የኢፌዲሪ የቴክኒክና ሙያ ስልጠና ኢንስቲትዩት
FDRE TECHNICAL & VOCATIONAL TRAINING INSTITUTE

TECHNICAL AND VOCATIONAL TRAINING INSTITUTE (TVTI)

School of Graduate Studies

Division of Electrical Electronics and ICT

DEPARTMENT OF ICT

POSTGRADUATE PROGRAM

Model Development based on Demographic and Psychographic factor to improve performance in case of TVET using Machine Learning

By

Habtom Atsebeha Berhe

Advisor: Tibebe Beshah (Dr.)

August, 2023

Addis Ababa, Ethiopia

DECLARATION

I, the undersigned, declare that this thesis entitled: “Model Development based on Demographic and Psychographic factor to improve performance in case of TVET” is my original work. I have undertaken the research work independently with the guidance and support of the research advisor. This study has not been submitted for any degree or diploma program in this or any other institutions and that all sources of materials used for the thesis has been duly acknowledged.

Declared by

Name: Habtom Atsebeha Berhe

Signature: ________________________

Department: Information Communication Technology

Date: August/2023

This thesis has been submitted for examination with my approval advisor.

Tibebe Beshah (Dr.) _____________________

Advisor’s Name Signature

CERTIFICATION

This is to certify that the thesis prepared by Habtom Atsebeha Berhe, entitled "Model Development based on Demographic and Psychographic factor to improve performance in case of TVET" and submitted in partial fulfillment of the requirements for the Degree of Masters of Arts in MSc in Information Communication Technology complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Name of Candidate:	Habtom Atsebeha Berhe
Signature:	_______________
Date:	August/2023
Name of Advisor:	Tibebe Beshah (Dr.)
Signature:	_______________
Date:	August/2023

Signature of Board of Examiner`s:

____________________	____________________	____________________
External Examiner	Signature	Date
____________________	____________________	____________________
Internal Examiner	Signature	Date
____________________	____________________	____________________
Dean, SGS	Signature	Date

Faculty of Electrical Electronics and ICT

ACKNOWLEDGMENT

I would like to express my sincere gratitude to my advisor Tibebe Beshah (Dr.), for his invaluable guidance and support throughout my master's program. His expertise and encouragement helped me to complete this research and write this thesis.

I would also like to emphasize my respect and gratitude to Polytechnics College Colleagues those who shared me the data and collaboration working with me who gives valuable information used for my thesis conduct without hesitation.

I also want to express my gratitude to my family and friends for their love and support throughout this process. This voyage would not have been possible without them.

Finally, I'd like to express my gratitude to all of the study subjects for their cooperation and readiness to share their insights. Without their help, this job would not have been feasible.

.

Table of Contents

List of Tables

List of Figures

ACRONYMS and ABBREVIATIONS

TVET – Technical Vocational Educational and Training

ICT – Information Communication Technology

ML – Machine Learning

AI – Artificial Intelligence

EGSECE – Ethiopian General Secondary Education Certificate Examination

CSV – Comma-separated Value

SVM – Support Vector Machine

CA – Classification Accuracy

ABSTRACT

TVET is important for the growth of human resources and, consequently, for the advancement and prosperity of a community. In this thesis, the factors that significantly affect TVET students' academic achievement and performance were modeled. The academic performance of TVET students has been found to be significantly influenced by a number of factors, including gender, age, monthly family income, study hours, stimulant use during the course of the study, and satisfaction with the area of study placement. According to this study, lower academic accomplishment was associated with a variety of factors, including having a negative opinion of TVET, being female, having a low family income, studying for a shorter amount of time, using stimulants while studying, and not being content with one's field of study. This thesis was conduct to find out the role of demographic and psychographic factors on the effectiveness of the performance of the student in a TVET. The students from those Polytechnic Colleges were randomly select in order to evaluate the effectiveness of performance of the student’s learning process. The primary data source is from the college’s registrar record of the student and survey using Google Form applications that was distributed to the students. Data was analyzed using Python Programming Language.

The experiment involving Support Vector Machines (SVM) has yielded valuable and productive outcomes. Achieving an 84% accuracy rate provides substantial support for the effectiveness of SVM in forecasting student performance. This process highlights the significance of harnessing advanced algorithms to decipher intricate educational data, ultimately facilitating deeper understandings of the elements influencing academic achievements.

Keywords: Performance, Demographic factors, psychographic variables, machine learning, SVM, TVET.

CHAPTER ONE

INTRODUCTION

1.1. Background of the Study

The focus of government policies on education, the strategies for socioeconomic development, the reports by international organizations on sustainable development, and the solutions to the current problems with socioeconomic conditions that favor Technical Vocational Educational and Training (TVET) are all largely in agreement. According to Makinde and Rafiu (2020), TVET is a method of preparing people for active participation in jobs that have practical significance and serve as reliable sources of skilled labor. According to Oviawe (2018), applied learning is prioritized over academic learning, skills over knowledge, and practice over theory in TVET. *[2]*

The quick development of information technology has made learning more effective. A moment of unimaginable change in universities all over the world has begun as a result of the rapid advancement of information, communication, and technology (ICT). With the development of technology and learning, new practices, techniques for conveying information, and communities of learners throughout the world are becoming available. How people would use ICT resources has been impacted by or predicted using demographic factors. Age, gender, education level, income, and skills are just a few of the demographic variables that are frequently mentioned as having an impact on how ICT is used. Demographic parameters including age, gender, teaching experience, the subject(s) taught, computer use experience, and educational background were taken into account in order to employ ICT resources in the classroom efficiently.

For a few years now, machine learning has been applied in educational informatics. To evaluate the performance of the TVET students, new machine learning methods were applied. *[3]*

1.2. Statement of the problem

Our country's national development and economic progress are dependent on high-quality training. Ethiopia is one of the developing countries in the world, so quality training has no choice to bring sustainable development. TVET sectors should stand to run quality training to attain the development of one country by monitoring and controlling student academic status periodically. But nowadays, the existing traditional work is not sufficient to exactly predict students ‘academic status and related patterns. As a result, forecasting students' performance assists trainers and academic institutions in planning teaching methodologies and learning processes, and the entire bureau of TVET recognizes each polytechnic colleges standing for resource allocation.

The Ministry of Education claims that a serious shortage of crucial data and information regarding TVET issues that is required to support planning, monitoring, and evaluation in the TVET system is now impeding the development of TVET. Currently, international professionals contribute the majority of TVET-related research. Ethiopia's research capabilities are still weak. As a result, Ethiopia needs to develop top-notch domestic TVET research capabilities if it hopes to eventually become independent. Education professionals have long been interested in their students' academic success. Academic greatness is a factor that educators and researchers have long been interested in recognizing and comprehending. Numerous studies have linked demographic, socioeconomic, family, and educational aspects to pupils' academic achievement.

The relationship between students' socio-demographics, academic preparation, and school features in TVET and their accomplishments is only the focus of a small number of scholars at the moment. The aim of this study was to identify the factors that affect academic success among polytechnic college students.

What key demographic and psychographic factors have an impact on academic achievement at polytechnic colleges? How do school characteristics affect academic performance?

1.3. Research Questions

To that purpose, the following questions have been addressed in this thesis:

- What possible traits of demographic and psychographic behaviors could be used to assess students' performance?
- What major demographic and psychographic factors significantly affect students' academic performance at TVET?
- What would be the best model that can classify TVET students' performance?

1.4. Objective of the Study

1.4.1. General Objective

General objective of this study is, to investigate the role of demographic and psychographic factors on the success of TVET student.

1.4.2. Specific Objectives

To achieve the general objectives, we will conduct the following activities.

- To Identify potential demographic and psychographic factors that could determine students success at TVET
- To select best algorithms for the experimentation
- To conduct experiment and build classification model that classifies a student's performance
- Developing a prototype to demonstrate the practicability of the solution
- To evaluate the performance of the prototype

1.5. Scope and Limitation of the Study

The scope of the research will cover all governmental TVET colleges in Addis Ababa. Many factors affect the success of this research, the most common ones are:

- Missing of very important attributes in the student performance roster form
- This study will not cover other regions and education sectors

1.6. Significance of the study

The purpose of the study is to improvement in academic performance of a student provides the general background of how a student is progressing. It also represents the amount of knowledge and skills developed by students. Students display their qualities of learning in academics by writing tests, involving in class/home works, writing examinations.

The first training set for a supervised learning system to predict whether a given student will ultimately pass or fail a certain module used key student demographics (such as age, sex, place of residence, etc.) and their grades on written assignments.

The results of the current study have a number of ramifications for comprehending the important variables affecting students' success and enhancing students' academic performance at Polytechnic Colleges.

The results of the study will be applied by:

- To improve students' interest in and success in school, several stakeholders, including TVET teachers, administrative teams at TVET institutions, policymakers, etc., must evaluate and effectively apply these findings.
- The study's conclusions will also be valuable for upcoming TVET research and other academic investigations.

1.7. Motivation of the Study

We chose this topic for academic performance research because quality education is a vital aspect of national development and economic prosperity in any country, particularly in a developing country like Ethiopia. As a result, it is Ethiopia's most pressing concern. Many students are attending the school and such students are getting lower achievers due to many reasons. To evaluate such a large amount of students using the traditional method is backward.
In this research study, Due to the widespread availability of vast amounts of data and the looming need to turn that data into valuable information and knowledge, data mining has gotten a lot of attention in the information industry and society as a whole in recent years. The information and knowledge gained can be used for applications ranging from education

performance, market analysis, production control, health information, disaster management, and science exploration. Therefore, the use of data mining in educational settings aids in the identification of determinants of academic achievement and better decision-making.

Most of these factors or problems are motivated us to discover unexpected patterns of knowledge using the existing data in schools to recommend ideas for Addis Ababa polytechnic colleges regarding the academic performance.

1.8. Organization of the thesis

The following five chapters make up this thesis: As an introduction, the first chapter gives background information on the research project, a summary of the subject matter, the objective of the study, its scope, and significance, as well as the technique employed. The second chapter reviews the pertinent research on the variables that affect student performance, the approaches and strategies employed, and how they are applied in the educational setting. The third chapter is devoted to explaining the methods, strategies, and SVM tools used in the study. This chapter covers topics such as the choice and preparation of data for use in the research process, as well as the operation of the algorithm underlying the technique. The thorough description of the data, model designs, and performance evaluations of the produced models are covered in length in the fourth chapter. These topics were chosen to demonstrate the work that has to be done to produce a better-quality dataset that is prepared to use data mining tools and techniques. In light of this, preprocessing activities such as data cleansing, transformation, and attribute selection are covered. Additionally, it describes the testing, performance assessment, and analysis of the outcome using particular hybrid data mining methodologies and algorithms.

The fifth chapter, which comes after the researcher has cited references, focuses on drawing conclusions and offering suggestions for future research topics.

CHAPTER TWO

LITERATURE REVIEW AND RELATED WORKS

2.1. Overview

There isn't a single, widely acknowledged definition of technical and vocational education and training (TVET). It's unclear if it refers to post-compulsory education alone or also includes work-based training. The diversity of courses and skills that help students get ready for the workforce can be broadly defined as TVET. According to P. Smith in "Building a world of learning for all," Prospects, vol. 36, no. 1, pp. 5-7, 2006, if TVET is the key to society's doors of greater equity, justice, and poverty reduction, then education is the key to economic and social growth. TVET's main goals are to enable self-employment and serve as a bridge for people to enter the workforce. Technical and vocational education and training typically trains students for employment that requires manual or practical labor, is usually non-theoretical, and is associated with a particular trade, profession, or vocation, hence the student's term.

Computers can learn without being programmed, thanks to a type of artificial intelligence called machine learning. The three main categories of machine learning are reinforcement learning, unsupervised learning, and supervised learning. The basic objective of supervised learning is to build a model from labeled or training data that will enable us to predict the behavior of new data. It is possible to further divide supervised learning into two types of tasks: classification tasks, where the outcome is projected to be a categorical value, and regression tasks, where the outcome is predicted to be a continuous value. The objective of reinforcement learning is to create a system or agent that gets better at what it does as a result of interactions with the environment. This method includes rewards and penalties, although it is quite similar to supervised learning. Unsupervised learning is the process of extracting useful information from unlabeled data without having any prior knowledge of the material.

With the use of machine learning (ML), which is a form of artificial intelligence (AI), software programs can predict outcomes more accurately without having to be explicitly instructed to do so. In order to forecast new output values, machine learning algorithms use historical data as input. Machine learning is frequently used in recommendation engines. Business process automation (BPA), predictive maintenance, spam filtering, malware threat detection, and fraud detection are a few additional common uses. Machine learning is significant because it aids in the development of new goods and provides businesses with a picture of trends in consumer behavior and operational business patterns. A significant portion of the operations of many of today's top businesses, like Facebook, Google, and Uber, revolve around machine learning. For many businesses, machine learning has emerged as a key competitive differentiation.

Decision trees are a supervised learning method that can be applied to classification and regression problems; however, they are most frequently employed to address classification issues. It is a tree-structured classifier, where internal nodes stand in for the dataset's features, branches for the rules of classification, and each leaf node for the result.

The decision node and leaf node are the decision tree's two nodes. Decision nodes are used to produce decisions and have a large number of branches, as opposed to leaf nodes, which are the outcomes of decisions and do not have any additional branches. The test or decision-making process is carried out using the features of the provided dataset.

It is a visual representation of finding every possible solution to a choice or problem under preset circumstances. Because it starts with the root node and expands on succeeding branches to build a structure resembling a tree, it is known as a decision tree.

A tree is built using the CART algorithm, which stands for Classification and Regression Tree Algorithm. A decision tree just asks a question, and depending on the answer (yes or no), it divides the tree into subtrees.

2.2. Academic Performance

The study's dependent variable, the academic success of the TVET students, was acquired from the Polytechnic College's registrar's office at the conclusion of the academic year. Following an analysis of relevant literature, sociodemographic factors such as the student's age, gender, training sector, department, level, batch, home address, family income, parental employment status, the highest level of education in the family, the student's monthly income, and marital status, as well as academic factors such as the student's EGSECE score, study hours per day, study location, English language proficiency, class size, class participation, and satisfactory ratings, are taken into consideration.

The Alan, W. [1] model confirms that, like all students receiving special education services, the academic performance of visually impaired students in the general education classroom is monitored annually, and that performance is very good in comparison to students with other impairments. This is only possible if the unique needs of each student in the classroom are met. If the instructional materials and teaching methods are altered or customized to the needs of the particular learner, learners with visual impairments can perform exceptionally well. For instance, if raised or embossed prints are employed as a teaching tool, the student can do well if his or her educational needs can be addressed through touch. The lecturers have a significant role to play in helping students with visual impairments perform better in the classroom.

2.3. Challenges of TVET

There is no doubt that TVET has a significant economic impact on a country. Two-thirds of the global workforce, it is estimated, has been trained by TVET teachers and trainers. Despite its importance, W. Bauer, "TVET instructors and teachers in Germany," in International Perspectives on Teachers and Lecturers in Technical and Vocational Education, P. Grollman and

F. Ranuer, Eds., Springer, Dordrecht, (e Netherlands, pp. 123–158, 2007, the social acceptance of TVET is often poor, and the status of TVET teachers and trainers as professionals is far from established. TVET continues to be considered a "second-class" education by many parents and students. One justification is that while TVET provides training, employment is not guaranteed. Students will be less inclined to enroll in vocational institutions as a result of these societal prejudices regarding this type of education.

TVET programs as a result of their constrained academic possibilities and low social standing. They explained the stigmatization of TVET in Ghana and other African nations as a result of colonial education's emphasis on the humanities and a postcolonial push to boost the share of intellectuals and white-collar employees.

The authors in Nigeria came to similar conclusions, finding that students did not prioritize TVET as their initial job choice due to its lower social standing and prestige in their society, which was supported by family and peers. The long-standing issues in Ethiopia were detrimental to students and TVET education. Students in Ethiopia who are unable to take part in the university preparation program because of their lower EGSECE results or because they did not fulfill the requirements for admission to the preparatory program, which is one indicator, can do so through TVET programs.

2.4. Factors Influencing Academic Achievement of Students

The academic success of students in post-secondary (i.e., tertiary) education has been linked to a number of elements. Among these are self-efficacy, anxiety, stress, health, study habits, social integration, academic support, course preferences, attitudes, prior academic success, student backgrounds, and levels of motivation, financial assistance, Age, gender, and intellectual ability are all factors in educational methods. Research by [16] found that gender, the number of courses taken, past performance, marital status, and age all have a big impact on academic achievement. They also notice how the various traits of the students interact with one another and with their academic achievement.

Accordingly, there is a sizable gradient between the educational attainment of each parent's child and their own. Moreover, this is backed by Students who skipped class on a particular day were shown to be considerably more likely than those who were present to get questions about the topic taught that day wrong. Additionally, numerous studies have revealed that peer influence has an impact on student achievement and that its effects are stronger than those of close family. [17 - 22]

2.5. Gap Analysis

TVET is very important for one’s country development especially for developing like Ethiopia. But there are factors that can impact the need and success of TVET students. Unfortunately, little research on the demographic and psychographic factors at TVET colleges has since been done.

The aim of this study was figure out the impact of demographic factors such as age, gender, etc. Therefore, this paper aims to examine whether student levels of satisfaction, or dissatisfaction and their challenge of success at TVET colleges differ according to selected demographic factors that define different categories of students at TVET colleges.

2.6. Related Works

This section of paper reviews the works that are similar to the presented work. The recent research works on the area that attempt to find good techniques for Model development based on demographic and psychographic factor to improve performance in case of TVET includes the works of Kidane Getahun, A. The study's objective is to identify the critical elements influencing students' academic development at polytechnic colleges. The typical Bahir Dar Polytechnic College students enrolling for the 2019–2020 academic year were the study's target demographic. 536 people took part in a cross-sectional survey using a stratified random sample. To analyze quantitative data, the author used WinBUGS 1.4 and SPSS version 25.

These studies by S. KOTSIANTIS, C. PIERRAKEAS, and P. PINTELAS suggested that the capacity to forecast a student's performance could be helpful in a variety of contexts related to university-level distance learning. Both the significant demographic characteristics of students and their performance on a few written assignments can be used to train supervised machine learning algorithms. The learning algorithm may therefore be able to predict the performance of incoming students, making it a useful tool for identifying anticipated underachievers. The goal of this study is to compare a few state-of-the-art learning algorithms. Two studies made use of six algorithms that were developed utilizing data sets provided by the Hellenic Open University. Among other noteworthy results, the Naïve Bayes algorithm was demonstrated to be the most appropriate to be used for the development of a software assistance tool. It has more than acceptable accuracy, an extraordinary level of overall sensitivity, and is the simplest method to implement.

These studies were conducted by Bressler, Linda A., Bressler, Martin S., and Bressler, Mark E. As online education continues to grow in importance as a method of delivering courses, academics pay more attention to the system's effectiveness and look into what influences whether or not students succeed. Studies analyzing student performance, in addition to counting the number of students enrolled in online courses or complete programs, offer more insight into creating best practices in higher education. Studies that look at the function of psychological elements and how they affect student success are becoming more and more common. Age, hope, and self-efficacy were three important psychological variables that the authors of this study looked into in order to see how they would affect the number of courses taken and grade performance.

The CRISP-DM methodology, which is governed by a series of steps, was employed by this researcher, Fatima Sayed Alsheikh. These stages are: business understanding; data understanding; data preparation; modeling evaluation; and deployment.

There have been numerous studies done to determine a model that could be effective for predicting students' academic success based on their social characteristics, employing Bayesian classifiers (naive Bayes, Bayes net), decision trees (j48), and random forests as classification techniques. According to the experimental findings, J48 is the best method for data classification. Additionally, it demonstrated how much social issues affect pupils' academic achievement.

2017 (Hamoud and Hashim) Based on Bayes algorithms, this study developed a model of a student's successful prediction and made recommendations for the optimal method based on performance information. In this model, the responses to the students' questionnaire were employed along with two built-in Bayes algorithms (naive Bayes and Bayes network). The questionnaire has 62 questions that cover the areas that have the biggest impact on students' performance. The inquiries deal with relationships, social interaction, academic achievement, and general health. 161 students responded to the survey, which was created using a Google form and open-source software (LimeSurvey). Weka 3.8 is the program used to create this model. There are two steps to the total model design process. Finding the questions that are most linked to the final class comes first, followed by applying algorithms and selecting the best one. Based on specific performance information, a comparison is made between these two Bayes algorithms.

In contrast to the Naive Bayes classifier, which has TP rate 0.667, FP rate 0.297, precision 0.706, and recall of 0.667, the Bayes net classifier has TP rate 0.655, FP rate 0.432, precision 0.643, and recall of 0.655. The naive Bayes algorithm is ultimately chosen as the best option for the accurate predictions made by pupils.

Singh and Kaur, 2016) By reducing the varied effects of three elements (the first being parameters that determine student performance, the second being data mining methodologies, and the third being data mining tools) on students' performance, the researchers want to maintain the educational quality of the institute. In this study, the WEKA tool is used to predict student performance using the Naive Byes and J48 decision tree classification algorithms. The researcher gains knowledge that describes student performance by using data mining techniques on student data. This information will aid in raising the standard of instruction, enhancing student performance, and lowering the failure rate. All of these things will aid in raising the institute's caliber. In order to acquire categorization results, this research is conducted on fictitious data. There are 52 instances in this data set, and each instance has 9 properties. J48 and Naive Bays yield accuracy of 61.53% and 63.59%, respectively.

2016's Al-Barrak and Al-Razgan In this study, the researcher used educational data mining to predict students' final GPA based on their grades in prior courses. The researcher collected students' transcript data from the database management system for female students who graduated from King Saud University's Computer Sciences College in 2012, totaling 236 students. Their overall course grades and ultimate GPA were included in the statistics. The researcher discovered classification rules to predict students' final GPA based on their grades in required courses after preprocessing the data. The researcher also evaluated the most crucial courses in the study plan that have a significant impact on the students' final GPA.

(Mueen, 2016) The main objective of this study is to apply data mining techniques to predict and analyze students' academic performance based on their academic record and forum participation. Data about students was gathered from two undergraduate classes by the researcher. Three classifiers were used by the researcher to compare, evaluate, and analyze the dataset. Naive Bayes, multilayer perception, and C4.5 (J48) are the three classifiers in question. The 38 accessible attributes were used to test each of the three classifiers. The dataset was partitioned into ten equal-sized subgroups at random by the researcher using tenfold cross-validation. Naive Bayes was found to perform better than the other two. Predictive power is demonstrated by the fact that Nave Bayes is also the winner in precision. Recall, which shows sensitivity, says. Overall prediction accuracy for naive Bayes was 86%. This study helps professors identify students who are likely to fail the course early on. 13 of those pupils can receive extra attention from the teacher, who will also work with them to improve their academic achievement. Numerous studies have been done in this area to find several elements, including student personal factors, family issues, and instructor factors. Student performance is influenced by a variety of elements or a combination of factors.

(Sumitha and Vinothkumar, 2016) The main objective of this research is to explore if it is possible to predict the performance of the student (output) based on the various explanatory (input). Data sets about 300 students were collected, dataset around 250 are being used as training dataset and 50 datasets as test data to design student model. Data are gathered from KLN College of Information Technology's I, II, III, and IV year B.E. CSE students (affiliated with Anna University). The students' actual responses to a questionnaire employed in this process are utilized to describe the connection between their academic performance and their learning behavior. Student demographic information, school information, attendance, cumulative grade point average (CGPA), and final grade from the previous semester are the factors utilized in the questionnaire to assess students' learning and academic behavior. The Naive Bayes, Multilayer Perception (MLP), REP tree, and J48 algorithms are used for categorization. Each classifier is put to use in one of two testing scenarios: using the provided test data or training data. The J48 method offers a maximum accuracy of 97% in efficiently identifying instances when developing the student model. The student model is created in Net Beans using Java coding. (Cheewaprakobkit, 2015) In this study, the researcher used WEKA open source data mining tool to analyze the attributes for predicting undergraduate students‟ academic performance in an international program. The data set comprised of 1,600 student records with 22 attributes of students registered between year 2001 and 2011 at a university in Thailand. Preprocessing included attribute importance analysis. The researcher applied the data set to differentiate classifiers (Decision Tree, Neural Network). Results show that the decision tree classifier achieves high accuracy of 85.188%, which is higher than that of a neural network classifier by 1.313%.

(Ahmad, Ismail and Aziz, 2015)A framework for predicting students‟ academic performance of first year bachelor students in Computer Science course. The data were collected from 8 year period intakes from July 2006/2007 until July 2013/2014 that contains the students‟ demographics, previous academic records, and family background information.

There are 497 records in the data set. To create the best students' academic performance prediction model, student data is subjected to decision tree, naive bayes, and rule-based categorization algorithms. By getting the maximum accuracy score of 71.3%, the experiment's findings demonstrate that the rule-based model is the best one among the others.

(2012) Baradwaj and Pal As there are many methods for classifying data, the decision tree method is used in this study to evaluate student performance. Data from the student management system, including attendance, class tests, seminars, and assignment grades, was gathered to forecast student performance at the end of the semester. The VBS Purvanchal University, Jaunpur (Uttar Pradesh), provided the data set for this study, which covered the MCA (Master of Computer Applications) department's sample methodology from the 2007–2010 academic year. The data's initial size is 50. Students and teachers can enhance student division with the use of this study. This study also aims to pinpoint the students who require extra help in order to lower the fail rate and take the necessary precautions for the examinations scheduled for the next semester.

Table 1: Summary of Related Works

Author, Title & Year	Objective/Purpose	Approaches/ Methodologies	Key Findings	Recommendation & Future Work	Remark
Kidane A. Getahun, "A Bayesian Approach to Investigating Factors Influencing Polytechnic College Students' Academic Achievement." (2022)	The purpose of the study is to discover the major variables affecting students' academic performance at polytechnic colleges.	The major data used in the current investigation was collected via a self-administered questionnaire. Three sections made up the questionnaire: questions gauging students' general happiness with school facilities, questions gauging their academic-related features, and demographic questions.	- The results of this study may be used as input for stakeholders, such as TVET teachers, administrative teams at TVET institutions, policymakers, etc., to assess and correctly incorporate these findings to boost students' interest in and success in school. - The study's findings would be useful for future TVET research as well as for other researchers.	(i) To help female TVET students compete with their male counterparts, they should get exceptional assistance and encouragement. (ii) It would be better for kids to put in more study time and consume fewer stimulants while they are in school. Additionally, teachers and student families ought to help and exhort kids to pay more attention to their studies.	This study tries to identify the elements that affect polytechnic college students' academic performance.
Wasiu Makinde, Tolulope Oluwatosin Bamiro, and others published "Service Quality of Teaching Vocational Education and Training (TVET) and Students Performance in The Federal Polytechnic Ilaro, Nigeria." (2022)	The survey found that the tangibility of TVET is highly ranked in terms of physical facilities, comfortable classrooms, suitable academic resources, and the look of physical facilities. Regarding the assurance of TVET, support staff fairness, lecturers' genuine concern for students, and lecturers' teaching abilities are highly ranked, as are the lecturers' good knowledge, politeness, and security precautions.	The study used the Krejcie and Morgan sampling procedure with a sample size of 375. The study included both primary and secondary sources of data collection. The acquired data were examined using frequency distribution tables, percentages, mean values, and linear regression analysis, among other descriptive and inferential statistics.	At the Federal Polytechnic, Ilaro, Nigeria, student performance was measured using responses to tangibility as a determinant of service quality in technical vocational education and training. The results showed that 144 respondents—or 39.5% of the total sampled respondents—disagreed with the attractiveness of physical facilities for learning at the Federal Polytechnic, while 220 respondents, or 60.5% of the total sampled respondents, agreed that physical facilities are attractive for learning TVET in the institution.	i. Management should emphasize areas of strength in TVET training to boost the institution's performance and global standing. ii. Management should work with private investors to upgrade the institution's physical facilities for TVET training in order to promote the nation's socioeconomic development.	Based on the five SERVQUAL model aspects that were assessed, the study found that students at the Federal Polytechnic Ilaro in Nigeria were satisfied with the TVET service quality.

Author, Title & Year	Objective/Purpose	Approaches/ Methodologies	Key Findings	Recommendation & Future Work	Remark
Deepak Kumar, Verma, Chaman, and Zoltán Illés In real-time automated web applications, "Students Demographic and Geographic Feature Identification Using Machine Learning Techniques" (2022)	The initial goal is to forecast the demographic characteristics of the students, including their gender, course, age group, and institution. The second goal is to locate students' geographic characteristics, such as their locality and home country. The third goal is to create and implement a novel algorithm for making predictions.	In order to accurately identify a student's demographic features (age group, course), we introduced a novel predictive method called Student Demographic Identification (SDI). SDI has been examined using primary, trustworthy samples. The classic machine methods Random Forest (RF), Logistic Regression (LR), and Radial Support Vector Machine (R-SVM) have also been contrasted with SDI. Performance indicators for these classifiers, like accuracy, F1-score, precision, recall, and Matthews Correlation Coefficient (MCC), were all greatly enhanced by the suggested technique.	The study's findings also showed that the GB and RF joint technique was effective at accurately predicting a student's gender with 98% accuracy. The implementation of the presented machine learning models for university classroom response systems is a task for the future. Additionally, social science scholars could detect demography or geographical features using our suggested algorithms and other models.	Our proposed algorithms and other models would be fruitful for social science researchers to identify demography or geographical features. There is also the possibility of designing a novel algorithm to identify a student's gender. Additionally, the SDI algorithm can also be used for other publicly available datasets.	This work offered groundbreakin g, significant predictive algorithms as a preliminary study to forecast students' geographic and demographic characteristics for real-time online applications like Google Forms and Microsoft Forms.
The authors are Kotsiantis, Sotiris, Christos Pierrakeas, and Panagiotis Pintelas. Their study is titled "Predicting Students Performance in Distance Learning Using Machine Learning Techniques." (2004)	For the purposes of our study, the Hellenic Open University's (HOU) informatics program provided the training data for the MLalgorithms. A student may enroll in up to three modules per year at HOU, which uses modules as its fundamental instructional unit. Twelve modules make up the informatics program, which leads to a bachelor's degree. In the academic year 2000–2001, 510 students were officially registered for the informatics course. 498 (90.7%) of the students chose the Introduction to Informatics (INF10) module. Because of this, the authors were able to concentrate on INF10 and solely gather information from the tutors who taught this module.	Six of the most popular machine learning methods, including decision trees (Murthy 1998), neural networks (Mitchell 1997), the Naïve Bayes algorithm (Domingos and Pazzani 1997), instance-based learning algorithms (Aha 1997), logistic regression (Long 1997), and support vector machines (Burges 1998), are applied in order to predict student performance. We will provide a quick overview of these supervised machine learning approaches in the next subsection. Kotsiantis et al. (2002b) provide a thorough description.	The Na ve Bayes algorithm comes out on top in the post-hoc analysis for the overall accuracy (Acc), followed by the logistic regression (72.32%), the BP (72.26%), and the SMO (72.17%).	In a subsequent paper, we plan to investigate whether using more complex discretization methods for WRI marks, such as the one described by Fayyad and Irani (1993) in Predicting Students' Performance 425, could improve classification accuracy. We will also investigate whether choosing a subset of characteristics can increase accuracy because FTOFs do not increase accuracy and the run time of inductive methods increases with the number of attributes.	This study tries to bridge the gap between existing ML approaches and empirical predictions of student performance.

Author, Title & Year	Objective/Purpose	Approaches/ Methodologies	Key Findings	Recommendation & Future Work	Remark
Murat Pojon predicts student performance using machine learning. June 2017	This thesis investigates the use of machine learning algorithms to forecast a student's success or failure. The thesis specifically compares feature engineering techniques and machine learning methodologies in terms of how much they enhance prediction performance.	In this thesis, three different machine learning techniques were employed. They are naive Bayes classification, decision trees, and linear regression. These learning algorithms' predictions were enhanced by the technique of feature engineering, which modifies and chooses the characteristics of a data set.	The findings demonstrate that machine learning may be used to accurately predict student achievement. The most accurate algorithms were naive Bayes classification for the first data set (98%) and decision trees for the second data set (78%). In the data utilized in this investigation, feature engineering was discovered to be a more significant factor in prediction performance than technique selection.	It is important to be aware of the limitations of this study. The study depends on public data sources because there was no access to a specific student data set. Additionally, with fewer than a thousand records each, both data sets were quite modest. A study with access to more complete data may produce more conclusive findings.	The variety of machine learning techniques is another area where future studies might make improvements. In this study, decision trees, linear regression, and naive Bayes classification were all applied. Many techniques, including artificial neural networks and clustering.
Martin S., Mark E., Linda A., and Linda A. Bressler The impact of demographic and psychological factors on online student success January 2015	The influence of psychological factors on student success is now being studied. Age, hope, and self-efficacy were three important psychological variables that the authors of this study looked into in order to see how they would affect the number of courses taken and grade performance.	Undergraduate and post-Bac students enrolled in AIS courses made up the sample. A total of 233 questions were completed, producing 219 usable surveys and a 94% response rate. The findings of the survey showed a varied group of students. Hispanic (n = 44), Pacific Islander (n = 5), Asian (n = 33), African American (n = 49), American Indian (n = 2), Caucasian (n = 77), Mixed (n = 4), and other (n = 5) were the ethnic groups represented among the respondents. The average age of students, who ranged in age from 20 to 55, was 31. Male students made up 49 of the respondents, while female students made up 170 of them. The research study also included 51 individuals who had already earned an undergraduate degree and 161 respondents who were completing 4-year degrees.	Age maturity may also greatly improve pupils' academic achievement, according to the first hypothesis, which examined whether there was a significant association between grade performance and age. A more mature student might have more life, job, and educational experience. The researchers discovered a strong correlation between age and grade performance.	Only students enrolled in online Accounting Information Systems courses were the subject of this study, which was restricted to students at a single university. In addition, more than 75% of the respondents to this survey were female and lived in metropolitan areas. Older female students might be more likely to experience pressure from home and job obligations. This university also accepts commuter students, many of whom are nontraditional and more likely than residential students to be employed or put in greater hours at work.	A factor that educators may be better able to use is self-efficacy. Teachers may be able to spot students who lack confidence early on and devise strategies to boost their self-esteem.

2.7. Conceptual model/Theoretical Framework

Theoretical Foundation The path-smoothing model created by Alan W. will serve as the research's compass. The approach emphasizes the critical methods for constructing a straightforward path for the learner to understand mathematics. According to this paradigm, a teacher should identify the type of issue a student is having in class. Additionally, when teaching, the teacher should try to divide the material into a manageable number of categories and introduce each one at a time. This paradigm makes the implicit premise that students will understand and be able to relate to the nature of the problem from the presentation. The main idea behind this strategy is to create safe passageways for pupils who have vision impairments. Since only one option is frequently given significant consideration, models that show solutions to issues in a sequence of steps are crucial. The model emphasizes that students must work on activities to put the techniques they are taught, which are meant to involve students more actively, into practice. The teacher often categorizes and rates the complexity of these methods. According to the concept, longer-term failure is addressed by repeatedly covering the same or related material throughout the course.

This conceptual framework focuses on how visually impaired students perform academically in polytechnic colleges. The arrows depict the reflected path the researcher will take while conducting the investigation. The framework starts by examining the visually impaired students, then moves on to the best practices for dealing with students who have visual impairments, instructional strategies, specially designed teaching materials, and the attitude of the lecturer toward students who have visual impairments. The academic performance of students with vision impairments will then be the conclusion.

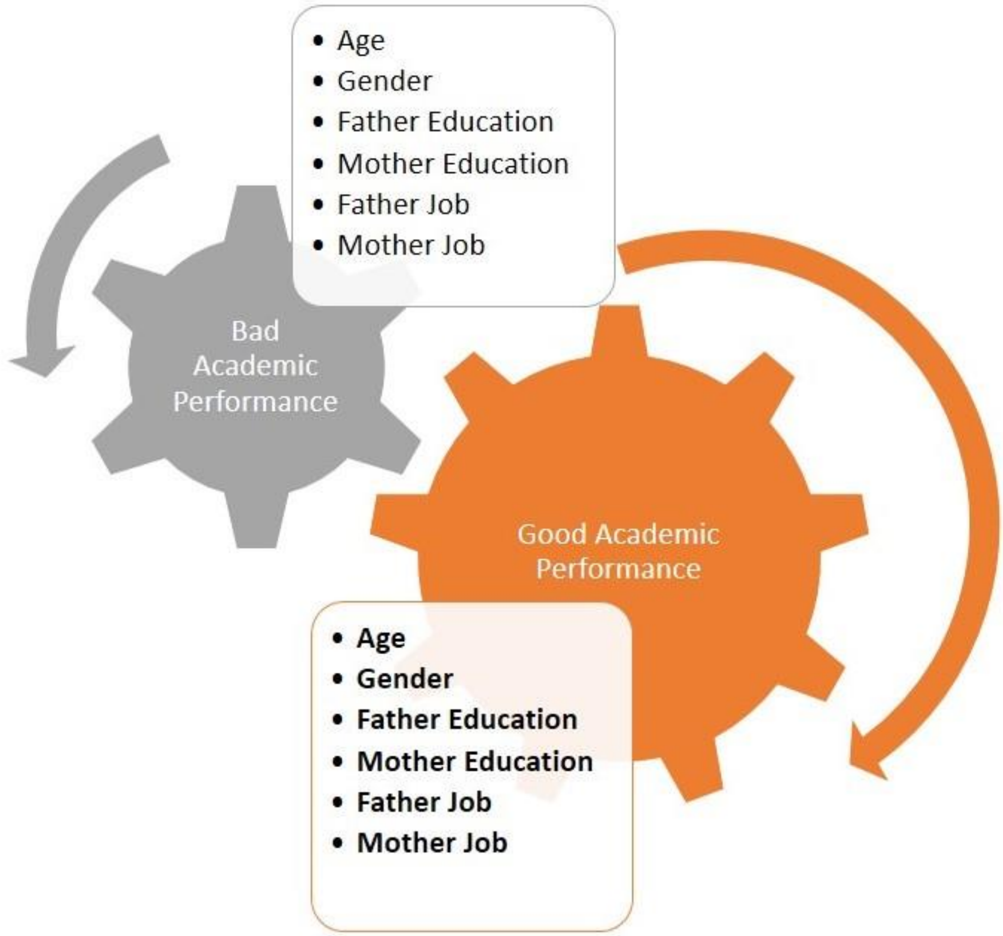


Figure 1: Conceptual Model Framework Diagram

CHAPTER THREE

RESEARCH METHODOLOGY

3.1. Research Design and Methods

Generally the research follows a quantitative experimental research approach. This chapter deals with the study design, methods, subjects/participants, data collection instruments, validity and reliabilities of the questionnaires, procedures of data collection and procedures of data analysis. The researcher will be examining three departments to conduct research on the problems facing during training trainees in practical skill acquisition in Polytechnic Colleges.

3.2. Architecture of Proposed Model

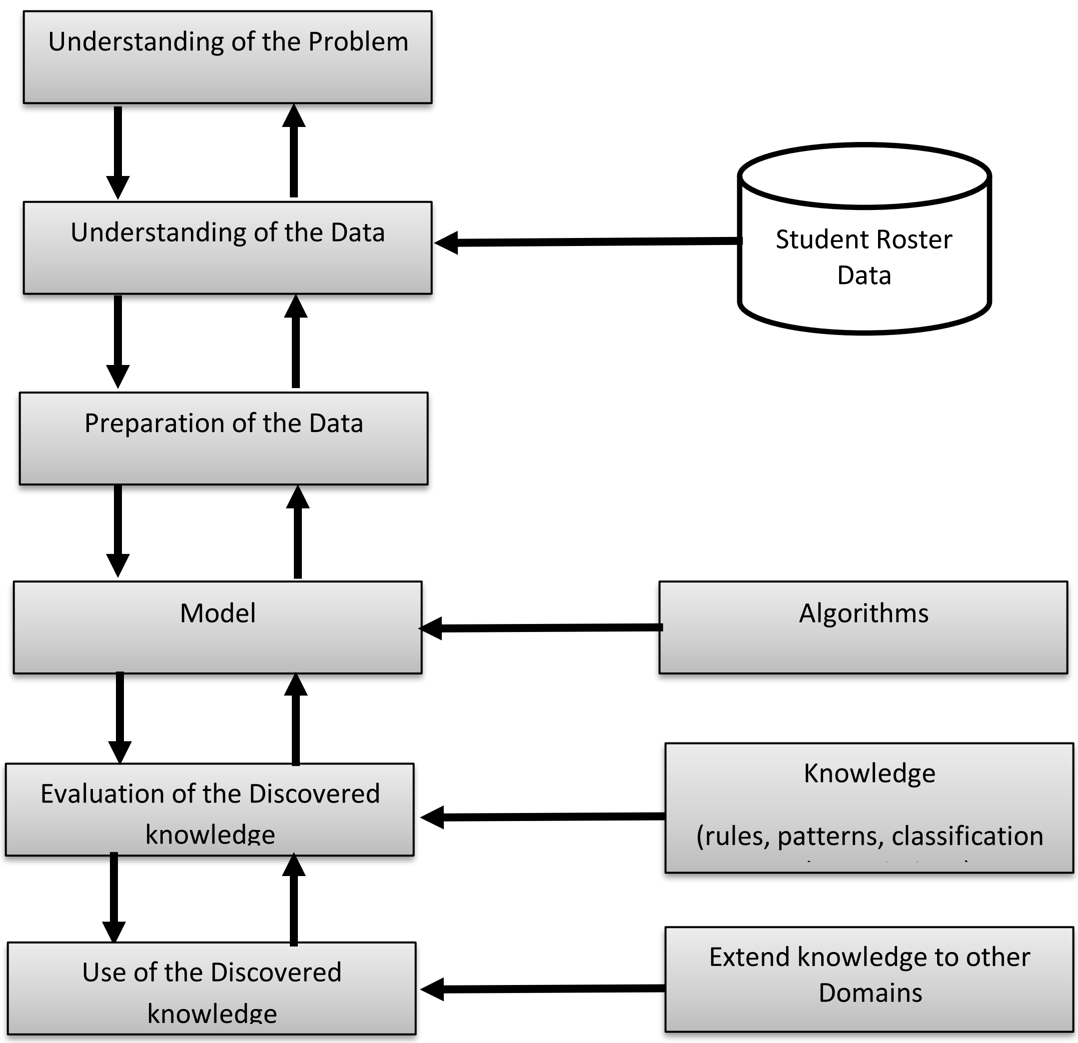


Figure 2: Architecture of proposed model

3.2.1. Understanding of the Problem

This process aids in choosing the suitable algorithm and approach for the investigation. The choice of the tool to be utilized is another crucial action to take in this situation. Due to its platform independence and user-friendly graphical user interface, the SVM tool is utilized in this study. Additionally, it contains several model algorithms, and we are also comfortable using this tool.

3.2.2. Understanding of the Data

Once the student data has been gathered from the current manual as well as electronic files, the qualities found in the data should be reliable for the experts to understand the purpose of each parameter and whether or not they are significant for the specific problem area. Then, characteristics were specified using various data points. The student result data has been organized in excel format collected in Addis Ababa Polytechnic Colleges have been selected using purposeful sampling method.

3.2.3. Preparation of the Data

A smaller dataset is produced during data preparation, which greatly boosts the model's effectiveness. In this step, the appropriate datasets perfect for the research were selected. These datasets were cleaned, integrated, and formatted to be adequate for use. Data cleaning tasks have been completed, including the removal of duplicate records using Excel, the correction of noisy data, the use of estimates to fill in missing values, and the removal of unnecessary attributes and records, or attributes and records that are unrelated to the problems.

The cleansed data would then undergo additional processing for dimensionality reduction. The decision tree attribute selection is used for dimensionality purposes. The input data collection used to construct tests and training data for classification algorithms is the last step.

We use the model technique for the study that has been done. To accomplish the studies stated goal, we chose the best model task and algorithm, and ultimately, we used the model algorithms. In this instance, we build the model using the classification technique or task. Although there are many algorithms available today for models, the right algorithms have been chosen based on the type of data we have, which is nominal, categorical, and numerical. Using the selected model tools, the training and testing methods are created, and the data model is built. Decision tree (J48) classification techniques were used. Algorithm for the Naive Bayes Classifier The accuracy of such algorithms is then evaluated.

3.2.4. Model

A model represents what was learned by a machine learning algorithm. The model is the "thing" that is saved after running a machine learning algorithm on training data and represents the rules, numbers, and any other algorithm-specific data structures required to make predictions.

A machine learning model is produced by the machine learning algorithm after it has learned from data using one of the aforementioned methods. The model was produced by applying an algorithm to the data.

Once you have the model, you may apply it to the data or other sets of data that are comparable to it to create new predictions. The accuracy and confidence with which the model makes predictions will depend on how well the algorithm was trained.

So, in the context of data science, what do algorithms and models mean? Making predictions that you can use to make data-driven business decisions is the aim of machine learning.

3.2.5. Evaluation of the Discovered Knowledge

The phase's models that are produced are evaluated. A test set of data is used to put the algorithm to the test, and recall and accuracy are determined after learning how many of the test sets are evaluated as having great performance. Other variables used to evaluate algorithm performance include classification%, accuracy, precision, recall, and time parameter (the amount of time it took to create the model). Finally, domain experts will be used to evaluate the model that has been constructed.

3.2.6. Use of the Discovered Knowledge

The model process, which describes how to apply the knowledge that has been obtained to diverse situations, comes to a close with this stage. Interested parties can use the new information or incorporate it into another system to enable additional action.

Finally, the knowledge acquired can be put to use by developing a prototype and testing each incoming claim before processing.

3.3. Population and Sampling

In order to achieve the objective of the research, the following methods will be employed. Generally the methodology used is quantitative experimental using machine learning approach. Thus various experiments will be conducted using algorithms selected and comparison will be made accordingly.

3.3.1. Sampling Techniques

In this study, domain experts and schools were chosen using the purposive sample method in order to collect knowledge from the polytechnic colleges. The selection criteria of domain experts for the study are based on their academic year. Therefore, 1500 students in the Polytechnic Colleges were selected for fill the questioners from those colleges for identifying the problem. The selected colleges were the most known and experienced TVETs in huge communities in Addis Ababa.

The research is conducted using primary data that has been taken from a Polytechnic Colleges database that is accessible to academics. Secondary information from student interviews is recorded using a Google form, along with student performance. As a result, interviews will be used in this study's quantitative and qualitative research designs to grasp the domain knowledge and interpret the findings.

3.4. Source of Data

There are two methods of data collection which was used in this study for preparing the prediction model. Those are primary and secondary. The primary method includes the data gathered through roster primarily. The secondary data was gathered from questionnaires from those polytechnic colleges student fill using Google form applications.

3.5. Data collection

- Primary Data Source

For the purpose of experimentation, data is gathered from the registrar student record and survey student information by using Google form applications. I used the scientific method is applied in the design of experimental quantitative research. It sets protocols that enable the researcher to put a hypothesis to the test and to methodically and objectively investigate causal links between variables.

Observable traits of an experimental quantitative study include:

- The variables' characteristics and relationships
- A precise question that can be answered; subjects divided into groups according to pre-established standards
- Experiments that alter the dependent variable
- Measuring the dependent variable both before and after changes to the independent variables

- Secondary Data Source

Questionnaire

Questionnaire enables to secure factual information about opinions and views and also appropriate instrument to obtain a variety of opinions with a relatively short period of time. Both closed-ended and open-ended questionnaires will be employing in this study. Most of the questionnaire will be closed ended, while some questionnaire items open-ended in order to give a chance for respondents to express their views, ideas and opinions using their own words.

3.6. Experimentation Tools

Different tools will be use during the analysis, among these tools: From the development tools we will be use Python. Python is a programming language that lets you work quickly and integrate systems more effectively.

3.6.1. Libraries and software tools

An analysis of available software tools and their libraries is performed to select the best tool for implementing the SVM algorithm for developing a model. Before selecting the tools, we considered some criteria that will assist us in selecting the appropriate software tools and libraries. The primary criterion is the programming language (python) that will be used to put the automated system into action. Additional considerations also use equipment that has an adequate level of knowledge material, like free YouTube videos, our knowledge and understanding of the language, freely available software, and tools' ability to be used in machines with limited resources (like RAM, Hard disk space of the pc and CPU). To incorporate a Decision Tree algorithm, we were using Python as programming language Tensor Flow as libraries in an Anaconda atmosphere. These tools meet all of the criteria for consideration and are written in

Python, which we are familiar with. Programming language Python 3.8.1 has been used, and it's unrestricted & has a range of unrestricted library, extensive documentation, & contributor. Following that, you will find a selection of helpful libraries & software tools.

- ✓ Numpy
- ✓ Matplotlib
- ✓ OpenCV

Anaconda

Anaconda the model is built with Anaconda, an open and free allocation of the Python programming languages for data science implementations that intends to streamline package management and implementation.is used to create the model.it includes IDEs like Spyder and Jupyter Notebook that should be used to create the coding. Jupyter Notebook has been used to integrate the code portion. As part of the free Python distribution Anaconda® created by Continuum Analytics specifically for scientific computing, Anaconda Navigator is a desktop GUI. Even for business usage, it's free to download then use. As a supplement, the software offers a number of programs those are essential for data science, mathematics, and design.

TensorFlow is a Google-developed, free, and open-source deep learning framework that is presently the most well-known and quickest [38]. There are two CPU and GPU distributions.

Keras is a greater neural network written in Python that displays the results of Tensor Flow.

MS Office Visio 2007 the system architecture is designed in Microsoft office Visio 2007. With perfect templates, this application was used to collaborate, and share data-linked diagrams, assisting in the simplification of complicated information.

Decision tree analysis involves visually outlining the potential outcomes, costs, and consequences of a complex decision. These trees are particularly helpful for analyzing quantitative data and making a decision based on numbers.

A decision tree is a powerful and popular tool for classification and prediction. A decision tree consists of nodes that form a rooted tree, which means it is a directed tree that has no incoming edges.

Why decision trees are popular?

- ✓ Easy to interpret and present
- ✓ Well defined Logic, mimic human level thought
- ✓ Random Forests, ensembles of decision trees are more powerful classifiers
- ✓ Feature values are preferred to be categorical. If the values are continuous then they are discretized before building the model

A- Decision Tree

A decision tree is a flow-chart-like tree structure, where each internal node is denoted by rectangles, and leaf nodes are denoted by ovals. All internal nodes have two or more child nodes. All internal nodes contain splits, which test the value of an expression of the attributes.

(Yadav, S. K., & Pal, S. (2012). The advantages of decision trees are that they represent rules which could easily be understood and interpreted by users, (Kabakchieva, 2013) the tow decision tree algorisms filters applied on the dataset are the j48 and the random forest.

SVM

Support Vector Machine: This is the most well-known abbreviation for SVM. A Support Vector Machine is a type of machine learning algorithm used for classification and regression tasks. It works by finding a hyperplane that best separates different classes in a high-dimensional feature space.

J48

The J48 algorithm is WEKA‟s implementation of the C4.5 decision tree learner. The algorithm uses a greedy technique to induce decision trees for classification and uses reduced-error pruning. The C4.5 algorithm was proposed in 1992, by Ross Quinlan, to overcome the limitation of the ID3 algorithm (unavailable values, continuous attribute value ranges, pruning of decision trees, etc.). C4.5 uses a divide-and-conquer approach to growing decision trees. The default splitting criterion used by C4.5 is gain ratio, an information-based measure that takes into account different number of test outcomes (Quinlan, R. J. (1996).

Random Forest

The random forest classifier A tree-based classifier consists of a combination of tree classifiers where each classifier is generated using a random vector sampled independently from the input vector, and each tree casts a unit vote for the most popular class to classify an input vector (Pal, 2005). The random forest classifier used for this study consists of using a random combination of features at each node to grow a tree. Bagging. In a random forest, each node is split using the best split among all variables. In a random forest, each node is split using the best among a subset of predictors randomly chosen at that node (Wiener, 2003).

B- Bayesian

Bayesian inference is one of the principal statistical techniques used in data mining.

Naïve Bayes

The naïve Bayesian classifier provides a simple and effective approach to classifier learning (Minaei-Bidgoli, 2004). Naive Bayes is a famous classifier which consists on conditional probabilities. Indeed, it employs Bayes' theorem and it supposes that features have a solid independence between each other. However, it has many advantages such as simplicity of use, quick convergence, and high scalability. Finally, Naïve Bayes needs less training data for building a model (Alaoui, Farhaoui and Aksasse, 2018)**.**

3.7. Validity and Reliability of the research

Two of the factors that affect the questionnaire's quality are validity and reliability. The degree to which the measure is truly measuring what was supposed to be measured is referred to as validity. A questionnaire's face validity, content validity, construct validity, statistical validity, ecological validity, and internal and external validity are just a few of the validity types that can be considered while evaluating it (Leavy, 2017). However, according to Creswell & Cresswell (2018), among these forms of validity, the three that need to be looked for are construct validity, predictive or concurrent validity, which looks at the scores predicting a criterion measured and does results correlate with other results, and content validity, which looks at the item measuring the content they are intended to measure. Which measures hypothetical constructions or notions by looking at the item. In general, validity aids the researcher in deciding whether or not the questionnaire is appropriate for survey research. On the other side, the consistency of the outcome is referred to as reliability. The two reliability tests that are most frequently used to examine the internal consistency of scales are factor analysis and Cronbach's alpha (Leavy, 2017). These tests assess how consistently sets of items behave (Creswell & Cresswell, 2018). The quantity of items and their average intercorrelation affect Cronbach's alpha. Consequently, a high Cronbach's alpha score might suggest a high level of reliability. As the respondent might recall their previous response and make an effort to be consistent, it may also suggest that the replies may influence one another. A low value, on the other hand, can be an indication of low dependability or even the measurement of different constructions (Wiley, 2020).

3.8. Evaluation and Deployment

At this step, the obtained model is more carefully evaluated, and the methods used to create the model are looked at to make sure they correctly meet the business objectives. The criteria for comparing algorithms can be determined by evaluating accuracy, speed, scalability, interpretability, and robustness. A confusion matrix is an approach that presents the predicted and actual classification based on multiple standers, such as the true positive rate (TPR), which "are examples correctly labeled as positives (Staeheli and Mitchell, 2010). In this study, the model was evaluated by measuring the accuracy of algorithms, which can be defined as the capacity of a classifier to predict the class label correctly. The false positive rate (FPR), according to Staeheli and Mitchell (2010), "refers to negative examples that are mistakenly classified as positive." A high FP rate signifies that the majority of the results returned by the algorithms are relevant, and a high TP rate means that the algorithm provides more relevant results than irrelevant ones.

The knowledge gained will need to be organized and presented in a way that the customer can use it. The final models from the previous phase are then applied on a testing assessing dataset predictive accuracy and consistency.

CHAPTER FOUR

EXPERIMENTATION AND RESULTS

4.1. Overview

The goal of this chapter is to propose solutions based on the findings of the study on model development based on demographics and psychographic factors to improve performance in Technical and Vocational Education and Training (TVET) institutions. This chapter will focus on practical recommendations for TVET institutions based on the results of our research.

The data are well understood, thoroughly investigated, carefully chosen, and sufficiently clean for model creation in Chapter 3. This chapter explains the thorough data analysis used to choose a modeling approach, how that approach was put into practice using the proper algorithms, and how the models were compared to determine which was the best at identifying the components. Data analysis is the process of evaluating data using analytical and logical reasoning to examine each component of the data provided (Ahmad, 2015). Data from various sources is gathered, reviewed, and then analyzed to form some sort of finding or conclusion.

4.2. Data Description

Model development based on Demographic and Psychographic factor to improve performance in case of TVET for this research we use the dataset is 5841 and there is 11 attributes also the data format in in csv file format.

D2 | Construction

	A	B	C	D	E	F	G	H	I	J	K
1	ID Numbe	Gender	Age	Department	Level	Father Edu	Mother Edu	Father Job	Mother Job	Course Title	Grade
2	BCM501/:	M	21	Construction	1	Other	Other	Government	Other	Demonstrate Work	A
3	BCM504/:	M	20	Construction	1	Bachelors	Other	Private	Other	Demonstrate Work	B
4	BCM505/:	M	23	Construction	1	Diploma	Other	Government	Other	Demonstrate Work	B
5	BCM507/:	M	22	Construction	1	Certificate	Other	Government	Own Business	Demonstrate Work	B
6	BCM508/:	M	19	Construction	1	Other	Other	Government	Other	Demonstrate Work	B
7	CM502/1:	M	18	Construction	1	Other	Certifcate	Other	Other	Demonstrate Work	B
8	CM503/1:	M	23	Construction	1	Other	Certifcate	Private	Other	Demonstrate Work	C
9	CM504/1:	F	20	Construction	1	Other	Certifcate	Own Business	Other	Demonstrate Work	B
10	CM505/1:	M	24	Construction	1	Diploma	Other	Government	Private	Demonstrate Work	C
11	CM518/1:	M	22	Construction	1	Diploma	Other	Private	Other	Demonstrate Work	C
12	CM506/1:	M	23	Construction	1	Certificate	Other	Own Business	Other	Demonstrate Work	C
13	CM507/1:	M	21	Construction	1	Certificate	Diploma	Other	Own Business	Demonstrate Work	B
14	CM509/1:	M	19	Construction	1	Other	Diploma	Other	Private	Demonstrate Work	C
15	BCM509/:	M	18	Construction	1	Other	Diploma	Government	Government	Demonstrate Work	B
16	CM510/1:	M	21	Construction	1	Other	Other	Government	Government	Demonstrate Work	B
17	CM513/1:	M	23	Construction	1	Bachelors	Other	Government	Government	Demonstrate Work	A
18	CM514/1:	F	24	Construction	1	Bachelors	Other	Private	Other	Demonstrate Work	B
19	CM515/1:	M	19	Construction	1	Other	Other	Own Business	Other	Demonstrate Work	B
20	CM516/1:	M	20	Construction	1	Diploma	Other	Government	Other	Demonstrate Work	C

Figure 3: Sample dataset

4.3. Formatting the Data

After the data are collected and organized from the selected polytechnic colleges it should be encoded from hard copy to soft copy in excel format. And then we have the data suitable to the intended data mining tool. In the conducted research we use python jupyter Notebook software.

However, the raw data set to a suitable format that has to be compatible with the python jupyter Notebook tool which is a comma-delimited format or comma-separated value (CSV). In this case, the data extension file .xlsx must be changed to .csv after the excel file changed to a comma-separated value it is simple to import the data set to the python jupyter Notebook software on the python jupyter Notebook explorer. Python jupyter Notebook is free to download and is a set of integrated tools algorithms to help you more productive for different data mining tasks such as classification. It includes classification, clustering, an association that supports direct data execution, and a variety of robust tools for plotting, viewing history, analyzing, and managing our workspace (Dr.Kishori, 2014).

4.4. Experimentation Work

At this stage the thesis work focus on what kind of algorithm can we use for the implementation and we have 11 attributes.

Attribute of the data set with their descriptions	
Attributes	**Descriptions**
Trainees ID.No.	Uniquely identify the trainees
Gender	The characteristics of women, men
Age	A period of human life, measured by years
Department	A distinct part of anything arranged in divisions
Level	Qualification framework in terms of progressive stages of achievement and complexity
Father Education	Academic achievement
Mother Education	Academic achievement
Father Job	Working environment
Mother Job	Working environment
Course Title	Trainees take a skill or knowledge
Grade	Result of the trainees

Table 2: Attribute of the data set with their descriptions

4.5. Preprocess Data

Row data may have incomplete records, noise value, outliers, and inconsistent data this leads to low-quality mining results. In this experiment, the data set must be prepared in the appropriate format.

4.5.1. Data Cleaning

In the conducted experiment, we uses the following sub techniques to handle data cleaning problems.

4.5.2. Missing Value Handling

Handling Missing Values python jupyter notebook tool uses NA for missing values. All mathematical operations with NA values return NA. Sometimes missing data may be recorded because, the data may not be entered due to misunderstanding, in this experiment the researcher uses the Python Jupyter Notebook tool to identify and select an attribute that contains an incomplete value and the number of missing values for each attribute. When such happens, the researchers tried to fill the missing value manually (Etsehiwot, 2020).

Remove Noisy Values In our dataset there were some noisy attribute values as well as attribute itself due to data entry problems, duplicate records, incomplete and inconsistent data problems.

In real-world datasets contain missing values may occur for several reasons such as human error, computer error, technology limitation, hardware malfunctioning, and many other factors. In this study some of the missing values were present, for instance, the values missed in the attribute age occur because the student registered by another person.

When applied the quality of knowledge extracted, learning, and decision problems depend directly upon the quality of training data. Due to the above reason missing values were handled properly. To handle missing value there are many methods like, ignore the tuple, Fill in the missing value manually, Use a global constant to fill in the missing value.

4.5.3. Data Transformation

The data is altered in this experiment to make it suitable for mining. We used attribute restriction or feature selection to choose or order the most crucial attributes throughout the class and normalization to scale the data for specific attributes, allowing the data to fall within a specified range.

4.6. Model Development

The main steps in making a model are choosing a modeling technique, creating a test design, building the model, and evaluating the model.

4.6.1. Selection of Modeling Technique

The choice of a specific modeling technique is the first stage in the model-building process. The purpose set out informs the choice of modeling technique. Classification techniques were employed to generate the model because the goal of the research is to create a predictive model for student performance. Utilizing the python jupyter notebook environment, the analyses were carried out.

The class outputs are the leaf nodes (Witten and Frank, 2005). One of the most popular and useful types of machine learning and data mining is SVMs. Recursive partitioning is a method for creating SVM models. J48 classification methods were employed in this investigation.

This chapter discusses all the findings and results. Firstly, results of different classification in experiment 1. Finally the results showed which factor is most affected student performance.

Choosing the most suitable algorithm for a new dataset is important task since there is no single classifier that yields the best results on all datasets (Osmanbegović & Suljić, 2012; Cristobal Romero, Espejo, Zafra, Romero, & Ventura, 2010). Therefore, as a first step, the predictive performance and intelligibility of the models generated.

4.7. Data Presentation

4.7.1. Categorical Variable distributions

All attributes show that how the categorical variable distributions is and each attributes describe their categorical variables.

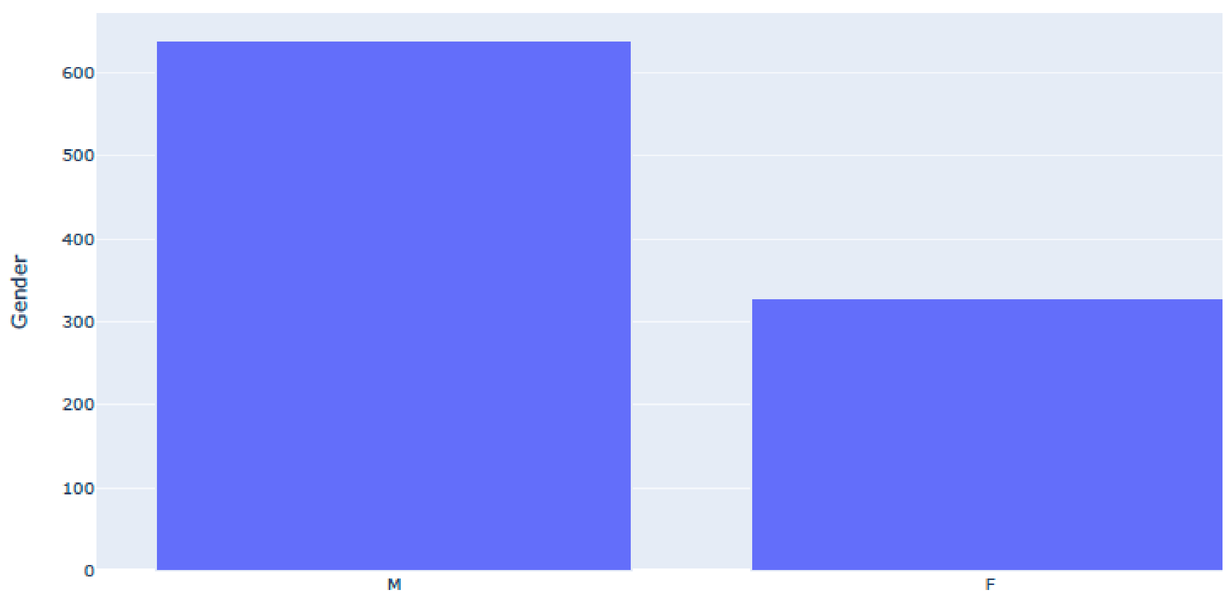


Figure 4: Categorical variables in Male and Female Distributions

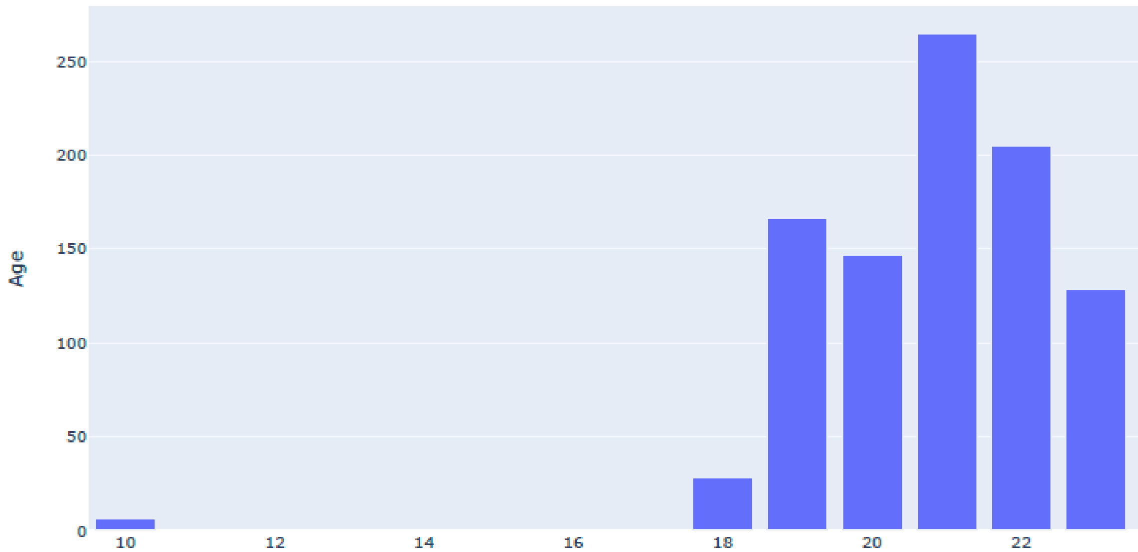


Figure 5: Categorical variables in Age Distributions

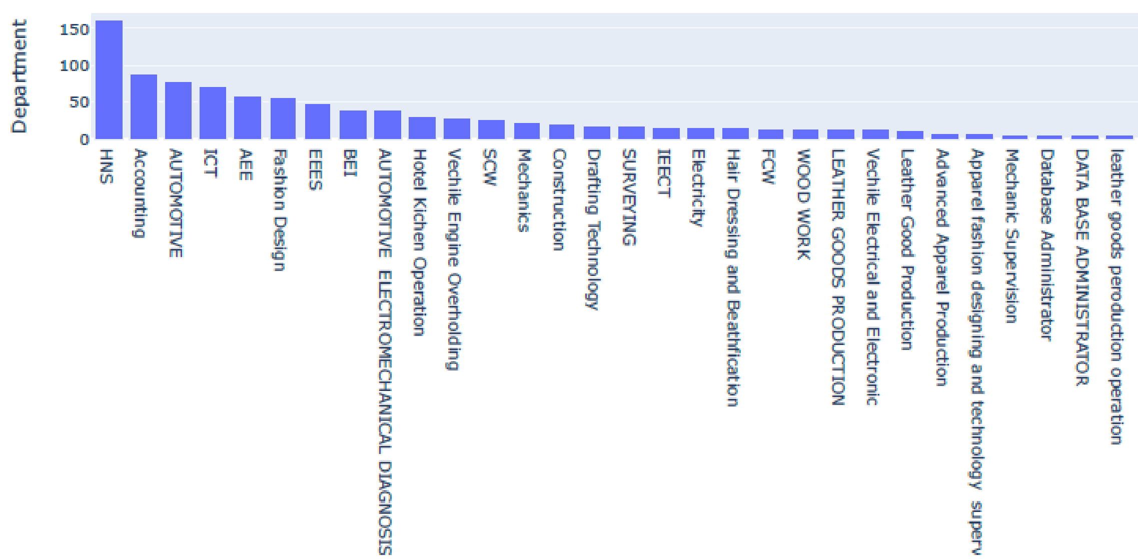


Figure 6: Categorical variables in Department Distributions

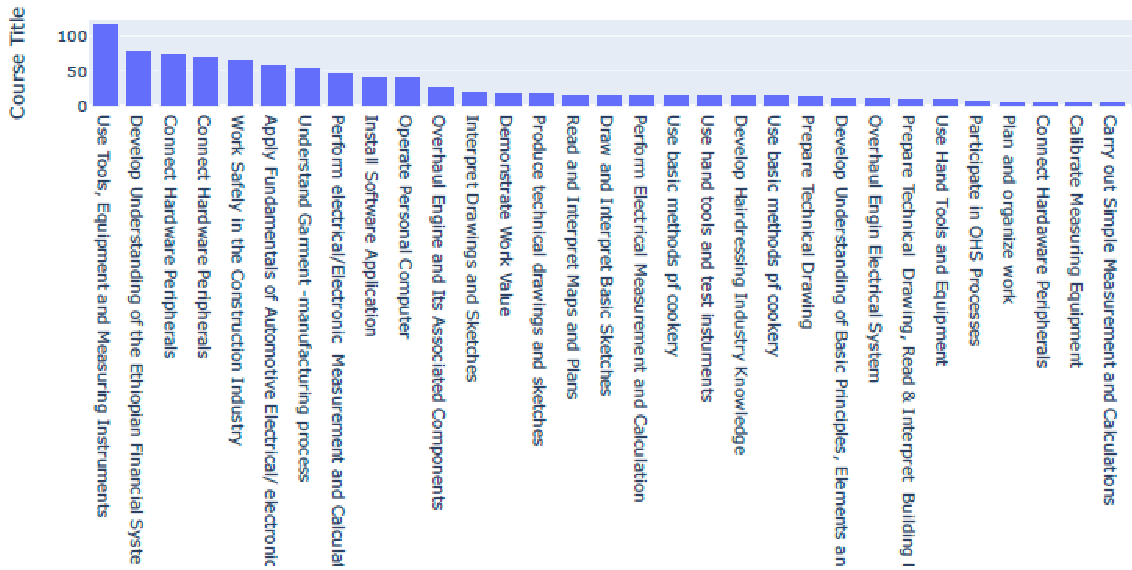


Figure 7: Categorical variables in Course Title Distributions

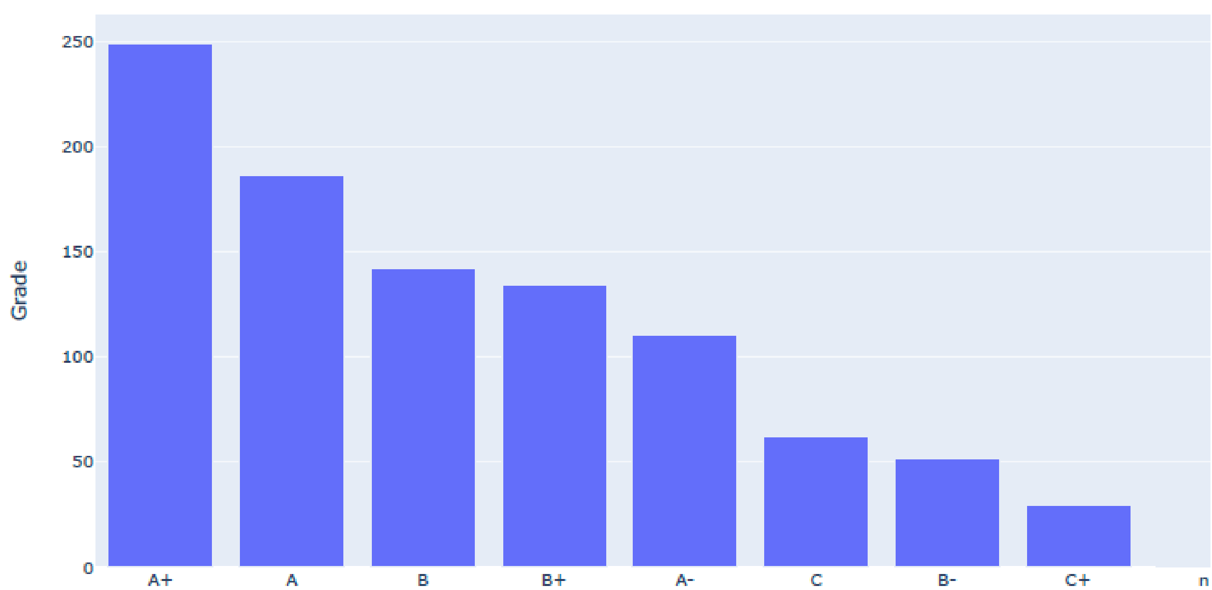


Figure 8: Categorical variables in Grade Distributions

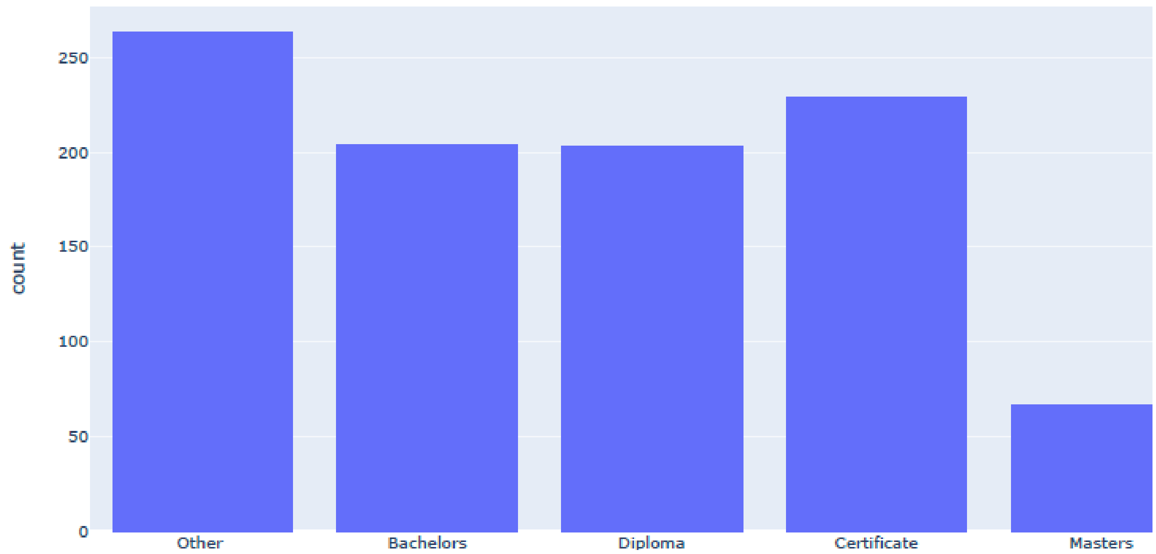


Figure 9: Categorical variables in Father Education Distributions

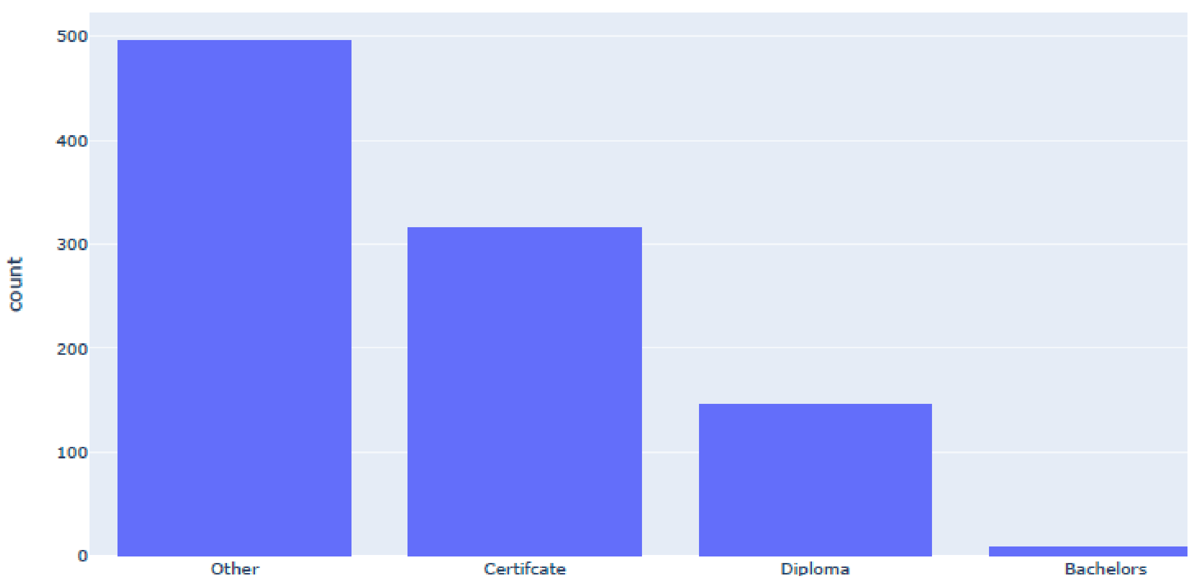


Figure 10: Categorical variables in Mother Education Distributions

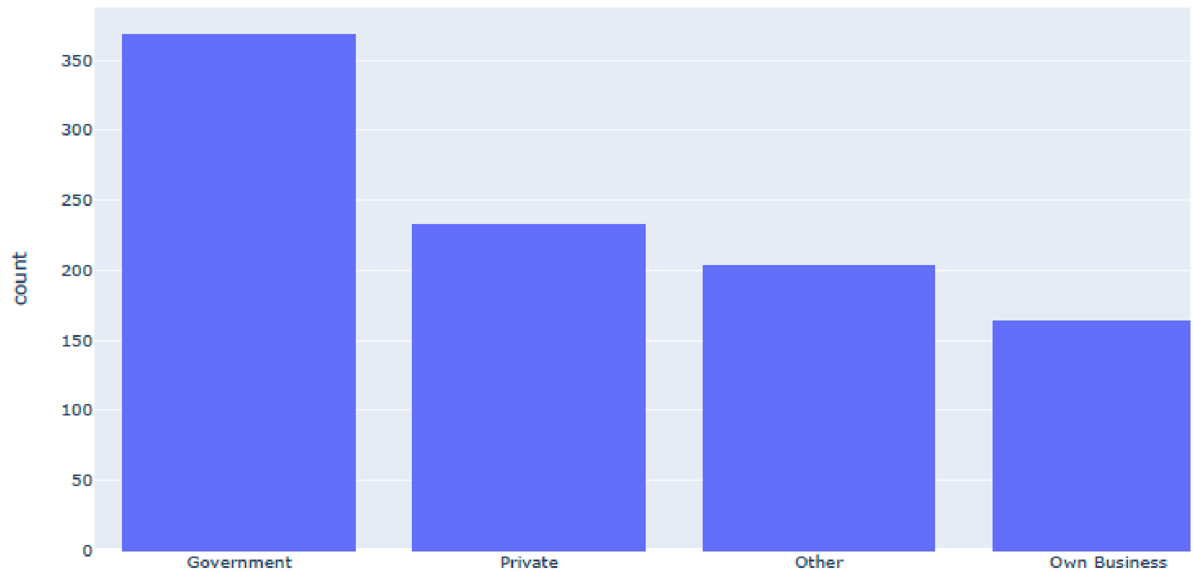


Figure 11: Categorical variables in Father Job Distributions

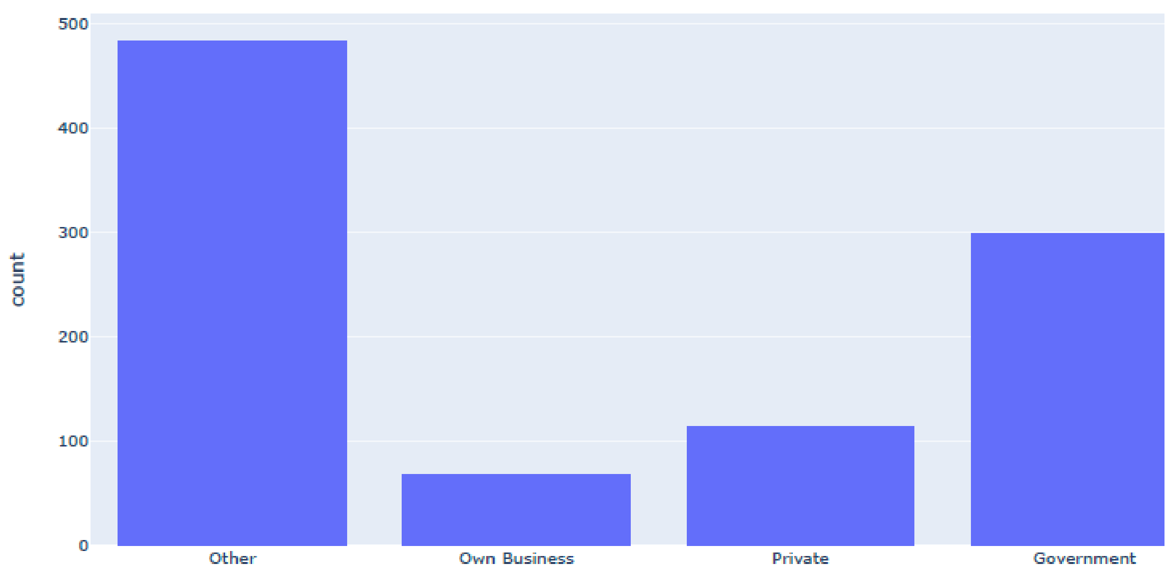


Figure 12: Categorical variables in Mother Job Distributions

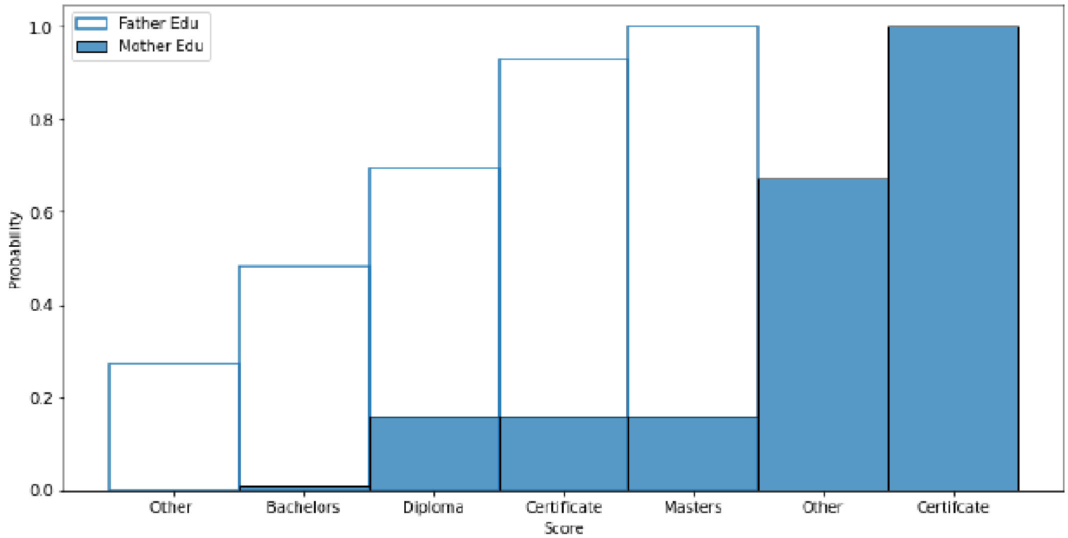


Figure 13: Comparison between Father Education and Mother Education

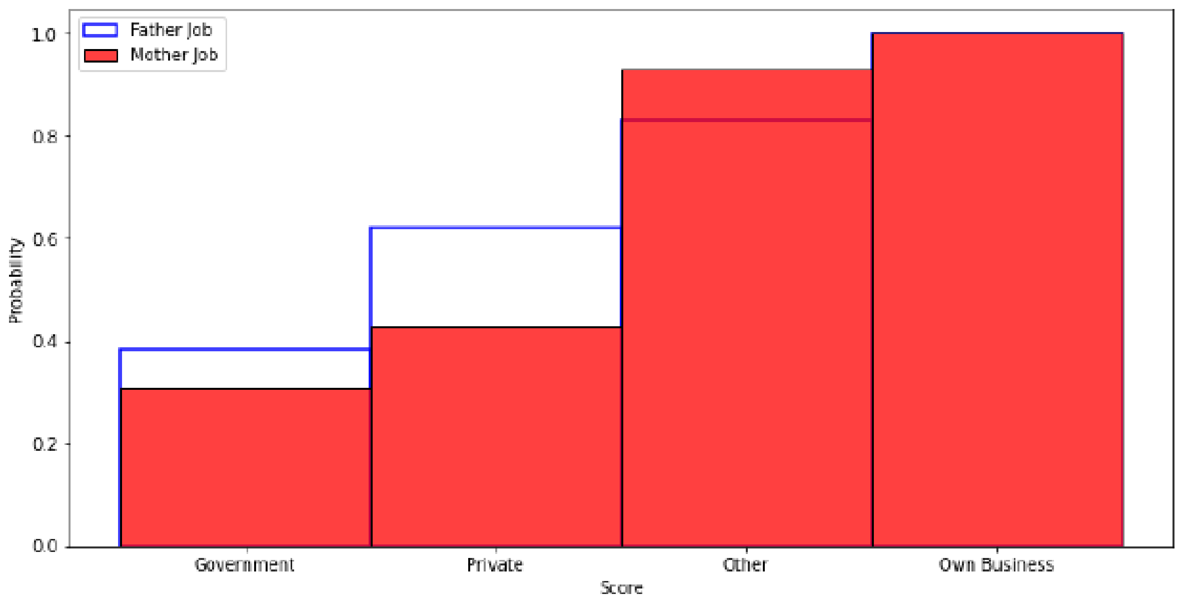


Figure 14: Comparison between Father Job and Mother Job

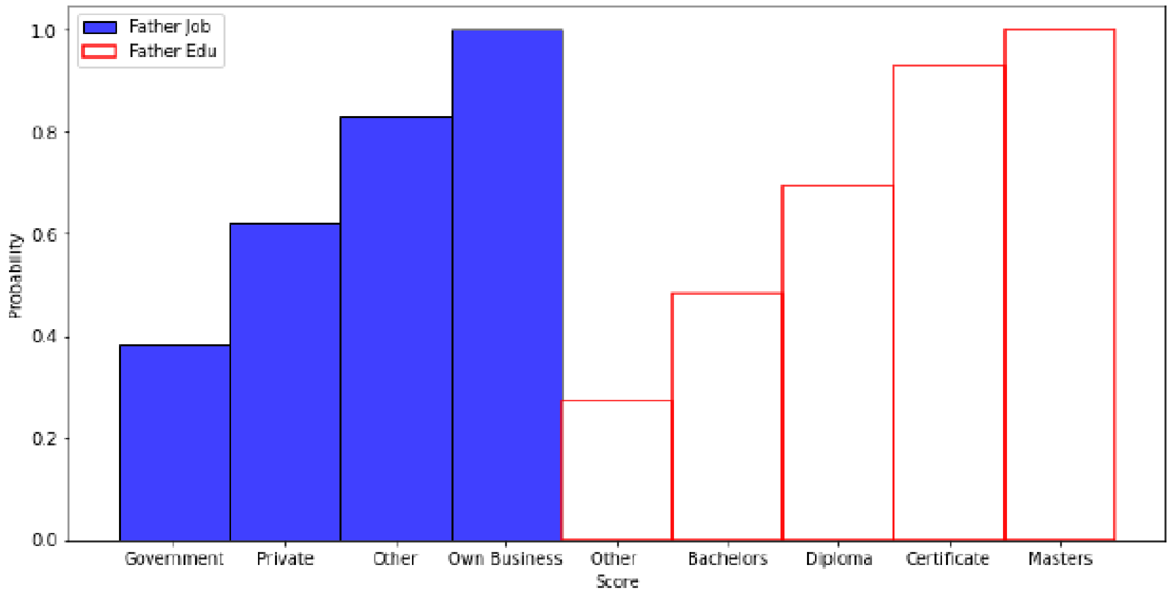


Figure 15: Comparison between Father Job and Father Education

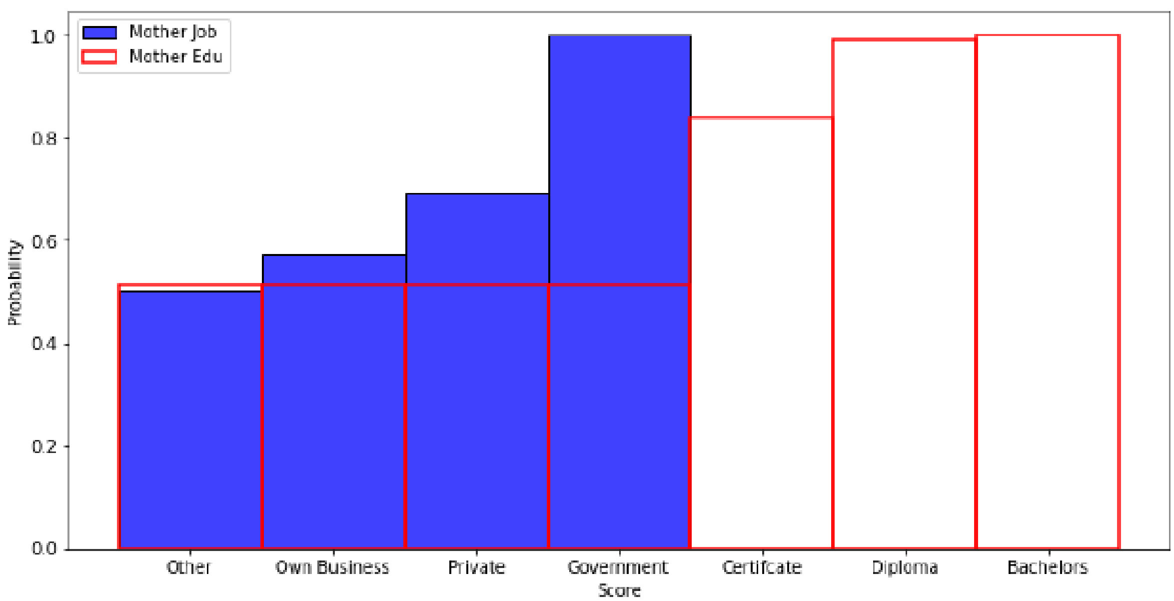


Figure 16: Comparison between Mother Job and Mother Education

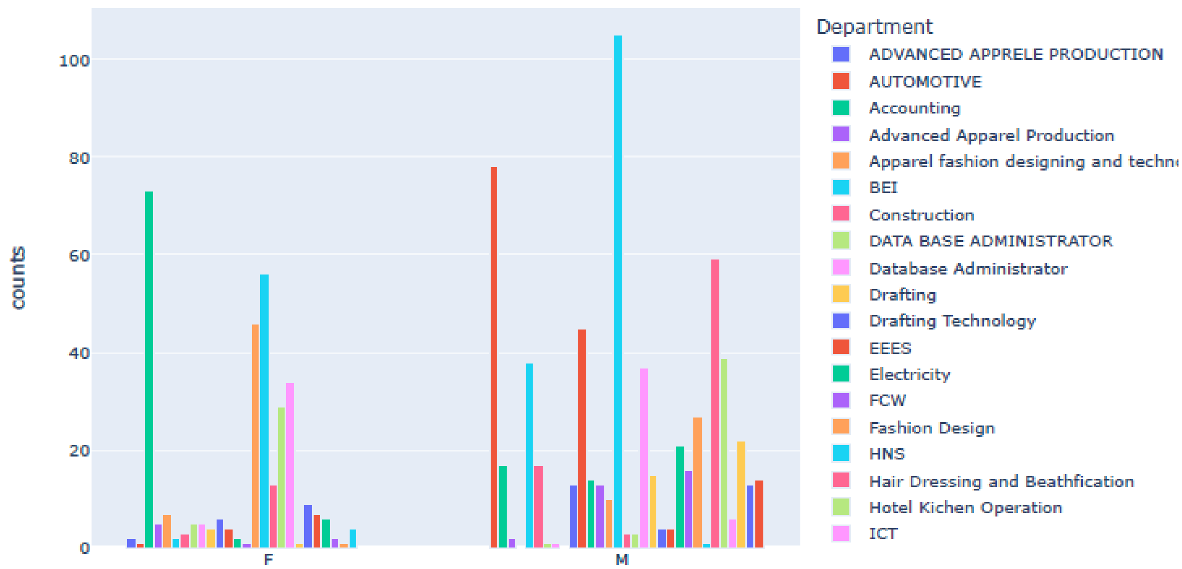


Figure 17: Relationship between categorical variables in Gender and Departments

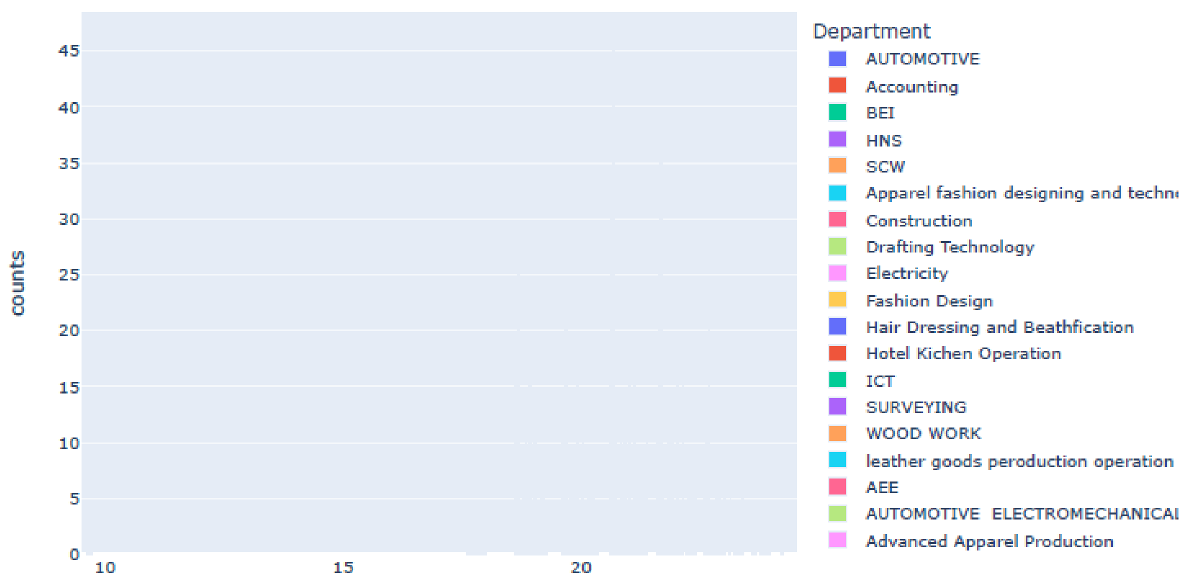


Figure 18: Relationship between categorical variables in Age and Departments

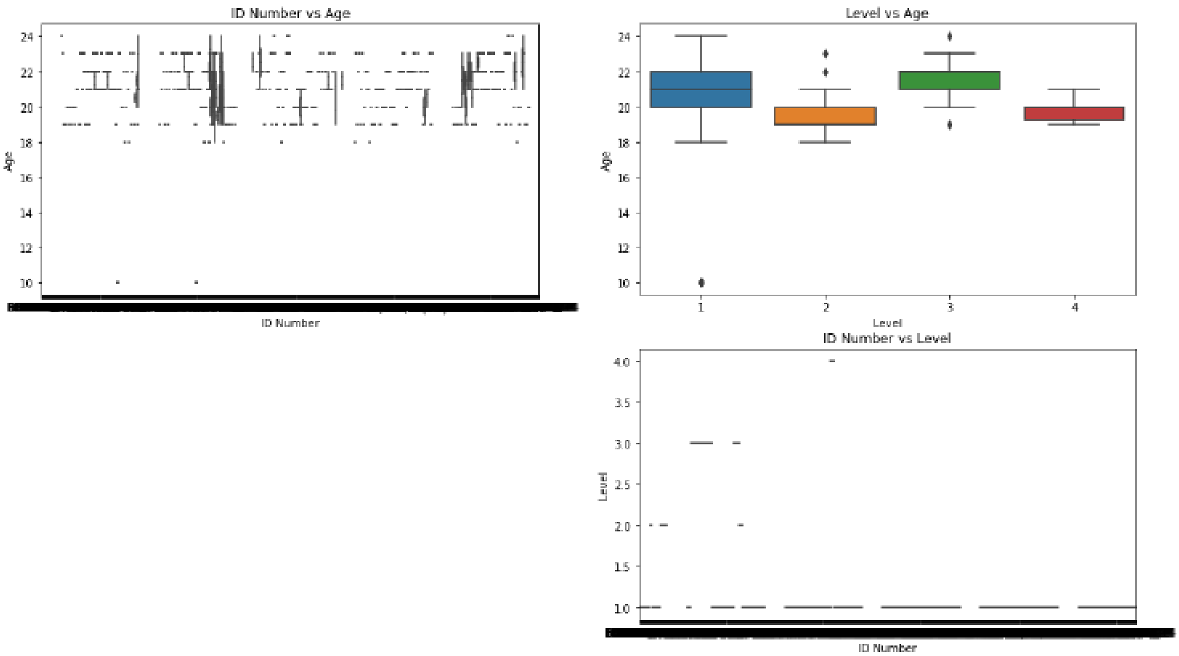


Figure 19: Relationship between categorical and continuous variables

4.8. Experimentation Setup

To implement Model Development based on Demographic and Psychographic factor to improve performance in case of TVET using SVM algorithm the dataset was divided in to train and test set with 80/20 split ratio.

The following steps were followed for build the chosen algorithm

- Import and install the necessary basic libraries and packages
- Adding or importing preprocess dataset
- Extract variable from the dataset assigning them to x and y
- Splitting the dataset in to training and test sets
- Initiates the selected classifier on the training dataset
- Using the trained classifier to make prediction on test dataset
- Evaluating the performance of the classifier

4.8.1. Training of the Model

Once the dataset is divided into training and test sets, a split ratio of 80:20 is applied. This means that out of the 5,841 datasets, 4,672 allocated for training purposes, while the remaining 1,169 are reserved for testing.

In [24]:
```
# Display the number of rows and columns in the DataFrame
num_rows, num_columns = data.shape
print("Number of rows:", num_rows)
print("Number of columns:", num_columns)
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# Print the number of samples in the training and testing datasets
print("Number of samples in training dataset:", X_train.shape[0])
print("Number of samples in testing dataset:", X_test.shape[0])
```

```
Number of rows: 5841
Number of columns: 11
Number of samples in training dataset: 4672
Number of samples in testing dataset: 1169
```

Figure 20: Training of the selected SVM model

4.8.2. Testing of the Model

After completing the training of SVM models using a carefully selected dataset, the next step involves evaluating their performance on an unseen test dataset. This test dataset constitutes 20% of the entire dataset, comprising a total of 1169 for total of 5,841. By assessing the models on this separate test dataset, we can get their ability to generalize and make accurate predictions on new, previously unseen data. This evaluation enables us to measure the effectiveness of the trained SVM models in real-world scenarios, providing insights.

```
# Calculate test accuracy
test_accuracy = accuracy_score(y_test, y_test_pred)
print("Test Accuracy:", test_accuracy)
```

```
Training Accuracy: 0.9056078767123288
Test Accuracy: 0.844311377245509
```

Figure 21: Testing the trained SVM model

4.9. Result of the SVM

In the pursuit of understanding and forecasting student performance, I embarked on an experimental journey employing the powerful Support Vector Machine (SVM) algorithm. With meticulous preprocessing and a dataset enriched with various attributes, the SVM model showcased its predictive prowess. The culmination of this endeavor bore witness to an impressive achievement of 84% accuracy, illuminating the path toward better grasping the intricate factors that contribute to student outcomes.

The SVM algorithm deftly navigated through the complexities of the dataset, leveraging its decision function strategies and probability estimations. The 84% accuracy achieved encapsulates the collective effort of not only data manipulation but also the algorithm's knack for discerning patterns and nuances within the information. As each attribute was meticulously one-hot encoded, the SVM exhibited its ability to interpret these diverse features and generate meaningful predictions. This significant outcome emphasizes the potential of machine learning, as encapsulated by the SVM, to decode the multifaceted tapestry of student performance determinants.

In conclusion, the SVM experiment has proven to be an enlightening and fruitful endeavor, with its 84% accuracy lending credence to its efficacy in predicting student performance. This journey underscores the importance of leveraging sophisticated algorithms to unravel the complexities of educational data, paving the way for enhanced insights into the factors that shape academic outcomes.

Table 3: Performance Metrics

Algorithms	Accuracy	Precision	Recall	F1-
SVM	0.84	0.86	0.84	0.85

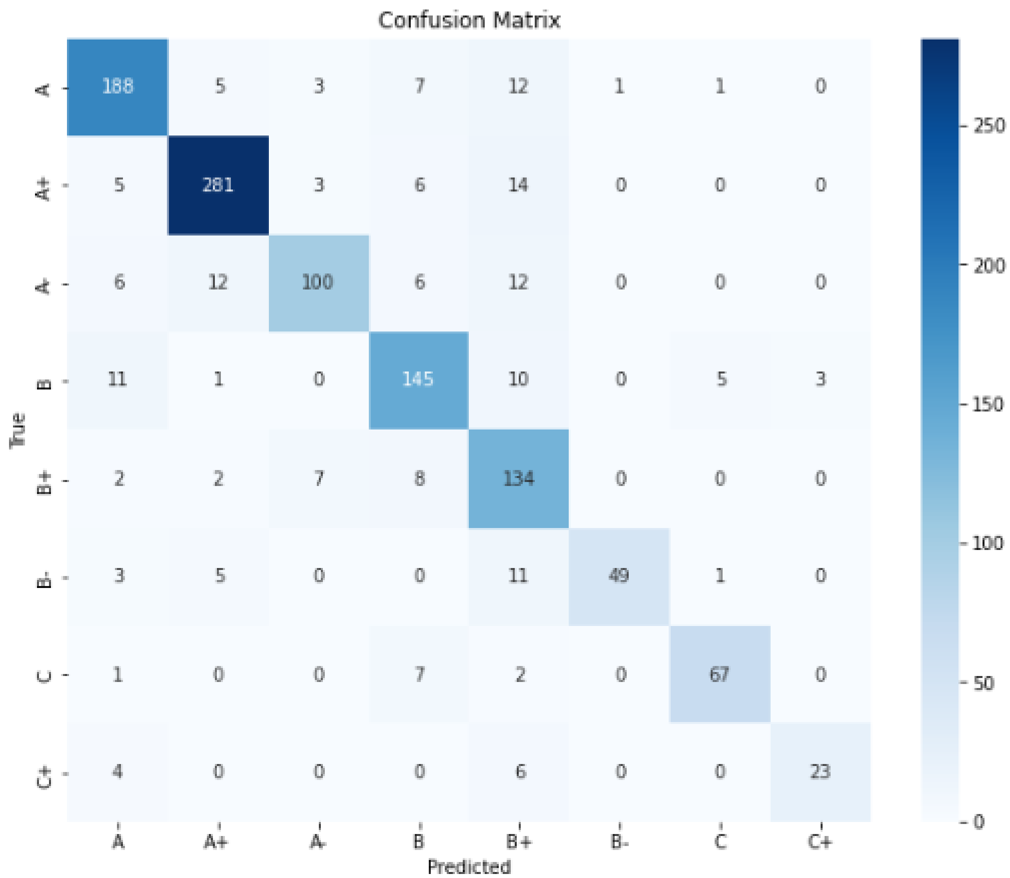


Figure 22: Confusion Matrix

Classification Report:

	precision	recall	f1-score	support
A	0.85	0.87	0.86	217
A+	0.92	0.91	0.91	309
A-	0.88	0.74	0.80	136
B	0.81	0.83	0.82	175
B+	0.67	0.88	0.76	153
B-	0.98	0.71	0.82	69
C	0.91	0.87	0.89	77
C+	0.88	0.70	0.78	33
accuracy			0.84	1169
macro avg	0.86	0.81	0.83	1169
weighted avg	0.86	0.84	0.85	1169

Figure 23: Classification Report

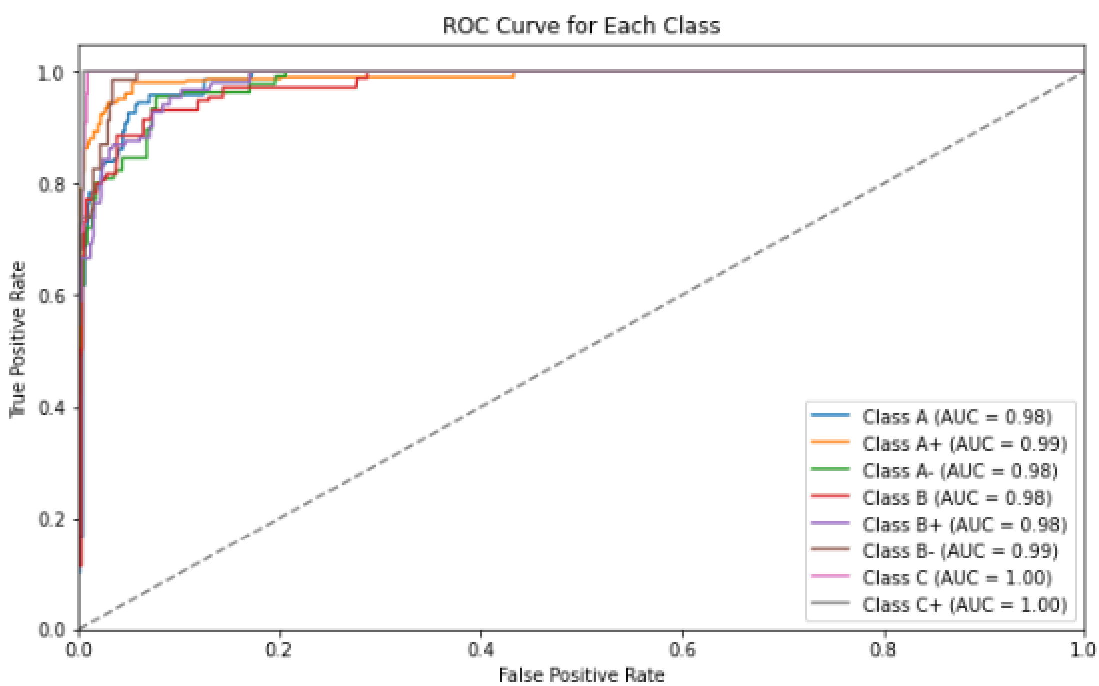


Figure 24: ROC Graph

4.10. DISCUSSION

Discussion of the results in light of the research questions indicated at the beginning of this document is provided here.

Research Question 1: What possible traits of demographic and psychographic behaviors could be used to assess students' performance?

In addressing the first research question, the study focused on identifying the potential demographic and psychographic factors that could influence students' academic performance in the context of Technical and Vocational Education and Training (TVET). By carefully selecting attributes that encompassed both demographic and psychographic information, the study aimed to create a comprehensive dataset that accurately represents students' diverse backgrounds and personal characteristics.

Through meticulous data collection and preprocessing, the study incorporated attributes such as age, gender, educational background, learning preferences, and personal motivations. These attributes were chosen based on their relevance to the educational context and their potential to contribute to the predictive power of the model. By combining demographic information (age, gender, educational background) with psychographic insights (learning preferences, motivations), the dataset aimed to capture a holistic view of students' profiles.

Research Question 2: What major demographic and psychographic factors significantly affect students' academic performance at TVET?

To address the second research question, the study performed extensive data analysis and modeling to uncover the major demographic and psychographic factors that have a significant impact on students' academic performance within TVET institutions. This involved employing machine learning techniques, specifically Support Vector Machines (SVMs), to reveal patterns and relationships within the dataset.

The results of the SVM model demonstrated that certain factors exhibited a stronger correlation with students' academic performance. Through careful feature selection and model evaluation, it was revealed that attributes like age, gender, and specific psychographic preferences were key factors contributing to the predictive accuracy of the model. The findings underlined the importance of considering not only traditional demographic information but also individual learning preferences and motivations in understanding student outcomes.

Research Question 3: What would be the best model that can classify TVET students' performance?

In tackling the third research question, the study delved into the selection of the most suitable machine learning algorithm for classifying TVET students' performance. After thorough consideration, the SVM algorithm was chosen due to its effectiveness in handling multi-dimensional data and its proven ability to classify complex patterns.

The SVM model demonstrated remarkable predictive power, achieving an accuracy rate of 84% on the test dataset. This high accuracy underscored the model's capability to distinguish between different levels of academic performance based on the chosen attributes. The SVM's performance demonstrated that it could successfully generalize its learning from the training data to new, unseen data points, making it a robust and reliable tool for predicting student outcomes.

In summary, the study effectively addressed the three research questions by meticulously crafting a dataset that encapsulated relevant demographic and psychographic factors, identifying key predictors of academic performance through advanced machine learning techniques, and ultimately selecting and demonstrating the efficacy of the Support Vector Machine model for classifying TVET students' performance. The findings shed light on the interplay between personal characteristics and academic success, offering valuable insights for educational institutions seeking to enhance student support and outcomes.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATION

5.1. CONCLUSION

The proposed solutions in this thesis are intended to guide TVET institutions in the development of effective, evidence-based strategies for improving the performance of their students. By improving demographic data collection and analysis, incorporating psychological factors, developing personalized learning programs, and adopting an integrated approach to program development and delivery, TVET institutions can help their students succeed and prepare them for the demands of the 21st-century workforce.

The purpose of the current study was to pinpoint the key elements that TVET students' academic success is influenced by. The study's findings indicate that factors such as gender, family income per month, prior attitudes toward TVET, and department placement satisfaction may have a substantial impact on TVET students' academic success. On the other side, academic success was positively correlated with male gender, family monthly income, past TVET attitudes, teacher satisfaction, and department placement. A negative correlation between stimulant use, and academic achievement was also found. This study shows that a number of variables, including female gender, low family income, shorter study sessions, the use of study aids, negative perceptions of TVET, and low satisfaction with the training sector (department), are linked to lower academic proficiency in Addis Ababa TVET colleges.

5.2. RECOMMENDATION

I. We would like to recommend the following ideas in light of the current study's findings: To help female TVET students compete with their male counterparts, they should get exceptional assistance and encouragement.

II. More study time and fewer stimulant use during class would be beneficial for students. In order to help kids prioritize their academics, teachers and student families should support and encourage them.

III. Since the TVET curriculum contains programs that help students get better at using the English language, it needs to be updated. Additionally, TVET instructors must get continual training in capacity-building.

IV. Through a variety of channels, including the media, the responsible authority must foster in society a positive perception of vocational education and its pupils.

V. Students' freedom of choice of training sector or department should be respected to the utmost extent possible rather than being restricted and limited (in terms of student accomplishment). Students who are uninterested in their chosen subject of study are unlikely to achieve academically because TVET is more practical than general education.

VI. The study's findings may be reviewed and properly incorporated by stakeholders, such as TVET instructors, TVET institution administration bodies, policymakers, etc., to boost students' interest in and academic accomplishment.

VII. Future TVET studies and research by other academics will benefit from the findings of this study.

VIII. The first proposed solution is to improve demographic data collection and analysis. Institutions should ensure that accurate and complete data is collected and stored in a secure and accessible manner. They should also analyze the data regularly to identify trends and patterns that can inform decisions about student support, program development, and resource allocation.

IX. The second proposed solution is to incorporate psychological factors into TVET curricula and programs. Institutions should provide training and resources to instructors and staff on how to identify and support students who may be struggling with psychological issues. They should also develop counseling services and programs to help students address mental health issues and develop coping strategies.

X. The third proposed solution is to develop personalized learning programs for TVET students. Institutions should use the data collected on demographics and psychological factors to provide targeted support and resources to students who need it most. They might use machine learning algorithms or other technologies to create customized learning plans for each student based on their individual needs and learning style.

XI. The fourth proposed solution is to adopt an integrated approach to TVET program development and delivery. Institutions should work collaboratively with industry partners, government agencies, and other stakeholders to identify the skills and competencies needed in the workforce and then design programs and curricula to address these needs. This approach should also incorporate strategies for supporting students' mental health and wellbeing, as well as addressing the specific needs of underrepresented groups.

XII. Conducting various tests with sizable datasets and utilizing various categorization approaches should be considered as future work.

Additionally, intervention should be used to increase kids' overall achievement levels. There may be some explanations for the student's performance that are outside the purview of this investigation. This may be caused by issues inside and outside of the institution, such as low parental involvement in school management, low instructor motivation due to low pay, a lack of student interest in different subjects, and some socioeconomic circumstances affecting the students. Therefore, fostering an environment that facilitates effective teaching and learning helps to guarantee improved academic performance.

Consideration should be given to motivational instructional design in order to increase student engagement with technology. The TVET bureau and other stakeholders should design and develop new technology alternatives to deliver to colleges in accordance with their demand, such as two-way communication and individual, self-directed, and self-paced learning environments, in order to address the intractability and fixed schedule problems of the current broadcasting system.

REFERENCES

[1] Getahun, Kidane A. "A Bayesian Approach to Investigating Factors Influencing Polytechnic College Students' Academic Achievement." Education Research International 2022 (2022).

[2] Makinde, Wasiu, and Tolulope Oluwatosin Bamiro. "Service Quality of Teaching Vocational Education And Training (TVET) And Students Performance In The Federal Polytechnic Ilaro, Nigeria." International Journal of Entrepreneurship and Business Management 1.2 (2022): 116-127.

[3] Verma, Chaman, Zoltán Illés, and Deepak Kumar. "(SDGFI) Student's Demographic and Geographic Feature Identification Using Machine Learning Techniques for Real-Time Automated Web Applications." Mathematics 10.17 (2022): 3093.

[4] Kotsiantis, Sotiris, Christos Pierrakeas, and Panagiotis Pintelas. "PREDICTING STUDENTS'PERFORMANCE IN DISTANCE LEARNING USING MACHINE LEARNING TECHNIQUES." Applied Artificial Intelligence 18.5 (2004): 411-426.

[5] Pojon, Murat. Using machine learning to predict student performance. MS thesis. 2017.

[6] Bressler, Linda A., Martin S. Bressler, and Mark E. Bressler. "Demographic and psychographic variables and the effect on online student success." Journal of Technology Research 2 (2011): 1.

[7] Alsheikh, Fatima Sayed. Student Performance Prediction Using Classification based on their social factors. Diss. Sudan University of Science & Technology, 2021.

[8] Bilal, Muhammad, et al. "The role of demographic and academic features in a student performance prediction." Scientific Reports 12.1 (2022): 12508.

[9] Ahmed, Reda M., Nahla F. Omran, and Abdelmgeid A. Ali. "Predicting and analysis of students' academic performance using data mining techniques." International Journal of Computer Applications 975 (2018): 8887.

[10] Gbollie, Charles, and Harriett Pearl Keamu. "Student academic performance: The role of motivation, strategies, and perceived factors hindering Liberian junior and senior high school students learning." Education Research International 2017 (2017).

[11] Badal, Yudish Teshal, and Roopesh Kevin Sungkur. "Predictive modelling and analytics of students' grades using machine learning algorithms." Education and Information Technologies (2022): 1-31.

[12] Emirtekin, Emrah, Melike Karatay, and T. Kı. "Online course success prediction of students with machine learning methods." Journal of Modern Technology and Engineering 5.3 (2020): 271-282.

[13] Lee, Nick, and Graham Hooley. "The evolution of "classical mythology" within marketing measure development." European Journal of Marketing 39.3/4 (2005): 365-385.

[14]Rozali, Mohd Zulfadli, et al. "Reliability and validity of instrument on academic enhancement support for student-athlete using Rasch Measurement Model." Asian Journal of University Education 18.1 (2022): 290-299.

[15]Viladrich, Carme, Ariadna Angulo-Brunet, and Eduardo Doval. "A journey around alpha and omega to estimate internal consistency reliability." Anales de psicología 33.3 (2017): 755-782.

[16]Umphress, Elizabeth E., John B. Bingham, and Marie S. Mitchell. "Unethical behavior in the name of the company: the moderating effect of organizational identification and positive reciprocity beliefs on unethical pro-organizational behavior." Journal of applied psychology 95.4 (2010): 769.

[17]Ahmad, F., Ismail, N. H. and Aziz, A. A. (2015) „The prediction of students" academic performance using classification data mining techniques", Applied Mathematical Sciences, 9(129), pp. 6415–6426. doi: 10.12988/ams.2015.53289.

[18] Al-barrak, M. A. and Al-razgan, M. (2016) „Predicting Students Final GPA Using Decision Trees: A Case Study", 6(7). doi: 10.7763/IJIET.2016.V6.745.

[19] Al-radaideh, Q. A. (2014) „Mining Student Data Using Decision Trees", (January 2006).

[20]Alaoui, S. S., Farhaoui, Y. and Aksasse, B. (2018) „Classification Algorithms in Data Mining: A Survey", 3(1), pp. 349–355.

[21]Alaoui, S. S., Farhaoui, Y. and Aksasse, B. (2018) „Classification Algorithms in Data Mining: A Survey", 3(1), pp. 349–355.

[22]Badr, A., Din, E. and Elaraby, I. S. (2014) „Data Mining: A prediction for Student "s Performance Using Classification Method", World Journal of Computer Application and Technology, 2(2), pp. 43–47. doi: 10.13189/wjcat.2014.020203.

[23]Baradwaj, B. and Pal, S. (2012) „Mining educational data to analyze student"s performance", Internation Journal od Advamced Computer Science and Applications, 2(6), pp. 63–69. doi: vol.2,No.6.

[24]Barahate, S. R. (2012) „Educational Data Mining as a Trend of Data Mining in Educational System", Proceedings of IJCA International Conference and Workshop on Emerging Trends in Technology, pp. 11–16.

[25]Borkar, S. and Rajeswari, K. (2013) „Predicting students' academic performance using education data mining", International Journal of Computer Science and Mobile Computing, 2(7), pp. 273–279.

[26]Cheewaprakobkit, P. (2015) „Predicting Student Academic Achievement by Using the Decision Tree and Neural Network Techniques", 12(2), pp. 2408–137.

Gadhavi, M. and Patel, C. (2017) „STUDENT FINAL GRADE PREDICTION", 8(3), pp. 274–279.

[27] Govindasamy, K. (2018) „ANALYSIS OF STUDENT ACADEMIC PERFORMANCE USING", 119(15), pp. 309–323.

[28] Hamoud, A. K. and Hashim, A. S. (2017) „Students " Success Prediction based on Bayes Algorithms Students " Success Prediction based on Bayes Algorithms", (November). doi: 10.5120/ijca2017915506.

[29] Hooshyar, D., Pedaste, M. and Yang, Y. (2020) „Mining educational data to predict students" performance through procrastination behavior", Entropy, 22(1), p. 12. doi: 10.3390/e22010012.

[30] J. Kovacic, Z. (2010) „Early Prediction of Student Success: Mining Students Enrolment Data", pp. 647–665. doi: 10.28945/1281.

[31] Kabakchieva, D. (2012) „Student performance prediction by using data mining classification algorithms", International Journal of Computer Science and Management Research, 1(4), pp. 686–690.

[32] Kabakchieva, D. (2013) „Predicting student performance by using data mining methods for classification", Cybernetics and Information Technologies, 13(1), pp. 61–72. doi: 10.2478/cait-2013-0006.

[33] Kaur, G. and Singh, W. (2016) „Prediction Of Student Performance Using Weka Tool", 17(January), pp. 8–16.

[34]Kaur, H. (2015) „EDM: A Review of Applications of Data Mining in the Field of Education", India, 4(4), pp. 409–412. doi: 10.17148/IJARCCE.2015.4492.

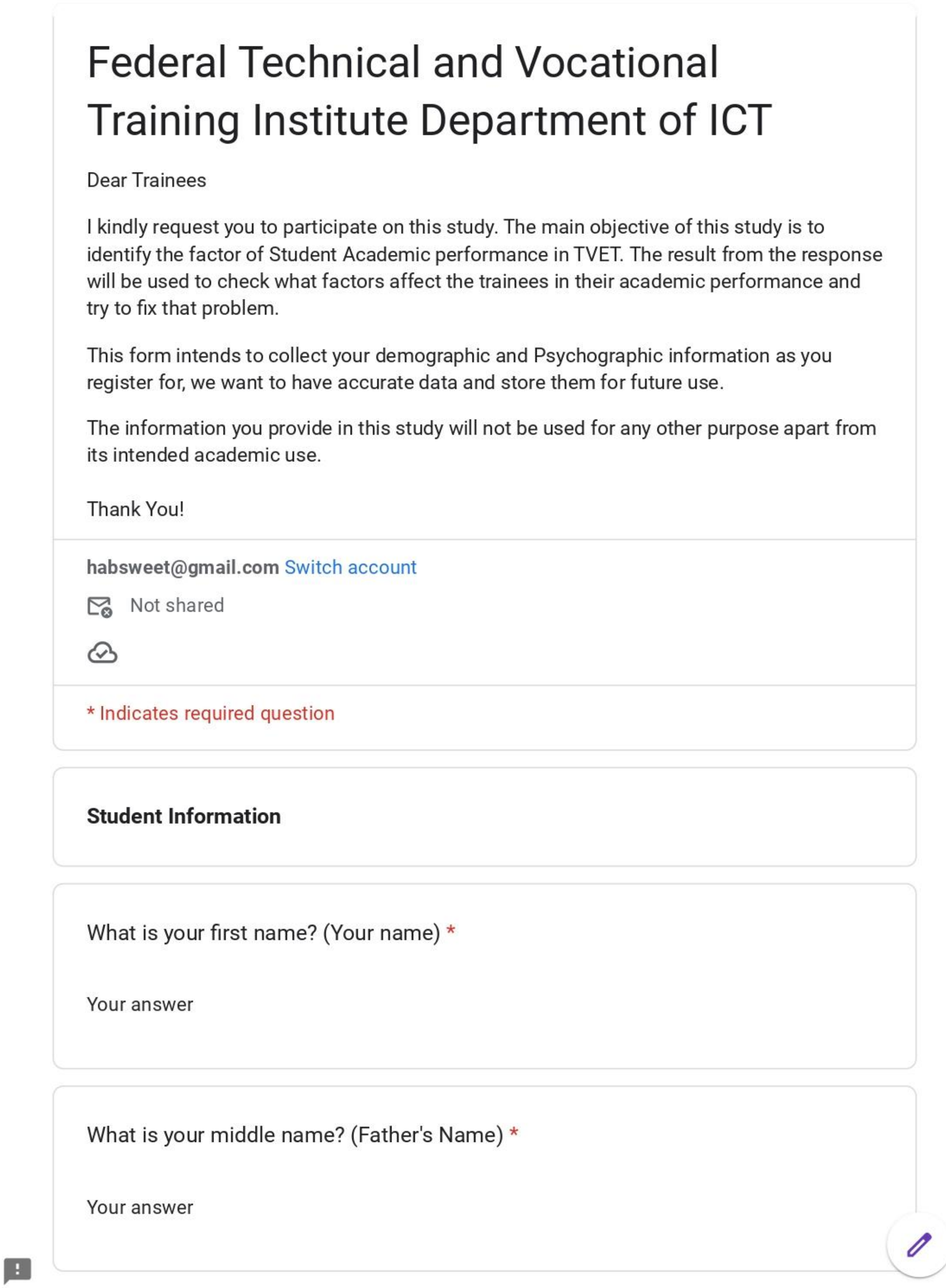
[35] Kumar, S. A. (2011). Efficiency of decision trees in predicting student"s academic performance.

[36] Minaei-Bidgoli, B. (2004) „Data Mining for a Web-Based Educational System", Thesis, p. 267.

[37]Mobasher, G., Shawish, A. and Ibrahim, O. (2017) „Educational Data Mining Rule based

[38]Pang, Bo, Erik Nijkamp, and Ying Nian Wu. "Deep learning with tensorflow: A review." Journal of Educational and Behavioral Statistics 45.2 (2020): 227-248.

APPENDIXES



Federal Technical and Vocational Training Institute Department of ICT

Dear Trainees

I kindly request you to participate on this study. The main objective of this study is to identify the factor of Student Academic performance in TVET. The result from the response will be used to check what factors affect the trainees in their academic performance and try to fix that problem.

This form intends to collect your demographic and Psychographic information as you register for, we want to have accurate data and store them for future use.

The information you provide in this study will not be used for any other purpose apart from its intended academic use.

Thank You!

habsweet@gmail.com Switch account

Not shared

* Indicates required question

Student Information

What is your first name? (Your name) *

Your answer

What is your middle name? (Father's Name) *

Your answer

What is your last name? (Grand-father's name) *

Your answer

What is your Student ID Number? *

Your answer

What is your Gender *

- Male
- Female

Age *

Your answer

Home Location: What is your region (city)? *

Your answer

Home Location: What is your sub-city? *

Your answer

Home Location: What is your woreda?

Your answer

What is your department (Occupation)? *

Your answer

What is your Sector (Field of study)? *

Your answer

Family Background

Note:

If you are living with your family please fill the following questions

What is educational status of your Father

- [ ] PHD
- [ ] Masters
- [ ] Bachelors
- [ ] Higher Diploma
- [ ] Diploma
- [ ] Certificate
- [ ] Other

What is educational status of your Mother

- [ ] PHD
- [ ] Masters
- [ ] Bachelors
- [ ] Higher Diploma
- [ ] Diploma
- [ ] Certificate
- [ ] Other

Job status of your Father

- [ ] Government
- [ ] Private
- [ ] Own Business
- [ ] Other

Job status of your Mother

- [ ] Government
- [ ] Private
- [ ] Own Business
- [ ] Other

If you live alone

What is your source of income

Your answer

How far the distance from your home to your TVET College

- ◯ Near
- ◯ Medium
- ◯ Far
- ◯ Too Far

Do you have a family support for studying

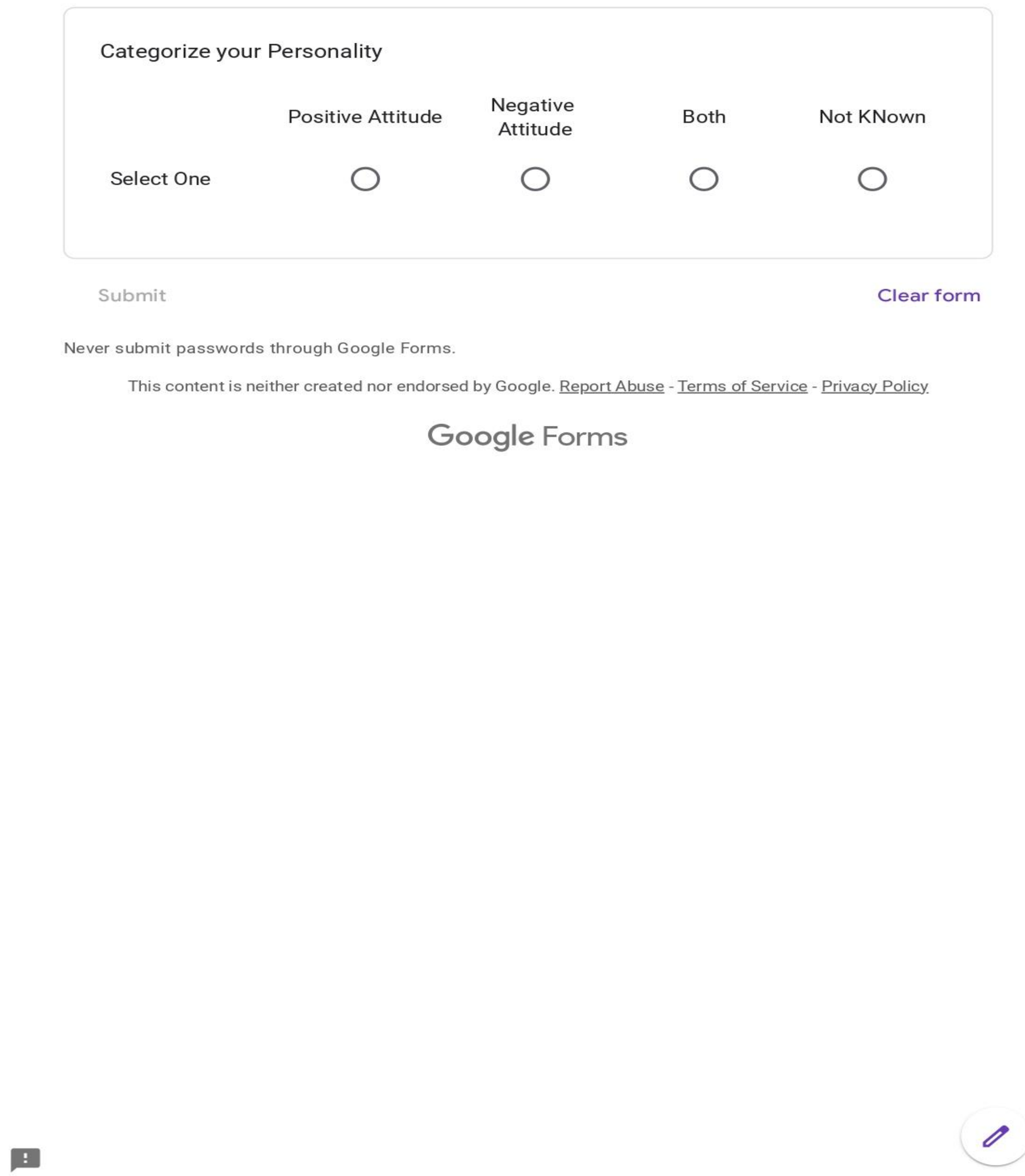
- ◯ Yes
- ◯ No

Are you Interested with your department/occupation

- ◯ Yes
- ◯ No

Does political affiliation affect your education

- ◯ Yes
- ◯ No



Categorize your Personality

	Positive Attitude	Negative Attitude	Both	Not KNown
Select One	○	○	○	○

Submit

Clear form

Never submit passwords through Google Forms.

This content is neither created nor endorsed by Google. Report Abuse - Terms of Service - Privacy Policy

Google Forms

Figure 25: Questionnaires