who literally can't leave without their phones.

High cost of delivery and implementation of AI

The technology has never at a point in time been cheap. Getting the foundational gadgets for the implementation of an artificial intelligence system in an academic institution is not cheap and as a result, many institutions shy away from it. To comfortably say that artificial intelligence has taken a root in education it has to be universally accepted. Most privately owned education institutions in Africa prioritize profit over the overall performance of the students due to the high cost of adopting and implementing an artificial intelligence tool.

Human Resource and Technological Know-How

The technological aficionados are somewhat limited in Africa in the department of educational technology. This poses another issue for implementing artificial

intelligence in education in the country. The handful of those that have the knowledge of how to create, manage and manipulate the technology are sometimes tempted by power dynamics as they control the system without them in many cases the system can't be managed properly.

Despite the concerns, a more beneficial counter-effect of AI deployment is the awakening of consciousness among educational institutions with regards to the training and development of educators and administrators. Adopting AI requires investments in infrastructure technologies that will facilitate the adoption and rollout of AI. As well as systems for the professional growth of the educators and administrators who will staff the schools. Institutions of higher learning cannot afford to engage in AI without also making significant financial commitments to supporting technologies and the professional growth of their staff, according to research by Gautam, A. (2019, July 7).

Being a fascinating and emerging technology, artificial intelligence has to face some challenges and risks for acceptance and implementation in secondary schools. Challenges need to be identified to give administrators and stakeholders the benefits of the implementation of artificial intelligence over its few flaws which could be easily corrected or worked through. This research work is aimed at surfacing values from a practical approach and successful implementation of artificial intelligence tools, coupled with the predictive analysis from data of the system that was implemented in the school and how it boost productivity.

CHAPTER SIX

CONCLUSION

Artificial intelligence is no doubt here to stay and to say we have scratched the surface of its very own applicability would be an offense to such a great tool. Now the applicability of artificial intelligence in education, on the other hand, would be if not already one of the key drivers of drastic educational growth and transition. It will overhaul the way education has been handled from knowledge transfer, methods of delivery, predictive approach for futuristic correctness and implementation, Career orientation from data gotten from predictive analysis and so much more.

While AI-fueled arrangements have been in the EdTech space for quite a while, the business has been delayed to take on them. In any case, the pandemic definitely moved the scene, compelling teachers to depend on innovation for virtual learning. Presently, 86% of teachers say innovation ought to be a centerpiece of training. Computer-based intelligence has the ability to upgrade both learning and instructing, assisting the schooling area with advancing to more readily help learners and educators the same Karandish, D. (2021, June 23). With all the possibility and fancy tricks that arises with Artificial intelligence and the benefits it possesses in education, there are however some technological laggards who stand on the side of artificial intelligence being a bad influence on education. Some valid points have been raised over the years.

Loss of skill: When people do certain activities multiple times, they ultimately become better at doing them. For that, when teachers are spending years teaching

students. They get better at their job day by day and gather much experience. But AI-based tools like Edu-bots scheme ahead by the algorithms which are not influenced by the repetition of activities. So, their experience does enhance the teacher.

Stimulating technological addiction: The classroom or school ground has been for eons a place where smartphones and tablets were kept aside for massive concentration by reading from physical textbooks and taking notes with pens and pencils. With the implementation of technology, the day-to-day traditional routine has erupted leaving technological junkies who literally can't leave without their phones.

High cost of delivery and implementation of AI: The technology has never at a point in time been cheap. Getting the foundational gadgets for the implementation of an artificial intelligence system in an academic institution is not cheap and as a result, many institutions shy away from it. To comfortably say that artificial intelligence has taken a root in education it has to be universally accepted. Most privately owned education institutions in Africa and the world prioritize profit over the overall performance of the students due to the high cost of adopting and implementing an artificial intelligence tool.

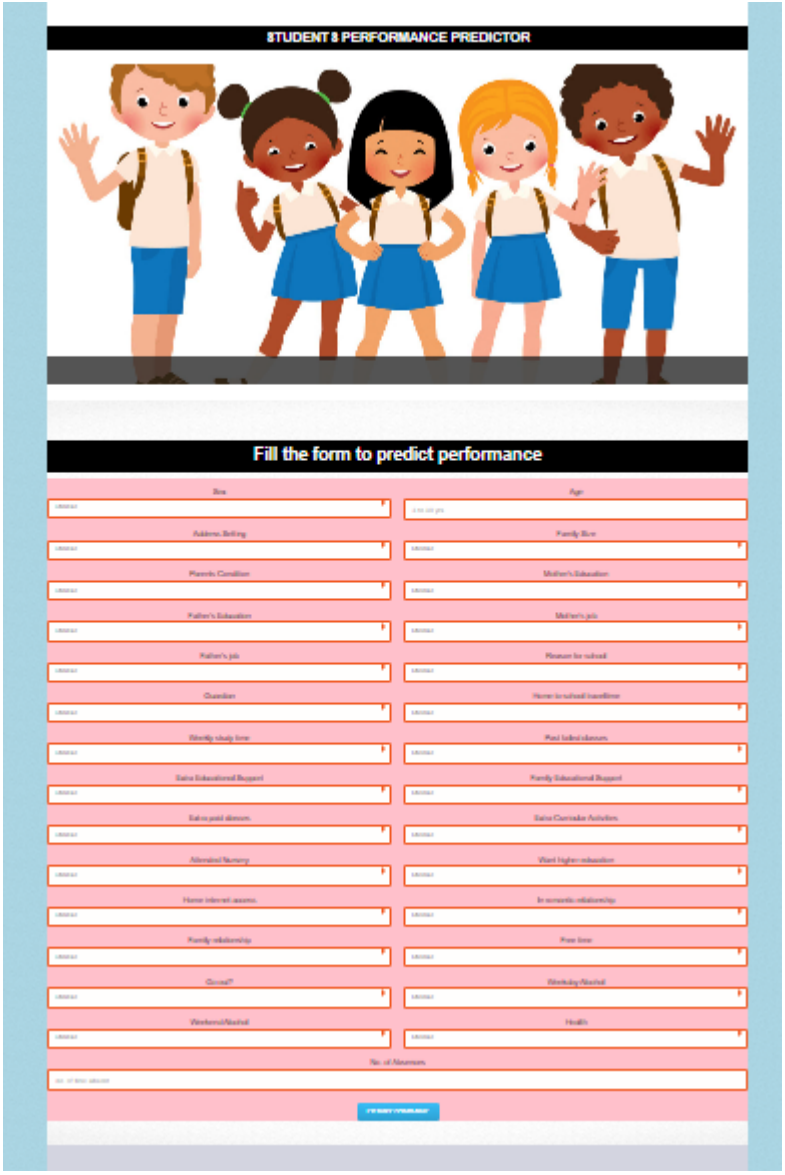
Human resources and technological know-how: The technological aficionados are somewhat limited in some parts of the world in the department of educational technology. This poses another issue for implementing artificial intelligence in education in the country. The handful of those that have the knowledge of how to create, manage and manipulate the technology are sometimes tempted by power

dynamics as they control the system without them in many cases the system can't be managed properly.

Despite the fears, a more positive counter-effect of Artificial intelligence adoption is the onset of awareness among educational organizations in terms of teachers and administrator training and development. Artificial Intelligence adoption comes with its own baggage of investments in technologies that will support the adoption and implementation. As well as Learning and Development programs and platforms for the teachers and administrators who will work in the institutions. There is no way that educational organizations can afford to invest in Artificial intelligence and not invest in ancillary technology and the development of their human workforce Gautam, A. (2019, July 7).

Being a fascinating and emerging technology, artificial intelligence has to face some challenges and risks for acceptance and implementation in secondary schools. Challenges need to be identified to give administrators and stakeholders the benefits of the implementation of artificial intelligence over its few flaws which could be easily corrected or worked through. This research work is aimed at surfacing values from a practical approach and successful implementation in the case of schools around the globe that had implemented artificial intelligence into their system, coupled with the predictive analysis from data of the system that was implemented in the school and how it boost productivity.

APPLICATION DEVELOPMENT



- **4.8 Application Development**

The concluded prediction model was then implemented in an application software to serve as a baseline for future predictions for students from other parts of the world. This application software was built primarily using open-source programming languages and frameworks

- **4.8.1 Front end**

The front-end of the application was built using HTML, CSS and Javascript, and it was built to be responsive and supportive of different screen sizes across Laptops and smartphones alike.The pictures were gotten from Google images and applied and fitted to the respective places in the app.The Front-end contains different methods including Divs, Navs, Headers, Footer,etc.The application features an easy to use interface that allows users to enter the variables as relates to them.

- **4.8.2 Back end**
- The flask framework, which was used to build the backend and is based on the Python programming language, enables efficient processing of data and data techniques. Here, the user's favourite modeller was used to assist with predictions and suggestions.
- Python-based Flask is a web framework used for quick and simple web application creation as well as for simple frontend and backend application configuration. It allows developers total control over how to access data. The Jinja templating engine and Werkzeug's (WSGI) toolbox are the foundations of Flask. Flask is intended for quick creation of REST APIs. When creating a blog website or any other kind of business website, Flask offers a variety of libraries, tools, modules, and functions like managing user requests, routing, sessions, form validation, etc. The implementation of Flask is likewise quite simple to understand. You can get going on this with only a few lines of code. Top tech businesses like Netflix, Reddit, Mozilla, and others utilise Flask.
- To train the model on the random forest classifier, functions were built.

Functions were written to find the best contributors to the predictions to be implemented in the app API as what to do to have better results and performance.

- **4.8.3 Deployment**

The application was deployed on Heroku (Salesforce cloud), thus allowing for an all-time access and also opening up future access for improvements on the modeler. This is because as users put in their data for prediction, their data is added to a separate data pool that would in the future be used to train the dataset for better improvement and prediction

REFERENCE

Al-Shuaibi, A. (2014). *The Importance of Education*. [online] ResearchGate. Available at:https://www.researchgate.net/publication/260075970_The_Importance_of_Education.

Azeez, B. (2014). *Reliable low latency I/O in torus-based interconnection networks*. [College Station, Tex.]: [Texas A & M University].

Barrett, B. (2017). *Globalization and change in higher education : the political economy of policy reform in Europe*. Cham, Switzerland: Palgrave Macmillan.

Chung, H.-S. and Kim, J.-M. (2014a). Semantic Model of Syllabus and Learning Ontology for Intelligent Learning System. *Computational Collective Intelligence. Technologies and Applications*, pp.175–183. doi:10.1007/978-3-319-11289-3_18.

Chung, H.-S. and Kim, J.-M. (2014b). Semantic Model of Syllabus and Learning Ontology for Intelligent Learning System. *Computational Collective Intelligence. Technologies and Applications*, pp.175–183. doi:10.1007/978-3-319-11289-3_18.

Computer Science, Engineering and Applications (ICCSEA 2012), May 25-27, 2012, New Delhi, India. Volume 1. Berlin ; New York: Springer.:

Correia, L.M., ReisL.P. and CascalhoJ. (2013). *Progress in artificial intelligence : 16th Portuguese Conference on Artificial Intelligence, EPIA 2013, Angra do Heroísmo, Azores, Portugal, September 9-12, 2013. Proceedings*. Heidelberg: Springer.

IEEE Guide for Wireless Access in Vehicular Environments (WAVE) Architecture - Redline. (2019). New York, USA: IEEE.

Irby, Beverly J., et al. *The Handbook of Educational Theories*. Charlotte, N.C., Information Age Publishing, 2013.

Kitchenham, B. and Mendes, E. (2004). Software productivity measurement using multiple size measures. *IEEE Transactions on Software Engineering*, 30(12), pp.1023–1035. doi:10.1109/tse.2004.104.

Shmueli and Koppius (2011). Predictive Analytics in Information Systems Research. *MIS Quarterly*, 35(3), p.553. doi:10.2307/23042796.

United Nations Statistics Division (2019). — *SDG Indicators*. [online] Un.org. Available at: https://unstats.un.org/sdgs/report/2019/goal-04/.

Wyld, D.C., Zizka, J. and Dhinaharan Nagamalai (2012). *Advances in computer science, engineering & applications : proceedings of the Second International Conference on*

Kitchenham, B. and Mendes, E. (2004). Software productivity measurement using multiple size me

asures. *IEEE Transactions on Software Engineering*, 30(12), pp.1023–1035. doi:10.1109/tse.2004.104.

Shmueli and Koppius (2011). Predictive Analytics in Information Systems Research. *MIS Quarterly*, 35(3), p.553. doi:10.2307/23042796.

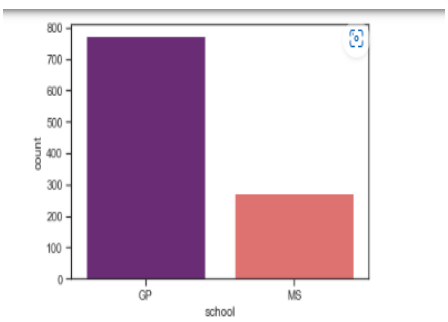
United Nations Statistics Division (2019). — *SDG Indicators*. [online] Un.org. Available at: https://unstats.un.org/sdgs/report/2019/goal-04/.

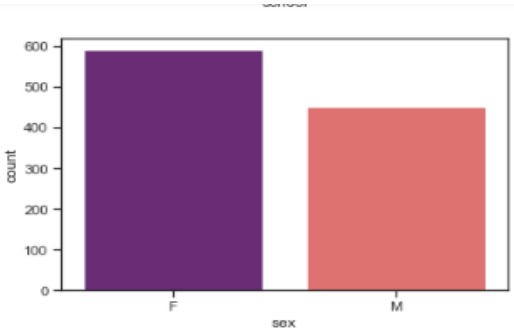
Wyld, D.C., Zizka, J. and Dhinaharan Nagamalai (2012). *Advances in computer science, engineering & applications : proceedings of the Second International Conference on*

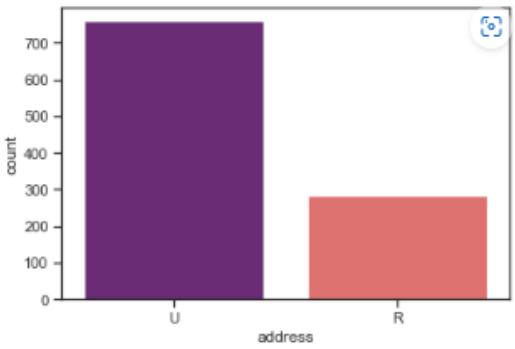
.

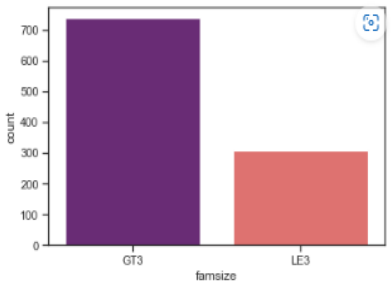
APPENDIX A

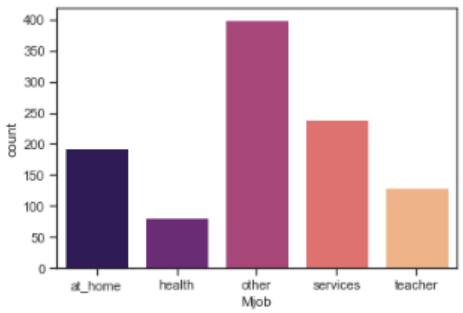
Categorical variable distribution

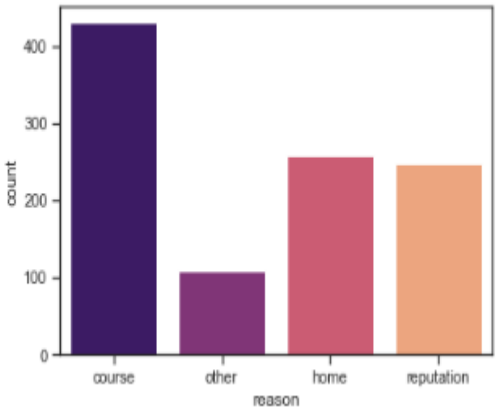


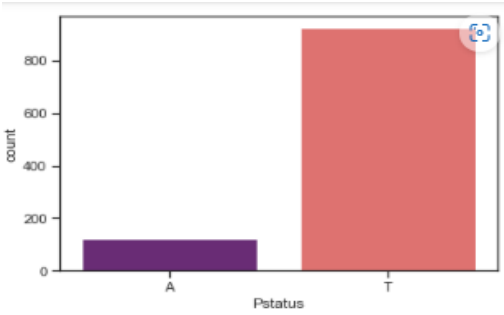


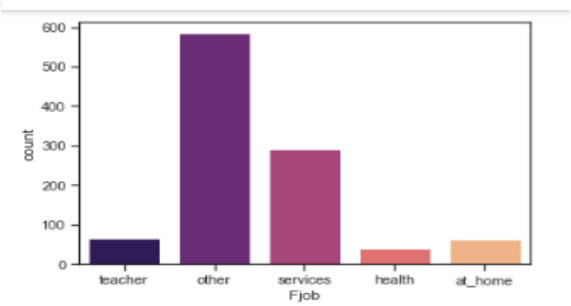


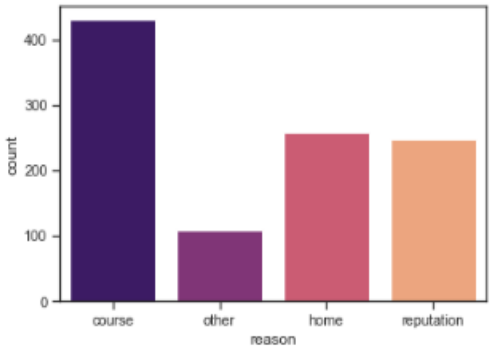


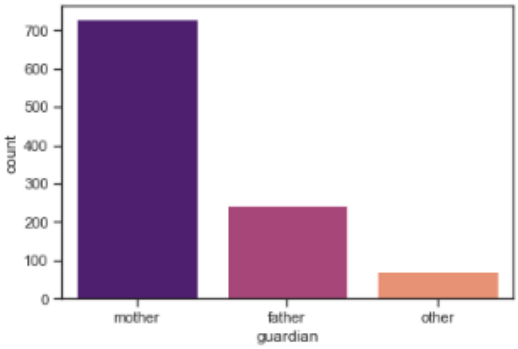


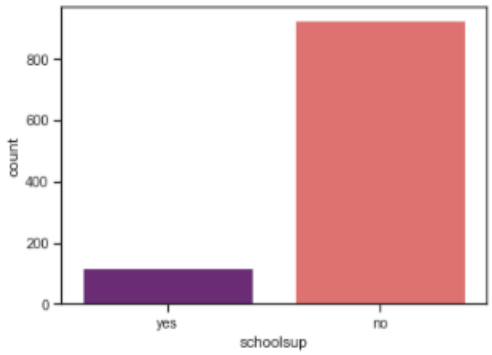


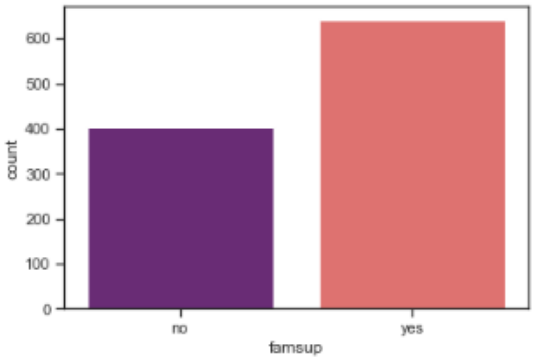

count
800
600
400
200
0
A
T
Pstatus

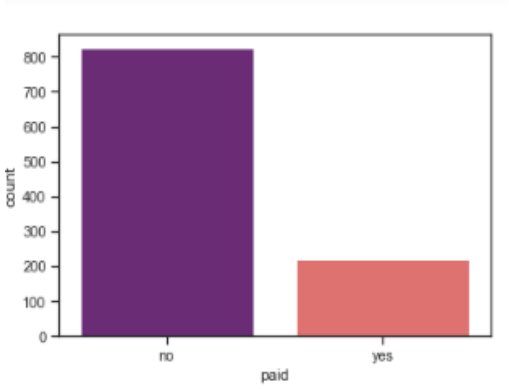

count
600
500
400
300
200
100
0
teacher
other
services
health
at_home
Fjob

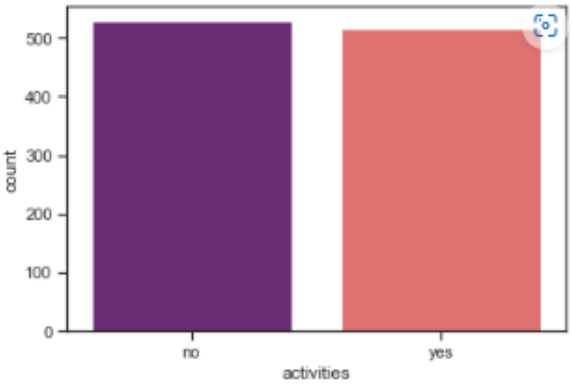

count
400
300
200
100
0
course
other
home
reputation
reason

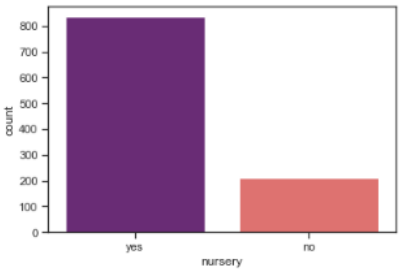

count
700
600
500
400
300
200
100
0
mother
father
other
guardian

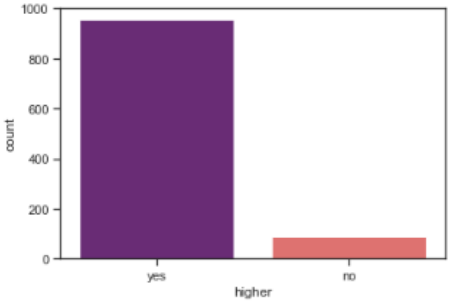

count
800
600
400
200
0
yes
no
schoolsup

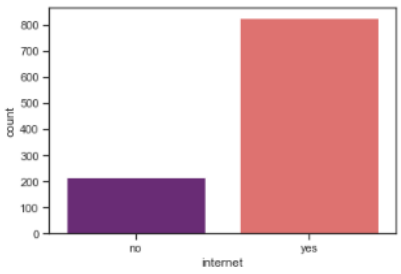

count
600
500
400
300
200
100
0
no
yes
famsup

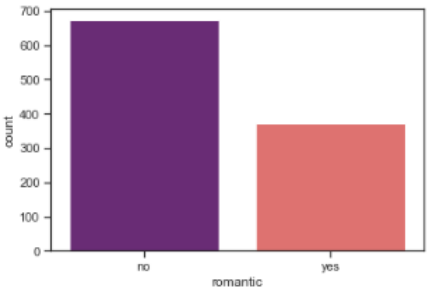

count
800
700
600
500
400
300
200
100
0
no
yes
paid


count
500
400
300
200
100
0
no
yes
activities

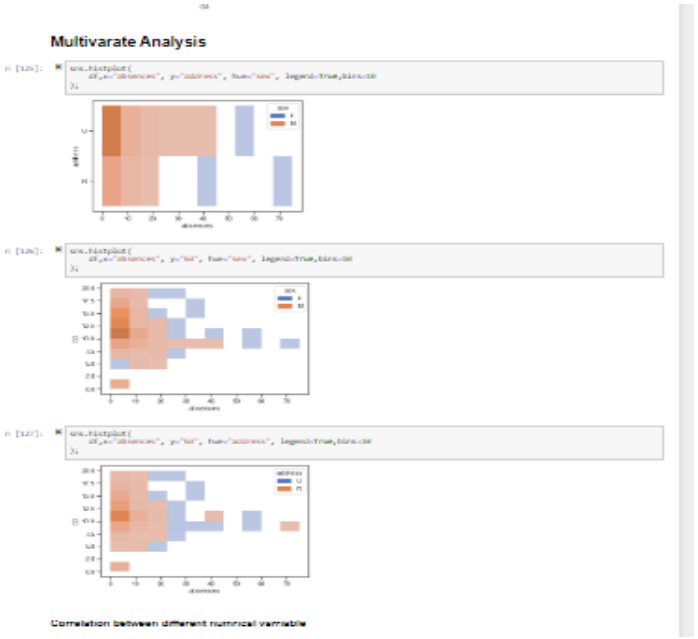

count
800
700
600
500
400
300
200
100
0
yes
no
nursery


count
1000
800
600
400
200
0
yes
no
higher

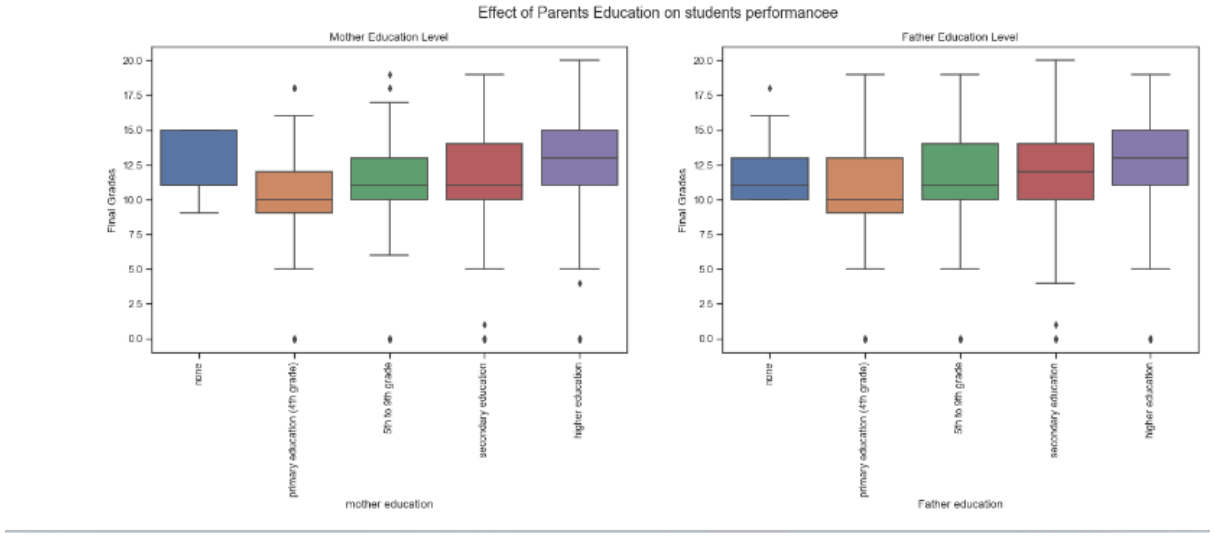

count
800
700
600
500
400
300
200
100
0
no
yes
internet

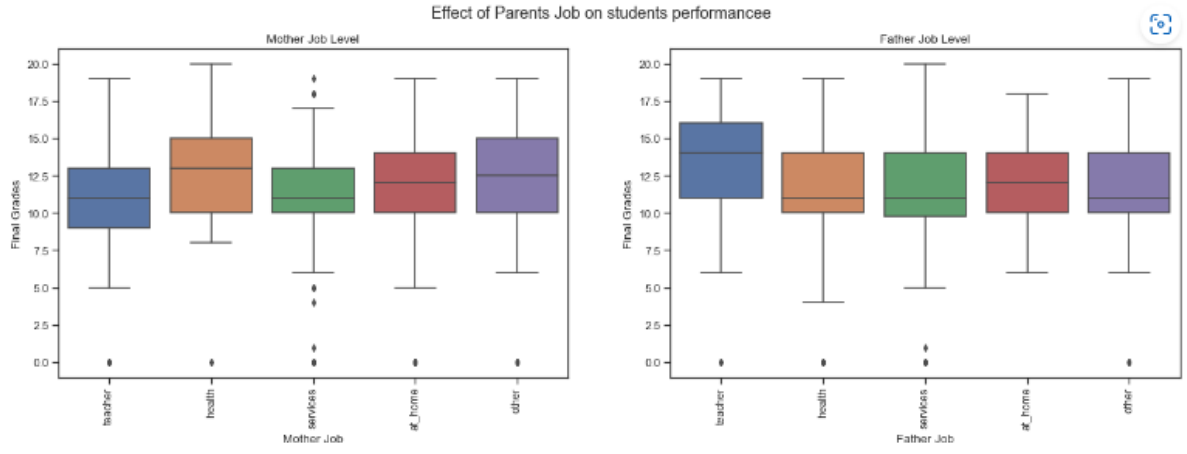

count
700
600
500
400
300
200
100
0
no
yes
romantic

Multivarate Analysis



Correlation between different numrical varriable





APPENDIX B

Sources of the Data set

https://archive.ics.uci.edu/ml/machine-learning-databases/00320/