A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF ÇANKIRI KARATEKİN UNIVERSITY

EFFICIENCY OF STACKED ENSEMBLE LEARNING FOR STUDENT'S ADAPTIVITY CLASSIFICATION TO ONLINE EDUCATION

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRONICS AND COMPUTER ENGINEERING

BY

MATHR ANWAR SHARIF SHARIF

ÇANKIRI

2023

EFFICIENCY OF STACKED ENSEMBLE LEARNING FOR STUDENT'S ADAPTIVITY CLASSIFICATION TO ONLINE EDUCATIONS

By Mathr Anwar Sharif SHARIF

June 2023

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science

Advisor **:** Asst. Prof. Dr. Selim BUYRUKOĞLU

Examining Committee Members:

Chairman **:** Asst. Prof. Dr. Selim BUYRUKOĞLU
Electronics and Computer Engineering
Çankırı Karatekin University

Member **:** Asst. Prof. Dr. Mustafa KARHAN
Electronics and Computer Engineering
Çankırı Karatekin University

Member **:** Assoc. Prof. Dr. Serkan SAVAŞ
Computer Engineering
Kırıkkale University

Approved for the Graduate School of Natural and Applied Sciences

Prof. Dr. Hamit ALYAR
Director of Graduate School

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Mathr Anwar Sharif SHARIF

ABSTRACT

EFFICIENCY OF STACKED ENSEMBLE LEARNING FOR STUDENT'S ADAPTIVITY CLASSIFICATION TO ONLINE EDUCATION

Mathr Anwar Sharif SHARIF

Master of Science in Electronics and Computer Engineering

Advisor: Asst. Prof. Dr. Selim BUYRUKOĞLU

June 2023

In the post-Corona educational context, it has become vitally important to assess the adaptability of students to online learning to ensure optimized educational outcomes. The seismic shift to digital learning platforms has intensified the need to investigate student adaptability and potential methodologies to gauge it accurately. This study seeks to extend our understanding of students' adaptability to online learning by applying various machine learning models. Through the development, comparison, and application of these models, this research intends to identify the best technology and most potent model that could enhance the education process significantly. The core objective of this research is to expand our knowledge on students' adaptability in the digital age using machine learning techniques, thus providing more nuanced insights into their online learning dynamics. For this purpose, four distinct machine learning models were formulated, each incorporating a specific type of machine learning algorithm, namely Bagging, Boosting, Stacking, and a model comprising multiple algorithms. The study's data set, named 'Student's Adaptability Level in Online Education,' encompasses 14 attributes and 1205 instances. These attributes provide an in-depth insight into students' online learning capabilities and their adaptability levels. They are chosen with a focus on providing a comprehensive understanding of the different facets of online learning adaptability. To construct the models, four feature selection methods were deployed: Relief F, ANOVA, Information Gain, and Chi-square. The chosen feature selection techniques are known for their efficacy in selecting and ranking features based on their importance,

thus aiding in model precision and eliminating redundancy. The models were evaluated using a 10-fold cross-validation process with classifiers in all models. This method is known for providing reliable estimates of model performance while minimizing the risk of overfitting, hence leading to the production of more accurate and robust machine learning models. Among the four models, the Gradient Boosting algorithm using the Chi-square technique demonstrated superior performance, reaching an accuracy of 0.866%. This indicates that the combination of boosting and Chi-square technique can accurately assess students' adaptability to online learning, suggesting its potential utility in the education sector. Further, the study found that the stacked learning model, employing Boosting classifiers at level-0 and Logistic Regression at level-1 using the Chi-square technique, yielded the highest accuracy rate of 0.874%. The results imply that the stacked model's application could provide an even more accurate assessment of students' adaptability to online education. This study, thus, introduces a potentially more accurate, automated, and efficient way of assessing students' adaptability to online learning. The findings not only enhance our understanding of student adaptability but also highlight the benefits and potential of applying machine learning techniques in education. With further advancements, such applications can contribute significantly to personalized learning, education management, and policy development in the era of online education.

2023, 62 pages

Keywords: Online teaching, Students adaptivity, Ensemble learning, Students classification, Stacked ensemble

ÖZET

ÖĞRENCİNİN ÇEVRİMİÇİ EĞİTİMLERE UYUMLULUK SINIFLANDIRMASI İÇİN YIĞINLANMIŞ TOPLULUK ÖĞRENİMİNİN VERİMLİLİĞİ

Mathr Anwar Sharif SHARIF
Elektronik ve Bilgisayar Mühendisliği, Yüksek Lisans
Tez Danışmanı: Dr. Öğr. Üyesi Selim BUYRUKOĞLU
Haziran 2023

Corona sonrası eğitim bağlamında, öğrencilerin çevrimiçi öğrenmeye uyumluluğunu sağlamak için optimum eğitim sonuçlarına ulaşmak hayati önem taşımaktadır. Dijital öğrenme platformlarına doğru olan ani geçiş, öğrenci uyumluluğunu ve bu konuyu doğru bir şekilde ölçebilecek potansiyel yöntemleri araştırma ihtiyacını yoğunlaştırmıştır. Bu çalışma, çeşitli makine öğrenmesi modellerini uygulayarak öğrencilerin çevrimiçi öğrenmeye uyumluluğu hakkındaki anlayışımızı genişletmeyi hedeflemektedir. Bu modellerin geliştirilmesi, karşılaştırılması ve uygulanması aracılığıyla, bu araştırma, eğitim sürecini önemli ölçüde geliştirebilecek en iyi teknolojiyi ve en güçlü modeli belirlemeyi amaçlamaktadır. Bu araştırmanın temel hedefi, makine öğrenmesi tekniklerini kullanarak öğrencilerin dijital çağdaki uyumluluğu üzerindeki bilgimizi genişletmektir, böylece çevrimiçi öğrenme dinamikleri hakkında daha ince ayrıntılı bilgiler sağlar. Bu amaçla, her biri Bagging, Boosting, Stacking ve çoklu algoritmalardan oluşan bir model olmak üzere dört farklı makine öğrenmesi modeli oluşturulmuştur. Çalışmanın veri seti, 'Çevrimiçi Eğitimde Öğrencinin Uyumluluk Seviyesi' adını taşır ve 14 özellik ve 1205 örnek içerir. Bu özellikler, öğrencilerin çevrimiçi öğrenme yeteneklerine ve uyumluluk seviyelerine derinlemesine bir bakış sağlar. Çevrimiçi öğrenme uyumluluğunun farklı yüzlerini kapsamlı bir şekilde anlamak amacıyla seçilmişlerdir. Modelleri oluşturmak için, Relief F, ANOVA, Bilgi Kazancı ve Ki-kare olmak üzere dört özellik seçme yöntemi kullanılmıştır. Seçilen özellik seçme teknikleri, özellikleri önemlerine göre seçme ve sıralama konusunda etkinlikleriyle bilinir, böylece model

hassasiyetine yardımcı olur ve gereksizliği ortadan kaldırır. Modeller, tüm modellerdeki sınıflandırıcılarla 10 katlı çapraz doğrulama işlemi kullanılarak değerlendirilmiştir. Bu yöntem, model performansının güvenilir tahminlerini sağlarken aşırı uydurmanın riskini en aza indirir, böylece daha doğru ve sağlam makine öğrenmesi modellerinin üretilmesine yol açar. Dört model arasında, Ki-kare tekniğini kullanarak Gradient Boosting algoritması, %0.866'lık bir doğrulukla üstün performans göstermiştir. Bu, boosting ve Chi-kare tekniğinin kombinasyonunun, öğrencilerin çevrimiçi öğrenmeye uyumluluğunu doğru bir şekilde değerlendirebileceğini göstermektedir ve bu da eğitim sektöründe potansiyel bir kullanımını önermektedir. Ayrıca, çalışma, Ki-kare tekniğini kullanarak level-0'da Boosting sınıflandırıcılar ve level-1'de Lojistik Regresyon uygulayan istiflenmiş öğrenme modelinin, %0.874'lük en yüksek doğruluk oranını elde ettiğini bulmuştur. Sonuçlar, istiflenmiş modelin uygulanmasının, öğrencilerin çevrimiçi eğitime uyumluluğunun daha doğru bir değerlendirmesini sağlayabileceğini göstermektedir.
Bu çalışma, böylece, öğrencilerin çevrimiçi öğrenmeye uyumluluğunun değerlendirilmesinde potansiyel olarak daha doğru, otomatik ve verimli bir yol sunmaktadır. Bulgular, öğrenci uyumluluğu hakkındaki anlayışımızı sadece geliştirmekle kalmaz, aynı zamanda eğitimde makine öğrenmesi tekniklerinin uygulanmasının faydalarını ve potansiyelini de vurgular. Daha ileri gelişmelerle, bu uygulamalar, kişiselleştirilmiş öğrenme, eğitim yönetimi ve çevrimiçi eğitim çağında politika geliştirmeye önemli ölçüde katkıda bulunabilir.

2023, 62 sayfa

Anahtar Kelimeler: Çevrimiçi öğretim, Öğrenci adaptasyonu, Toplulukla öğrenme, Öğrenci sınıflandırması, Yığılmış topluluk

PREFACE AND ACKNOWLEDGEMENTS

In the beginning, I would like to dedicate this effort to the soul of my beloved mom. Additionally, I would like to express my gratitude to my thesis advisor, Asst. Prof. Dr. Selim BUYRUKOĞLU, for his patience, guidance, and understanding.

Mathr Anwar Sharif SHARIF
Çankırı-2023

CONTENTS

LIST OF ABBREVIATIONS

ANN	Artificial neural network
AUC	Area under curve
DT	Decision tree
EDM	Educational data mining
FP	False positive
FS	Feature selection
IG	Information gain
KNN	K-Nearest neighbor
LR	Logistic regression
ML	Machine learning
NB	Naïve bayes
RF	Random forest
SVM	Support vector machine
TP	True positive

LIST OF FIGURE

LIST OF TABLES

1. INTRODUCTION

Online educational systems are now becoming part of mainstream education. The pandemic of Covid -19 forced schools and educational institutions to be compelled to make appropriate modifications to be able to continue to deliver education. Thus, the teaching and learning activities were shifted to online and e-learning (Amir *et al.* 2020).

The development of education technology has proven its superior ability to face societal, environmental, and health challenges. This technology has provided its services (in light of the pandemic) represented by the many educational platforms that have been widely used over more than two years. These platforms have also taken up a large space and made a qualitative leap in the field of education to this day.

It cannot be overlooked that there are obstacles that faced this type of education and the sudden and total transformation of it. Many students felt stress and anxiety during the pandemic and transforming education into online learning. Psychological issues and uncomfortable often delay students from adapting to online education (Chakraborty *et al.* 2021). Also, not all students, and even some of the teaching staff, were fully aware of how to use online learning platforms. So, this thing took time to become easy to deal with.

In spite of these challenges of turn, online learning enhanced the learning program for students. Where internet technology makes it possible to distribute the content of education to a large number of educated at the same time (Baashar *et al*. 2021). Online learning platforms also offer many advantages to learners such as control over the time spent learning and control over the content. Thus, the teaching process can be adapted according to the objectives of learning (Coman *et al.* 2020).

The existence of these aforementioned challenges forced researchers and those interested in the educational field to know the extent to which students adapt to e-learning, and this is the focus of interest and significance of this study. Where this study tried to classify the level of students' adaptation to online learning using ensemble learning methods. Using ensemble machine learning, four models were developed: a model that combines various algorithms (single, bagging, and boosting), a model for single algorithms, a model for bagging algorithms, and a final model for Boosting algorithms. All of these models have developed a stack model.

The results of these models were compared with each other, depending on the known performance measures in evaluating the classification algorithms. To the knowledge of the researcher, there is no study that used the stacked model to find out and classify the level of students' adaptation to online learning. Therefore, this study was unique to many goals, motives, and results that will be mentioned later.

This study used a dataset (Students' Adaptability Level in Online Education), from Bangladesh, from December 10, 2020, to February 5, 2021. This sample included 1205 students including a group of males and females in basic and academic education, as well as governmental and non-governmental educational institutions. The methods of selecting the features were also used on the data set in order to extract the important and influential features in improving the results.

1.1 Motivation

- Developing a model based on data mining by using stacked ensemble learning to classify student adaptation to online education.
- Anticipating the adaptation of students to online education. That adaptation allows parents and teachers to intervene early and take some actions such as guiding students and monitoring their progress, and trying to create conditions and make possible changes. Also, that helps in the development of smart teaching systems and the necessary decision-making process by the school administration.

- Student learning outcomes are goals that measure the extent to which they acquire knowledge, skills, and values. It is a more comprehensive measure of student performance than just assessment scores.

1.2 Aims and Objectives

The primary aim is to contribute to clarifying and understanding the effectiveness of online learning by analyzing the relevant factors related to online education. One of the aims is to provide educators with what they can do to support all students in their classroom education and adapt them to it.

1.3 Objectives of the Study

The objectives are:

- The use of smart methods and techniques has been developed to classify students' compatibility with the online learning situation.
- The use of feature selection improves the understanding of the problem. A reduction in dataset dimensionality improves performance either by increasing generalization capacity and classification accuracy or by decreasing model complexity and learning speed.
- Measuring the efficiency of stacked ensemble learning according to student adaptation to online education. The stacked ensemble classifier provides a useful solution by training different basic classifiers called level-0 and then combining their outputs with a Meta classifier called level-1.
- Comparison of the performance of models and techniques used in the study in different aspects, including their accuracy, strengths, and weaknesses.

1.4 Research Questions

The main purpose of this research is to explore how to use ensemble learning algorithms to classify students' adaptively to online learning. To achieve this goal, the research tries to answer some research questions (RQ) that are prepared are:

RQ1: What types of ensemble learning algorithms are used to classify students' adaptively?

RQ2: What is the best ensemble learning algorithms to classify students' adaptation?

RQ3: Can feature selection techniques be useful in improving the performance of the algorithms and raising the results values?

RQ4: Can the stacked ensemble learning approach provide a better result than the others in classification for student adaptivity to online education during covid - 19?

1.5 Contributes

- Analysis of the data set to help educational institutes improve curricula, and activate the role of the teaching staff.
- Determination of the optimal feature selection technique in order to obtain the highest classification results.
- Use machine learning algorithms, implement many stacked ensemble learning methods, and choose the optimal stack model to guide the researchers and those interested in developing the educational field.
- Enhance online learning and its role by understanding the status of students' adaptation to the online learning curriculum.

1.6 Thesis Organization

The first part of this chapter (introduction) contained the motives and reasons that motivated the researcher to carry out this study, as well as the important questions

that the content of this study would answer. The second chapter is (literature review), which was divided into four parts according to the results of this literature and according to the type of algorithms used in it. The third chapter (method), it contained a mention of the data set that was used and the inclusion of a Table representing the features of this data, then the methods for feature selection and how they were used, and an overview of them. As well as an explanation of the algorithms used and the mechanism of their work and the parameters by which these algorithms were established, as well as an explanation of the idea of developing the models that were established. Then the fourth chapter (results) contains the results of the algorithms used and their evaluation measures. Finally, the fifth chapter (conclusions), which analyzes, discusses, and summarizes the results drawn from the study and the most important recommendations for future studies that concern the same topic.

2. LITERATURE REVIEW

2.1 Introduction

The use of educational data mining has allowed the assessment and classification of datasets of students and their academic levels (Bhusal 2021). All that is for various analytical purposes such as classification, predicting and forecasting, segmenting, object detection, etc. They influence many industries and sectors, including whether a student's engagement and learning performance can impact their academic success (Sandra *et al.* 2021). The field of educational data mining has aroused the interest of many researchers and research institutions.

The development of ensemble techniques in the fields of machine learning and data mining has drawn a lot of interest from the scientific community over the past several decades. In machine learning ensemble approaches, different learning algorithms are combined to provide higher predicted performance than any one of the individual learning algorithms could (Buyrukoğlu and Savaş 2022). Theoretically and experimentally, it has been demonstrated that combining several learning models results in much greater performance than using only one. Ensemble learning algorithms have been used to solve a variety of real-world problems, including face and emotion recognition, text classification, medical diagnosis, and financial forecasting, including classification areas. These algorithms are considered to be a dominant and cutting-edge method for achieving maximum performance in the literature (Yağcı 2022).

In all educational institutions, a very accurate prediction of student performance is helpful in a variety of situations for detecting slow learners and differentiating between students with poor academic accomplishment and weak pupils who are likely to have low academic achievement (Charbuty and Abdulazeez 2021). Teachers, parents, and educational planners will benefit from the models' final product by learning if a student's conduct may be linked to good or bad outcomes in the past and by receiving advice on how to fix issues (Ghosh and Janan 2021). The

analysis of students' data using data mining methods contributes to reducing the failure of educational institutions and raising their educational value among the circles.

In this chapter, the researcher reviews various studies with recent publications, which are related to the subject of classifying students, divided according to the types of algorithms used for classification.

2.2 Overview of Students' Classification

Since the success of educational institutions depends mainly on their outputs in the sense of the number of successful students and their high rates, so the interest in studying the classification of students' performance has attracted the attention of researchers in this field (Chen *et al*. 2019). Educational data mining is the best way to achieve this goal. The classification was the data mining approach used most frequently (Kawade *et al.* 2020).

In general, the classification process is formulated based on two parts: the first is the data set that is being worked on, and the second is the algorithm or set of algorithms that are used to build the model. Classification enables the most accurate prediction of the target class (Dhankhar and Solanki 2021). To build a model that is being trained, the classification algorithm determines the relationship between the input attribute and the output attribute (Sudais and Asad 2022).

In order to maintain track of their students, teachers, and courses, educational institutions, is retained a lot of info (Farhana 2021). This information includes personal and academic details regarding teachers, students, and other parties, as well as syllabus, test questions, circulars, and other materials (Fahd *et al*. 2021). To better the lives of their students and professors, several institutions and independent groups have started using educational data mining for students' classification (Ragab *et al.* 2021).

One of the key areas for research in machine learning is classification. To classify students, supervised learning and features chosen via feature selection methods are used, which can also be assessed using classification algorithms (Smirani and Boulahia 2022).

2.3 Proposed Machine Learning Approaches for Students' Classification

Due to the wide use of artificial intelligence, facilitations have become apparent in most areas of life (Alamri *et al*. 2020). In the education field, Artificial Intelligence based learning systems needs enhancement and improvements to help the students in homes or classrooms as well as people who required private education (Pallathadka *et al.* 2021). The study of educational data makes for improvements in the student's performance. Education Data Mining (EDM) approach analyses the data of students. This information is analysed through the classification of this data using a decision tree or association rules.

Education aids students in learning and expanding their existing knowledge, while AI offers clever ways to comprehend mechanics and intelligent behaviour (Soni *et al.* 2018). In the modern technology context, it is believed that fixed classrooms, repeating lectures, and static printed textbooks are insufficient and ineffective. People who use technology often, in particular, are not interested in traditional classroom settings or printed textbooks (Hassan *et al*. 2020). One of the most significant requirements for any educational institute is the performance of its students. The performance of students can be anticipated based on their past academic performance. The results suggest that student's interests and skills may influence how well they achieve (Hussain and Khan 2021). This kind of analysis enables teachers to concentrate more of their attention on the students who most require it. The performance of a teacher's students is commonly used to gauge the teacher's success. Each institution must evaluate the quality of its professors (Karo *et al*. 2022). The performance, feedback, and other factors of a student's performance can be used to evaluate a teacher. An institution can enhance the form of its instruction with the help of this kind of investigation (Siddique *et al.* 2021).

2.3.1 Single-based algorithms

A study developed by (Perez and Perez 2021) predicts students' program completion using the Naïve Bayes classifier technique. After pre-processing, cleaning, balancing, and transforming this dataset, ten predictors were used for predicting. The variables assessed with the feature selection technique, (Weight by Correlation) were the last inputs used to build the model. Only four (4) predictor factors were chosen following correlation analysis out of a potential ten (10) predictor variables. The following major factors were shown to have an impact on program completion: parents' monthly income, parents' educational attainment, and High School GPA factors. In the test dataset (n=191), the accuracy values produced an 84 percent rating with an 80.46 percent class precision and an 83.33 percent class recall. It is important to consider the zero-probability problem, which confronts the model based solely on this classifier. Therefore, it is advisable to ensure that there is no specific category in the test data and that it does not exist in the training data.

Another study by (Jawthari and Stoffová 2021) used an updated method of the K-Nearest neighbour (KNN) algorithm. The proposed technique has addressed the problem of subjectively encoding nominal variables and the problem of the majority vote. A straightforward and efficient KNN method is developed by putting two similar metrics. Nominal variable encoding is not necessary for the procedure. The distance weight vote rule, which gives the closest neighbours more weight, is another factor taken into account by the improved technique. The best accuracy resulting where k = 1 from standard KNN was 66% while the best version was the one that used the Jaccard vote where this version has 80% accuracy resulting from k = 6. As a result, the suggested KNN algorithm performs more accurately than regular KNN. The suggested strategy improved the KNN's sensitivity to outliers by taking alternate voting rules into account. The contribution of this work is the creation of a distance function for classification decisions without the need to convert nominal variables to numeric values. Experiments revealed that the suggested technique, which uses the Jaccard distance, consistently beat the traditional KNN by 14%. The upgraded KNN method performed well in terms of accuracy but lagged behind scikit-learn KNN in

terms of speed. The limitations of the study were that the updated algorithm was slow compared to scikit-learn KNN. So, it's better to improve the speed of the algorithm by incorporating fast comparing techniques.

Two classification methods were used in a study by (Tripathi *et al.* 2019), Naive Bayes and Support Vector Machine algorithms. For the examination of the overall effectiveness of the educational domain, these categorization systems produced high-quality and realistic results. In this study, the Naive Bayes classifier has higher accuracy (92%), in comparison to the Support vector machine (SVM) classifier (84%). The Support Vector Machine classifier's execution time is contrasted with that of the Naive Bayes more quickly than the Support Vector Machine classifier. This module makes use of the Apache Prediction IO machine learning server. The prediction IO platform, utilized in the creation, assessment, and usage of search engines, was one of the sub-components. Additionally, an event Server, an open-source machine learning logical layer that brings together a variety of platforms, was employed.

The purpose of the study by (Lagman *et al.* 2019) is to determine how educational data mining, especially student graduation, may make use of the Naive Bayes method. The goal of the project is to develop a model that can early forecast and identify students who are likely to graduate late by using the Naive Bayes algorithm to predict student graduation. The development of the Naive Bayes Testing using Training Sets and the Naive Bayes Data Model are the two key components of the empirical testing for Nave Bayes. Waikato knowledge analysis, a data mining tool, was used to assess naive Bayes analysis on all of the proposed predictors. The overall acceptability of the Descriptive and Predictive Analytics of Student Graduation Prototype based on the respondents' responses recorded an overall mean of 4.69 which has an interpretation of very acceptable and can be concluded that the software can now be used for implementation. The Nave Bayes algorithm has an accuracy rate of 85.22 in predicting student graduation. The limitation of the study is the use of only one algorithm, as it was possible to apply a set of algorithms to test the data set.

Regarding the study by (Lau *et al.* 2019), it described a technique that combined traditional statistical analysis with neural network modelling and performance prediction. Eleven input variables, two layers of hidden neurons, and one output layer make up the neural network's model. The error performance, regression, error histogram, confusion matrix, and area under the receiver operating characteristic curve are used to assess the performance of the neural network model. The neural network model was able to predict events with an accuracy of 84.8 percent. The MathWorks MATLAB program was used to carry out the artificial neural network (ANN) modelling and assessments. There were eleven variables in the input layer. The modelled ANN has two hidden layers, with 30 neurons in each hidden layer. This design has proven to produce the best results over several simulations with various hidden layer and neuron values. The study lacked some important features in the process of analysing and evaluating students' levels, including the role of lecturers and other classrooms that were not included in the study.

After reviewing a section of the scientific papers related to the subject of the researcher's study, regarding the single-based algorithm, in the next part, the researcher reviews a section of the scientific papers related to the bagging algorithms.

2.3.2 Bagging algorithms

The first study (Yağcı 2022), looks at forecasting pupils' academic performance only based on grades, without considering socioeconomic or demographic factors. The goal of this project was to create a novel model based on, machine learning algorithms to forecast undergraduate students' final exam scores based on their midterm test marks, faculty, and department. Machine learning classification algorithms were used to identify the classification algorithms that performed the best in predicting students' academic success.

The following methods were used to predict students' academic performance: RF, ANN, Logistic regression LR, SVM, Naïve Bayes NB, and kNN. Ten-fold cross-

validation was used to assess the prediction accuracy. This study focused on the reasons for success and failure; by using statistical methods such as time series and logistic regression. Also, using neural networks, decision trees, and vector machines, has taken into account the predictions in a strong way. It also increased the efficiency and accuracy of the results by using random forests. This study relied on programming and developing its model by orange machine learning software.

The random forest algorithm achieved the highest classification accuracy of 75%. Additionally, just three types of characteristics were used to make the predictions: faculty data, department data, and midterm test marks. It was possible to add other variables in the entry for comprehensiveness and expansion of the research area. The limitations of the study are the failure to take into account the students' environment and the situation of their families and note the impact of these variables on their academic achievement. In addition, the study did not include any boosting algorithms and was limited to single-based and bagging algorithms.

Another study by (Ahmed and Sadiq 2018), the dataset contained from the students was utilized to extract relevant information using the Random Forest classification technique. The new data has been put to the test on all of the forest's trees to forecast pupils' success. The outcome of this algorithm had an accuracy of 83.56%. On the other side, using an identical methodology and dataset, the WEKA tool reports an accuracy of 80.82%. The Random Forest algorithm's suggested approach has also been evaluated using data from earlier research. According to the comparative results, the suggested approach may provide an accuracy of 92.65%, which is higher than the accuracy of 91.2% achieved by another study that was conducted. The Random Forest algorithm is a clear example of a black box algorithm, which is the bedrock of this research. The negative of this study is the small number of samples in the data set, which did not exceed 73 students.

In another study by (Mythili and Shanavas 2014), explores the potency of machine learning algorithms in the influence on the results of students. Using classification algorithms on the dataset, estimating the precision of the resultant prediction model,

and visualizing the model were all made possible by the classification panel. Weka was used to enforce the decision tree classifier C4.5 (J48), Random Forest, Neural Network (Multilayer Perceptron), and Lazy-based classifier (IB1). The 10-fold cross-validation is selected under the "Test settings" section. The results of such a classification model deal with accuracy, confusion matrices, and execution time. The Random Forest algorithm had the highest accuracy (89%). Thus, the conclusion could be reached that the Random Forest performance was better than that of different algorithms.

The study by (Hasan *et al.* 2018), used the Decision Tree method to determine student performance and Moodle access time is evaluated using the WEKA data mining tool. The results of simulations reveal that the Random Forest Tree method was more accurate than other Decision Tree algorithms. In this work, different classification algorithms were examined, including Decision Tree J48, REP Tree, Random Forest, Logistic Model Tree, Hoeffding Tree, Decision Stump, and Naive Bayes. It might assist the stakeholders in making adjustments and initiating early intervention to boost student experience and enhance module performance. It was discovered that Naive Bayes, and Random Forest provide a good agreement for the training set. Due to the module's reduced mean absolute error and relative absolute error, while Random Forest's accuracy was 100%, it was determined that Random Forest was the most appropriate choice. The group of students for this study, numbering only 22 students, makes that one of the limitations.

A hybrid model was put out in this study (Ghosh and Janan 2021), to forecast first-year university students' success on their final exams. The RFC (Random Forest Classifier) is used with the fuzzy ANFIS (Adaptive Neuro-Fuzzy Inference System) to create that model. In comparison to other conventional machine learning algorithms, the experimental findings showed that the suggested model is useful and practical for the precise prediction of students' development. The RF model's prediction accuracy is 96.88 percent accurate. Additionally, it has been discovered that two important elements that significantly influence students' achievement are their effort and their ability to write quality notes. Limiting the work to a single

algorithm constrains the results, so it is possible to implement a support vector machine (SVM), deep learning algorithms like Recurrent Neural Networks (RNNs), and Convolutional Neural networks (CNN), for comparative analysis.

After completing a review of some scientific papers related to the bagging algorithms in the field of classifying the students' level, there are reviews papers related to boosting algorithms in the same field.

2.3.3 Boosting algorithms

The first study by (Elden *et al.* 2013), a fast-learning method employing AdaBoost ensemble merging and a straightforward genetic algorithm known as "Ada-GA" were evaluated. In addition to researching the issue of employing a web-based tutoring system to predict student achievement (ASSISTments Platform dataset). It was shown that the genetic algorithm successfully increased the accuracy of the combined classifier performance. Particularly in very big classrooms, the Ada-GA approach proved to be quite helpful in detecting at-risk individuals early on. The outcomes demonstrated that this technique has effectively increased detection accuracy while lowering computing complexity. The use of a genetic algorithm with boosting in this study is another alternative boosting technique that produces better solutions (fewer weak classifiers and a slight increase of classification accuracy) than the classical AdaBoost produces. As well as better than the rest of the algorithms that have been tested on the data set used in this study, which includes (J48, VFI, IBk, Naïve Update, and k-Means). The accuracy of the proposed algorithm was 82.07%, slightly higher than the Ada boost algorithm, which had an accuracy of 81.85%. It is also higher than the rest of the algorithms that have been tested on the data set.

The classification models in this study by (Ketui *et al.* 2019), were applied to identify suiTable subjects for science students. The experiment was set to improve the student's performance that comparing the performance of five classification models, and after that estimating the appropriate academic success in each major. The study used 17,875 academic achievements of 483 students. The classifiers used

include Decision Tree, Decision Tree Weight-based, Iterative Dichotomiser 3, Random Tree, and Gradient Boosted Tree. There are two classifications assigned to each leaf node (Good and bad). In this experiment, a cross-validation test known as 10-fold cross-validation was used to partition the dataset. The classification findings showed that the Gradient Boosted Tree (GBT) algorithm score of 92.41%, is the best model, with the Decision Tree model coming in second (91.03%). The weight-based Random Tree and Decision Tree are both at the lowest. Non-employment of any of the single-based algorithms is one of the limitations of this study, as several are characterized by their high classification accuracy.

Another study by (Hu and Song 2019), used XGBoost algorithm to locate additional courses. The main idea of the study was to transform the score prediction regression problem into a classification issue. Prediction the fourth semester's scores by using the outcomes of the preceding three semesters. The fifth semester's outcomes were predicted using the data from the preceding four semesters. Then the outcomes of the sixth semester were predicted using the grades from the preceding five semesters. The study sample was 145 students. The prediction for XGBoost algorithm in this study accuracy is over 55%. The limitations of this study are that not to use another algorithm and compare its results with an XGBoost algorithm to evaluate its performance.

In this study (Guang-yu and Geng 2019), XGBoost gradient lifting decision tree algorithm was used. The data was collected on learning and life behaviour of the first semester and weighted average scores. The number of students was 566. This study explores the link between student education indicators and course performance using data mining technology based on the XGBoost gradient lifting decision tree method. It also examines the internal relationship between student conduct and academic success. The results of this study, students with high performance in discipline competitions, science, and technology competition, and academic workshops generally have higher academic performance. A comprehensive analysis was done on the weights of 10 Decision Trees' significant characteristic characteristics. The average information gain was used to indicate the significance of each XGBoost

model. The predictive accuracy rate was 73%. The limitations of this study did not contain the important means of evaluating the model, such as F1, Recall, and many others. It relied on mentioning the accuracy of the model only.

2.3.4 Proposed stacked ensemble learning for student classification

In fact, not many studies have developed a stacked model for classifying students' levels. The researcher summarized everything available in this field. The first study by (Smirani *et al.* 2022), proposed a model called Stacked Generalization for Failure Prediction (SGFP) to improve students' outcomes. The SGFP model employs a Multilayer Perceptron to combine three ensemble learning classifiers: Light Gradient Boosting Machine (LGBM), Extreme Gradient Boosting machine (XGB), and Random Forest (RF) (MLP). The model depends on high-quality training and testing datasets that are automatically gathered from the Blackboard Learning Management System's analytic reports. Student categorization into three performance-based groups is the primary output of SGFP (class A: above average, class B: average, class C: below average). The SGFP model employs the Blackboard Adaptive Release tool to create three learning pathways that students must automatically follow based on the class to which they are assigned in order to prevent failures. The SGFP model and basic classifiers were contrasted (LGBM, XGB, and RF). The results demonstrate that SGFP has greater mean and median accuracy. Additionally, it successfully classified pupils with an average sensitivity and accuracy of 97.3% and 97.2%, respectively. In addition, SGFP had the highest F1 score, coming in at 97.1%. Results showed that student success rates in the academic year 2020–2021 had significantly increased. The failure rate jumps from 12% to 1.14% for class C students, for whom more evaluation procedures and materials are tailored. Only two out of 176 pupils who followed laid out learning routes failed. The success percentage was 98.86% as a result. It is possible to adopting different categorization models and algorithms, as well as assessing and contrasting the outcomes.

In a different study by (Sridhar *et al.* 2020), a stacked ensemble model that forecasts a student's odds of admission to a certain university has been proposed. The

suggested model considers several elements of the students. The dataset used consists of the applicant scores. The data scraping was achieved by the python library 'BeautifulSoup'. It had a total of 50, 000 samples and 22 features. The proposed neural network is assessed by comparing it with existing supervised algorithms like Random Forest, Decision Trees, K-Nearest Neighbour, Logistic Regression, Naive Bayes Classifier, Support Vector Machine (SVM), Quadratic Discriminant Analysis, and Linear Discriminant Analysis. The proposed ensemble neural network outperforms all existing supervised learning techniques in terms of accuracy. Level-1 contains five sub-models, each one having the same architecture. All of them were Neural Network sub-model that use the 'ReLU' activation function. There are 5 hidden layers, each having 64, 128, 128, 64, and 32 nodes respectively. Level-2 also used Neural Network as a meta-learner. The level 1 ensemble model generates a two-level optimal link between the output of the level 0 ensemble model and the input of the level 1. With an accuracy of 91%, the suggested approach offers the best performance.

In another study by (Niyogisubizo *et al.* 2022) related to education, suggested a unique stacking ensemble based on a combination of Feed-forward Neural Networks (FNN), Extreme Gradient Boosting (XGBoost), Gradient Boosting (GB), and Random Forest (RF) to predict student dropout from university courses. When employing testing accuracy and the area under the curve (AUC) assessment metrics under identical circumstances, the suggested technique has shown to perform better than the base models. The first data set includes 261 samples and 12 characteristics of students who registered for the database systems primer class. The training and testing datasets were created by randomly dividing the input data into 80% and 20% halves. The over-fitting issue is addressed via ten-fold cross-validation. The suggested strategy outperformed the base models with overall accuracy, recall, and F1-Score outcomes of 0.93, 0.93, and 0.92, respectively. Results for training accuracy span an 86.67% to 96.67% range. The testing accuracy varies from 76.67% to 92.18%. However, the suggested approach performs better in terms of prediction, with a training accuracy of 93.59% and a testing accuracy of 92.18%. This study did not take into account other influencing factors that may affect the results, as the

students' data that were used lacked the environmental, life, and psychological factors of the student. These factors can be added in future studies and compared with this study in order to improve the results. Additionally, other methods of computing, such as deep learning and other hybrid models, can be utilized to forecast student dropout.

In this chapter, a set of studies are reviewed that deal with methods of classifying students' levels using all kinds of ensemble machine learning: single-based, bagging, boosting, as well as stacked ensemble learning. In general, the evaluation results of these studies were decent, making the use of ensemble learning methods broad in scope.

The next chapter presents a stacked ensemble learning approach. Beginning, the data set used in the study was explained, how to collect and obtain it, its features, and the most important processing that was introduced to it to become the form in which it was used. Then, a review of the methods of work and the algorithms used: their names, the idea of work, and the parameters that the researcher used in each algorithm.

3. MATERIALS AND METHODS

The proposed method includes four different models that were developed, each model contained algorithms for a type of machine learning (Bagging, boosting, and stacking), in addition to a model that had a variety of algorithms (single and ensemble learning). These four models used four methods of selecting features: Relief F, ANOVA, information gain, and chi-square. It combines the output of various classification algorithms to make a more accurate prediction. 11 classifiers were chosen and applied. Among these classifiers, five are single and the rest are ensemble. Each of the classifiers was tuned to get the best combination of parameters for optimizing the result. The dataset was split into training (70%) for model building and testing (30%) for model verification subsets. The researcher used the K-fold cross-validation resampling method (k = 10) in this study, to tune the model parameters. The researcher's focus was to investigate the effectiveness of the three most popular ensemble learning methods (bagging, boosting, and stacking) to improve classification accuracy and model performance.

Figure 3.1 illustrates the structure of the approach including dataset description, stacked ensemble learning models, and evaluation metrics.

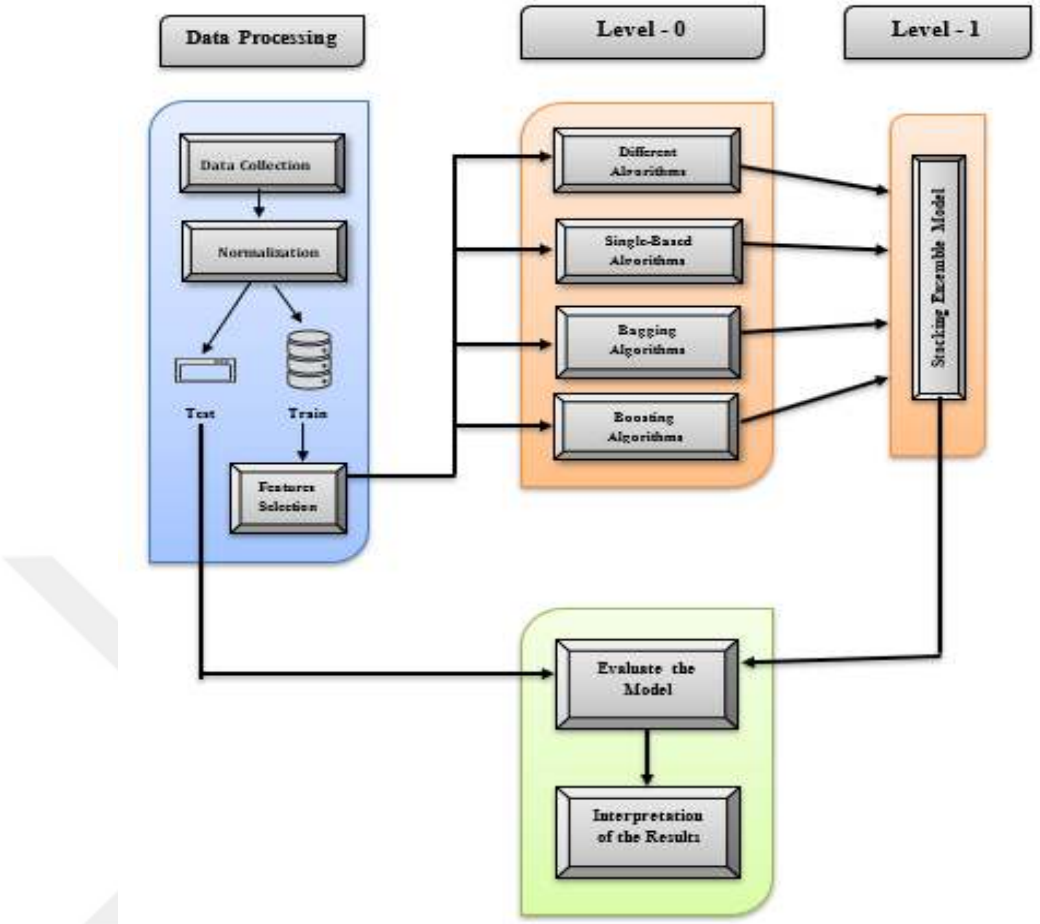


Figure 3.1 Structure of the proposed method

3.1 Data Collection

In recent years, the quick development of machine learning for data processing has been one of the greatest technological advances. The data set used in the study is one of several aspects that are important to this development. The raw data range of values frequently has several different scales. In this situation, without data normalization, several machine learning algorithms' objective functions won't operate well.

The data used in this study is called (Students Adaptability Level in Online Education), from Bangladesh, from December 10, 2020, to February 5, 2021. It contained 14 attributes and 1205 instances, arranged in the description Table 3.1

below. It contains 663 boys and 542 girls engaged in online education for the specific sample. Non-Government institutions were 823 whilst Government institutions were just 382. The students divided into poor 242 and Mid 878 are the financial conditions counts for non-government entities, whereas Poor 62 and Mid 288 are the financial conditions counts for governments.

Table 3.1 Description of the data set

ID	FEATURE NAME	CLASS	VALUES	LABEL
1	Gender	Integer	Boy (1) Girl (0)	Gender of student
2	Age	Numeric	1-5 (0) 6-10 (1) 11-15 (2) 16-20 (3) 21-25 (4) 26-30 (5)	Age of student
3	Education level	Integer	School (0) College (1) University (2)	Education level for student
4	Institution type	Integer	Government (0) Non-government (1)	The type of institution in which the student studies
5	IT student	Integer	Yes (1) No (0)	Is the student studying information technology?
6	Location	Integer	Yes (1) No (0)	
7	Load-shedding	Integer	Law (1) High (0)	
8	Financial condition	Integer	Rich (2) Poor (1) Mid (0)	Financial condition for student's family
9	Internet Type	Integer	Wifi (1) Mobil data (0)	
10	Network Type	Integer	2G (1) 3G (2) 4G (3)	

3.2 Data Pre-processing

In the data mining process, one of the most important steps is data pre-processing and data normalization (scaling). Some machine learning algorithms and methods do

not achieve satisfactory results and do perform not well on raw data. The data set used in this study is complete and does not contain missing data, so the first step of implementation converted the fields from string to numeric in order to facilitate the algorithms' handling of numerical results and to improve the results. Then, the dataset was divided into training and testing parts (training 70% and testing 30%).

3.3 Features Selection

The feature selection effectively prepares data for a variety of data-mining and machine-learning issues. Building easier-to-understand models, enhancing data mining efficiency, and creating clean, intelligible data are all goals of feature selection (Khaire and Dhanalakshmi 2019).

In this study, the researcher has applied four feature selection techniques, it is Relief F, ANOVA (The one-way analysis of variance), Information Gain, and Chi-square.

- Relief F: It is the ability of an attribute to distinguish between classes on similar data instances. Relief methods may seem simple at first appearance, but it is not always clear how to interpret feature weights or how feature dependencies can be inferred without explicitly taking feature subsets into account (Katarya 2019). Relief's core principle is that feature relevance may be estimated by examining how effectively feature values can separate concept (endpoint) values from instances that are similar to close to each other (Urbanowicz *et al.* 2018). In this study, the number of features use with this kind was 10 features.
- ANOVA: The one-way analysis of variance (ANOVA) tests is mostly used to determine whether or not all classes of Y have the same mean as X (Khaire and Dhanalakshmi 2019). Also, which is known as a robust and simple method to test the difference in means between groups, even when the number of observations is uneven in each group. Besides, it is easily generalized to more than two groups without increasing the Type 1 error. 8 features were the number of features used with this feature selection type.

- Information Gain: also referred to as Mutual Information Maximization (MIM), uses a feature's correlation with class labels to gauge its significance. The presumption is that a feature can aid in achieving superior classification performance if there is a substantial association between it and the class label (Li *et al.* 2017). Importantly, each feature's MIM score is assessed individually (Khaire and Dhanalakshmi, 2019). Consequently, only feature correlation is considered, while feature redundancy is entirely disregarded. Once the MIM scores for each feature are determined, we select the features with the highest scores and incorporate them into the chosen feature set. This process is repeated until the desired selection of features is reached. This technique is widely used in the field of machine learning as a feature selector (Krishnan *et al.* 2022). The Information Gain (IG) of a term is calculated by determining the amount of information gained about a predicted category from the presence or absence of the term in a text (Rizaldy and Santoso 2017).

- Chi-Square: The best terms tk for the class ci are those that are distributed most difference in the sets of positive and negative instances of class ci. This concept is captured by the Chi-square statistics formula in Equation (3.1), which is connected to information theoretic feature selection functions.

$$Chi - square = \frac{N(AD-CB)^2}{(A+C)(B+D)(A+B)(C+D)} \quad (3.1)$$

Where N = All number of documents in the corpus, A = Number of documents, B = Number of documents containing the term (tk) in other classes; C = Number of documents and D = Number of documents that do not contain the term in other classes (Bahassine *et al.* 2020).

3.4 Machine Learning Algorithms

In this study, a variety of supervised learning algorithms were used, where the models were built somewhat differently in order to know the effect of each group of algorithms on the data and their results in the process of classifying students.

3.4.1 Single based algorithm

1. Support Vector Machine (SVM): It is one of the most important Supervised Learning algorithms, mainly used in classification problems, and also used in regression (Lau *et al*. 2019). Its mechanism of action is based on finding the best separating hyperplane between classes (Manu 2016). It is used here to classify data into a set of classes (Tarik *et al.* 2021). SVM models are related very closely to classical multilayer perceptron neural networks (Manu 2016). It has been shown that an upper bound on the expected generalization error can be reduced by maximizing the margin and so establishing the greatest distance between the separating hyperplane and the examples on either side of it (Osisanwo *et al.* 2017). In the creation of SVM model, the parameter value for C was 1, and the 'rbf' kernel function was used.

2. Logistic Regression: This classification method employs a single multinomial logistic regression model with a single estimator and builds its model using the class. In a particular way, logistic regression often identifies the location of the boundary between the classes and notes that class probabilities vary with proximity to the boundary (Ketui *et al*. 2019). When the data set is larger, this advances more quickly towards the extremes (0 and 1). This method is more complex than a simple classifier. It is adaptable and can produce predictions that are stronger and more precise, but those exact projections could be inaccurate. Logistic regression is one of the most popular tools for discrete data analysis and applied statistics (Osisanwo *et al.* 2017). The parameter value of C = 10 in the creation of this algorithm.

3. Artificial Neural Network (ANN): It is a parallel operation structure suitable for processing a large amount of data, to which it is easy to add or change information by adjusting the connection strength through learning and training (Krishnan *et al*. 2022). It was developed from biological neural networks found in the human body. The three neural network layers are the input, middle (also known as hidden), and output layers. A number of neurons make up each layer. When a critical point is crossed by the activation function, the neurons in each layer double the weighting factor applied to the input value before outputting the result (Lagman *et al*. 2019). Through the iterative backpropagation process, these weighting factors are modified with an eye on decreasing mistakes (Li *et al.* 2017). In this model, the researcher uses neurons in hidden layers 500 (Figure 3.2),

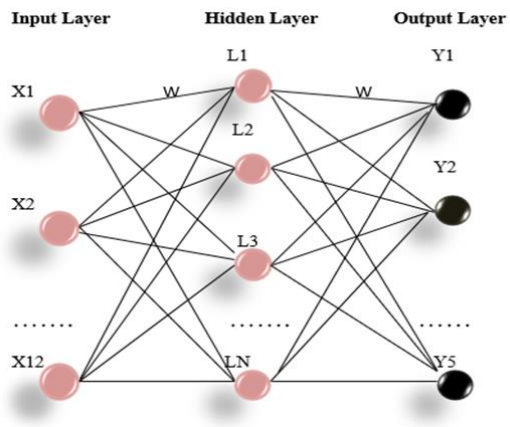


Figure 3.2 Structure of ANN algorithm

4. K-Nearest Neighbor (KNN): It is one of the simplest machine learning algorithms, and it falls under the category of supervised learning. The idea of this algorithm is that it depends on the classification of the test point based on the points near and surrounding it from the test points, hence the name (Kumbhar and Mali 2016). Nearest neighbors mean the data points that have a minimum distance in the feature space from the new data point. And 'k' is the number of data points that we have to take into account when we apply the algorithm. Therefore, the magnitude of 'k' and the measure of distance are two important

considerations that must be paid attention to while using the KNN algorithm (Li *et al*. 2017). The parameter of neighbours was 5, and also Euclidean metrics were used in the creation of the KNN.

3.4.2 Bagging algorithm

The ensemble approach includes repeatedly training the same algorithm with various subsets of the training data (Martin *et al*. 2021). The forecasts of all the sub-models are then averaged to get the final output prediction. By reducing the variation of classification mistakes, bagging often increases classification accuracy (Pal and Pal 2013). This part has explained the outline of the mechanism of action for the bagging algorithms that were used in this study.

1. Decision Trees: This algorithm contains a set of "Decision Rules", through which the class (test result) is found. Its working mechanism is the division into nodes from top to bottom, and the data set is divided into small subgroups until the target node (class) is reached. The top node is called the "Root Node", the child node is called the "Decision Node", and the nodes that do not split are called the "Leaf Node", which represents the end of this branch of the tree (Mulyati and Setiani 2018). It often uses post-pruning approaches to assess the effectiveness of decision trees as they are being pruned with the use of a validation set. Any node may be deleted and have the most prevalent class of the sorted training instances allocated to it (Binkhonain and Zhao 2019). The number of instances in leaves was 2 and the maximum depth of trees = 100.

2. Random Forest (RF): It is an algorithm belonging to the supervised learning technique. It includes several decision trees on different subsets of the dataset that is being worked on and takes the average instead of relying on a single decision tree. The random forest takes a prediction from each tree and the output depends on the majority of votes for the predictions. It is used for both classification and regression problems in ML. The basis of this work is the concept of group learning which means combining several classifiers to solve a

complex problem and improve the performance of the model (González *et al.* 2020). Among the set of hyper-parameters, that can be tuned for Random Forest, in this study the researcher used: number of trees=20 for initial and then it changed depending on the model.

3.4.3 Boosting algorithm

The sequential ensemble technique known as "boosting" is used to transform weak learner algorithms into strong learner algorithms. The following is an explanation of how the Boosting algorithms work, that are used in this study work.

1. Gradient boosting: The iterations of the gradient boosting machine, a stage-wise additive model, can be thought of as the steepest descent minimization with respect to a particular loss function. Through numerical optimization in the function space, the predictive function is calculated. GBM is made to employ either a decision tree or a linear regression as its base learner. GBM continues to fit trees until the maximum number of estimators is reached. The initial guess and the total of all scaled outputs from the ensemble's trees forecast the result for fresh measurements x. The attributes used in creation of the models were: number of trees: 100.

2. EXtreme Gradient Boosting (XGBoosting): The XGBoost algorithm is a technique for enhancement. The improvement algorithm's core tenet is that numerous experts evaluate complex tasks on an individual basis before performing an appropriate synthesis to obtain a decision. The conclusion reached by synthesis is superior to any one expert's opinion alone. The regression tree model is the foundation of the XGBoost algorithm (Yuvalı *et al*. 2022). The main idea is to build hundreds of regression tree models by continually extracting various variables, then linearly combining those models to create the final model. The following hyper-parameters tuned for XGBoost in this study: number of trees=100 and limit depth of individual trees=6.

3. Gradient Boosting with Categorical features (CatBoost): is the newest GBM-based algorithm that is relevant according to the literature. The new ordered boosting technique that avoids prediction shift and support for categorical features are CatBoost's two key characteristics. CatBoost offers a variety of solutions that enable categorical features (Pallathadka *et al*. 2021). These processes are designed to be used during tree splitting, not during preprocessing. CatBoost uses one-hot encoding to combine features with a limited number of categories. With regard to the quantity of appearances of the categories, it can also convert categorical features to numerical values (Zhang *et al*. 2019). In this study, the hyperparameter used for this algorithm were: number of trees=50 and learning rate=0.200.

3.4.4 Stacked ensemble

The stacked ensemble consists of two phases, a layer with learners/estimators (level-0) and meta learner (level-1). The output of level-0 is passed as input to the next layer level-1 (Qiu *et al*. 2022). The architecture of the stacked model is shown in Figure 3.3.

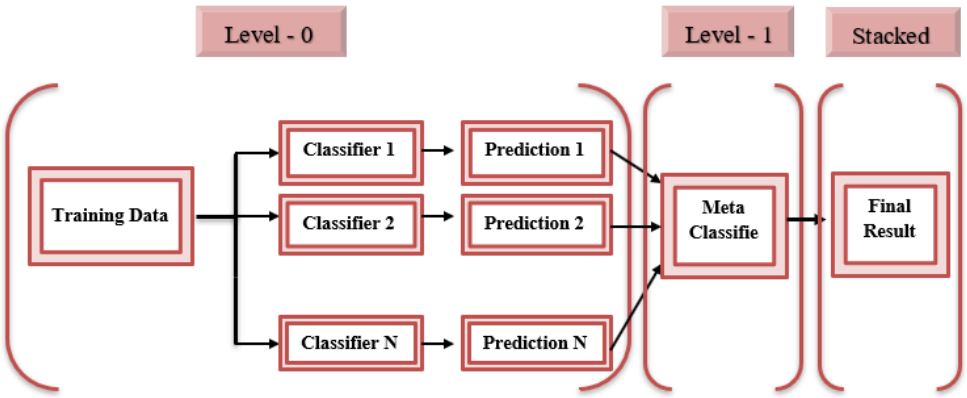


Figure 3.3 Architecture of stacked ensemble

In this study, the stack model was created according to the type of algorithms used to create the base model, where four models were developed at level-0. The first model contains a variety of ensemble learning algorithms in the base model, the second

model consists of several single based algorithms, the third model contains two bagging algorithms, and finally, the fourth and final model contains a group of boosting algorithms. All these models were employed by taking all the features, and then four methods were used to select the features and obtain the results in order to compare them.

3.5 Evaluation Matrices

Classification metrics are evaluated from false positives (FPs), true positives (TPs), false negatives (FNs), and true negatives (TNs).

TP (True Positive): When a positive event is anticipated and what really occurs matches the prediction, the data point in the confusion matrix is True Positive (TP).

FP (False Positive): When a good outcome is anticipated but a negative outcome actually occurs, the data point in the confusion matrix is said to be false positive. A Type 1 Error is what this is called.

FN (False Negative): When a negative event is anticipated but a good outcome actually occurs, the data point in the confusion matrix is considered false negative. This situation is a common example of a Type 2 Error and is just as risky (Sekeroglu *et al*. 2019).

TN (True Negative): When a negative event is anticipated and it really occurs, the data point True Negative (TN) in the confusion matrix is present (Nair *et al*. 2022).

3.5.1 Accuracy

The capacity of a predictor to properly identify all samples, whether positive or negative, is measured by accuracy (Jiao and Du 2016). It is perhaps the most popular and initial option for assessing an algorithm's effectiveness in classification tasks. It

may be characterized as the proportion of correctly identified data items to all observations (Ng *et al*. 2021). In some scenarios, especially when target variable classes in the dataset are unbalanced, accuracy is not the best performance statistic.

$$Accuracy = \frac{TP+TN}{(TP+TN+FP+FN)} \tag{3.2}$$

3.5.2 Sensitivity (recall)

Recall or the true positive rate (TPR) are other terms for sensitivity (Jiao and Du 2016).

$$Recall = \frac{TP}{(TP+FN)} \tag{3.3}$$

3.5.3 Precision

It simply means how many of the observations that an algorithm anticipated to be positive are really positive.

$$Precision = \frac{Tp}{(TP+FP)} \tag{3.4}$$

3.5.4 F-score (F1)

Another summary statistic in machine learning is the F1 score, which is calculated as the harmonic mean of the sensitivity (recall) and PPV (precision).

$$Fscore = 2 * \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{3.5}$$

3.5.5 ROC curve

The receiver operating characteristic (ROC) curve is one of several statistical estimators that are used to assess the effectiveness of classificatory models. It shows how a quantitative data model performs by displaying its sensitivity (percentage of true positives) against the proportion of false positives (1-specificity) for various test values (Nawai *et al*. 2021). The test is considered strong in terms of its ability to differentiate between groups the closer the ROC curve gets to the upper left corner of the graph (Pintelas and Livieris 2020). Additionally, the ROC graph's diagonal reference line represents a completely random area where a test is unable to distinguish between healthy and sick people (sensitivity = specificity). The applied reference value affects the sensitivity and specificity. The specificity increases while the sensitivity drops when the reference value is raised (Xiao *et al*. 2022). Using their ROC curves, it is also possible to simultaneously compare the performance of two or more classificatory models.

3.5.6 Confusion matrix

It is a Table design that makes it possible to see how well a supervised learning algorithm is performing (Muñoz-Carpio *et al*. 2021). Also, it is considered a tool to measure the performance of classifiers in their ability to categorize items with several classes. Calculating classification accuracy for 2-classed attributes using a confusion matrix is relatively simple, but for multi-class attributes, it is quite laborious (Patro and Patra 2015). With the aid of its four components—True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN), confusion Matrix serves as the foundation for assessing the performance of any classifier (Sudrajat *et al*. 2017).

The data used and the details related to it were clarified from the number of features and the number of students included in this study. After that the normalization that was used for the data set was explained, and the methods for choosing features were explained in order to improve the results. Then the researcher explained the methods

of machine learning and the mechanism of work of the algorithms used. Then determine the initial parameters adopted by the algorithms when they were created. Finally, the researcher clarified the methods of evaluating the models used and calculating them.

4. RESULTS AND DISCUSSION

The machine learning algorithms used were explained in the previous chapter. They used some tools to find out how well they performed their work. These are named performance evaluation metrics (Narassiguin 2018). Many metrics have been introduced in this study, and each one considers certain aspects of an algorithm's performance. Orange Data Mining Tool has been used to classify various machine learning algorithms in this study. It is helpful programming and data analysis tool. It is supported by python and many other programming languages. It is also easy to learn.

The researcher uses several metrics for classification problems to obtain all information about the performance of algorithms and can make a comparison between the types of models that have been developed. These metrics are f1-score, precision, recall, accuracy, confusion matrix, and ROC-AUC score. The results of these tools will be presented in tabular order in this chapter. The results are arranged in Tables according to the types of algorithms in the development of the model. Before commencing, the important features selected using feature selection methods will be indicated.

4.1 Selected Features

The feature selection technique's goal is to reduce the number of features while maintaining high performance by discarding those least useful (Vikström 2021). This can be done by looking at each feature's usefulness and selecting some arbitrary amount from the best features (Shardlow 2016). The number of features selected for each type of feature selection technique differed in this study. The technologies used and the features selected for each type are listed below.

4.1.1 Relief F

This technique was chosen due to it selects the best attributes based on contribution between instances (Xiao *et al*. 2022). The number of features of this technique in this study was 10 features. Table 4.1 below contains selected features. The reason for choosing these features and this number is because they gave the best results for this method. As the researcher tried different numbers of features and this was the best.

Table 4.1 Selected features for Relief F

FEATURE NO.	FEATURE NAME
1	Class Duration
2	Network Type
3	Age
4	Internet Type
5	Financial Condition
6	Education Level
7	Gender
8	IT Student
9	Institution Type
10	Device

4.1.2 ANOVA

It is a statistical evaluation carried out on a reference dataset to learn about feature attributes (Shakeela *et al.* 2021). It calculated the ratio of the variance within groups for each attribute and the variance among groups. 8 features were selected, shown in Table 4.2 These features give the best scores compare with another number of features.

Table 4.2 Selected features for ANOVA

FEATURE NO.	FEATURE NAME
1	Gender
2	Age
3	Education Level
4	Institution Type
5	IT Student
6	Location
7	Load-shedding
8	Financial Condition

4.1.3 Chi-square

To determine whether one variable's distribution varies from another's, this was used this technique. Specifically, it is used to test whether the occurrence of a specific term and the occurrence of a specific class are independent (Kumbhar and Mali 2016). 10 features were included in this study within this method as shows in Table 4.3.

Table 4.3 Selected features for Chi- square

FEATURE NO.	FEATURE NAME
1	Financial Condition
2	Class Duration
3	Institution Type
4	Self LMS
5	Location
6	IT Student
7	Internet Type
8	Educational Level
9	Age
10	Gender

4.1.4 Information gain

Reducing the uncertainty linked with identifying the class attribute while the value of the feature is unknown, is the motivation to use this technique (Zhang *et al*. 2021). It measures the amount of information by bits, about the class prediction, if all that is

known about a feature is if it exists and what classes it belongs to. The number of features selected was 7. Selected features are explained in Table 4.4.

Table 4.4 Selected features for information gain

FEATURE NO.	FEATURE NAME
1	Class Duration
2	Financial Condition
3	Age
4	Institution Type
5	Location
6	Device
7	Educational Level

4.2 Results for Model 1 Set of Single and Ensemble Learning

Model 1 was developed using various classifiers. It contains single based algorithms including SVM and KNN and ensemble algorithms including DT, RF, and XGBoost. Table 4.5 explains all the results for this model with all features. The higher accuracy was for the KNN algorithm with a value = 0.821. The value of AUC for this algorithm was 0.935. Also, the F1, Precision, and Recall scores for the same algorithm were 0.819, 0.819, and 0.821 sequentially.

Table 4.5 Results of model 1. with all features

MODEL1-ALL FEATURES	AUC	K=10 F	F1	PRECISION	RECALL
KNN	0.935	0.821	0.819	0.819	0.821
DT	0.916	0.805	0.805	0.805	0.805
SVM	0.869	0.748	0.748	0.748	0.748
RF	0.806	0.676	0.636	0.643	0.676
XGBoost	0.835	0.702	0.687	0.716	0.702

Table 4.6 contains the results for model 1 by Relief F technique with 10 features. KNN algorithm has the best accuracy score with a value = 0.864. RF algorithm has the second best result for accuracy with a value = 0.851. The best scores for F1, Precision, and Recall also for KNN with scores sequentially = 0.862, 0.868, and 0.864.

Table 4.6 Results of model 1 by Relief F technique with a choice of 10 features

MODEL1-RELIEF F/10 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
KNN	0.869	0.864	0.862	0.868	0.864
DT	0.950	0.841	0.841	0.842	0.841
SVM	0.899	0.821	0.818	0.822	0.821
RF	0.952	0.851	0.850	0.851	0.851
XGBoost	0.927	0.827	0.825	0.831	0.827

For the results of model 1 by ANOVA technique with 8 features, RF has the best score for accuracy with value = 0.763, While the lowest score is for SVM algorithm with value = 0.641. Table 4.7 showed all results for every algorithm. These were the highest values obtained after changing the number of selected features many times.

Table 4.7 Results of model 1. by ANOVA technique with a choice of 8 features

MODEL1- ANOVA/8 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
KNN	0.846	0.715	0.713	0.712	0.715
DT	0.888	0.754	0.751	0.754	0.754
SVM	0.728	0.641	0.641	0.650	0.641
RF	0.896	0.763	0.760	0.762	0.763
XGBoost	0.883	0.746	0.741	0.751	0.746

About Chi-square technique with 10 features for model 1, also KNN has the best score with a value = 0.863, a small difference from the XGBoost algorithm whose accuracy = 0.862. The high score for F1, Precision, and Recall was: 0.713, 0.712, and 0.715 for KNN. The results for all algorithms of this model were listed in Table 4.8.

Table 4.8 Results of model 1. by chi-square with a choice of 10 features

MODEL1-CHI-SQUARE/10 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
KNN	0.872	0.863	0.862	0.864	0.863
DT	0.951	0.844	0.843	0.843	0.844
SVM	0.893	0.824	0.822	0.826	0.824
RF	0.951	0.843	0.842	0.842	0.842
XGBoost	0.961	0.862	0.861	0.861	0.861

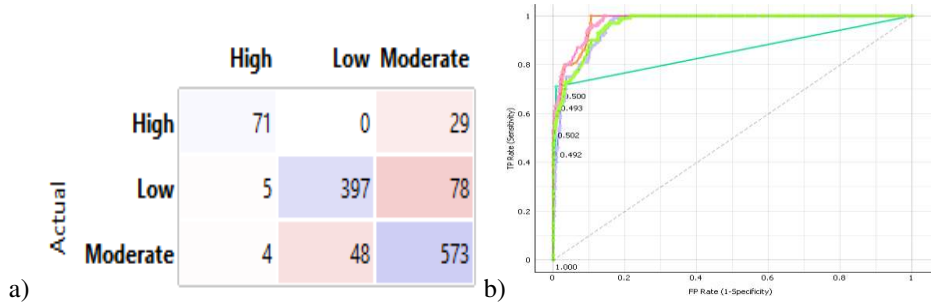
The last technique used was Information Gain, which takes 7 features. Their results were very similar, as the highest accuracy value = 0.759 for RF, while the single algorithms were equal to an accuracy = 0.758 for each SVM and KNN. The results of all algorithms are listed in Table 4.9.

Table 4.9 Results of model 1. by information gain with a choice of 7 features

MODEL1-INFORMATION GAIN/7 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
KNN	0.878	0.758	0.757	0.759	0.758
DT	0.886	0.747	0.747	0.746	0.747
SVM	0.818	0.758	0.b754	0.755	0.758
RF	0.888	0.759	0.758	0.757	0.759
XGBoost	0.886	0.755	0.755	0.754	0.754

The results of the first model show the superiority of algorithm KNN in most of the developed models. It achieved the highest accuracy by Relief F technique with a slight difference from the Chi-square technique (0.864 in Relief and 0.863 in Chi-square) over the rest of the types of feature selection techniques. The SVM algorithm got the least accuracy with the ANOVA technique which is = 0.641.

Figure 4.1 shows the confusion matrix to Relief F (10 features) due to the superiority of its results compared with other techniques in this model, and the ROC curve for KNN algorithm that gets the best score = 0.502.



a) b)

Figure 4.1 a) Confusion matrix for Relief F and b) ROC curve for KNN algorithm

4.3 Results of Model 2 Single Based Algorithms of the Student Adaptation

Model 2. Consist of 3 single based algorithms which are; KNN, SVM, and Neural Network. The four feature selection techniques were used when the model was developed. When all features were taken, the highest accuracy was for the KNN algorithm, which was = 0.795. The least accurate was 0.641 for SVM algorithm. Also, the value of F1, Precision, and Recall for KNN was good, which is higher than the values of the other algorithms used. The values, in sequence, are as follows: 0.819, 0.795, and 0.795. AUC = 0.910 for KNN algorithm. Table 4.10 contains the results of all feature selection techniques used in this model.

Table 4.10 Results of model 2. with All features

MODEL2/SIGNLE BASED ALL FEATURES	AUC	K=10F	F1	PRECISION	RECALL
KNN	0.910	0.795	0.819	0.795	0.795
SVM	0.781	0.640	0.636	0.662	0.641
Neural Network	0.802	0.703	0.700	0.709	0.703

Table 4.11 contains the results of using the Relief F technique in 10 features of Model 2. The highest accuracy was for the KNN algorithm which achieved 0.859. The least accurate was the algorithm with a value of 0.758. The scores of F1, Precision and Recall were equal to the value of the accuracy of KNN algorithm. The AUC of this algorithm was 0.955.

Table 4.11 Results of model 2. by Relief F technique with a choice of 10 features

MODEL2/SIGNLE BASED RELIEF-F/10 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
KNN	0.955	0.859	0.858	0.859	0.859
SVM	0.893	0.823	0.821	0.822	0.823
Neural Network	0.856	0.758	0.752	0.761	0.758

With an accuracy reaching 0.859, also KNN algorithm outperformed others when using the ANOVA technique with 8 features. Table 4.12 contains the results of all algorithms.

Table 4.12 Results of model 2. by ANOVA technique with a choice of 7 features

MODEL2/SIGNLE BASED ANOVA/8 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
KNN	0.845	0.710	0.705	0.701	0.710
SVM	0.725	0.641	0.641	0.650	0.641
Neural Network	0.809	0.690	0.688	0.696	0.698

SVM algorithm has the highest score for model 2 using a chi-square with 10 features. It's achieved 0.818 accuracy. KNN did not excel with this technique, but came after SVM by achieving accuracy = 0.788. KNN achieves the highest value in AUC with a score of 0.904. All the results of the algorithms are explained in Table 4.13.

Table 4.13 Results of model 2. by chi-square technique with a choice of 10 features

MODEL2/SIGNLE BASED CHI-SQUARE/10 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
KNN	0.904	0.788	0.783	0.797	0.788
SVM	0.892	0.818	0.816	0.819	0.818
Neural Network	0.792	0.700	0.692	0.701	0.700

When using the Information Gain technique, the performance of both KNN and SVM algorithms is very similar. KNN achieved an accuracy of 0.759 a slight difference from SVM, whose accuracy was 0.758. Other results are showed in Table 4.14.

Table 4.14 Results of model 2. by information gain technique with a choice of 8 features

MODEL2/SIGNLE BASED -INFORMATION GAIN/8 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
KNN	0.876	0.759	0.758	0.7760	0.759
SVM	0.819	0.758	0.754	0.755	0.758
Neural Network	0.855	0.733	0.728	0.730	0.733

For which all the algorithms are single based algorithms. Figure 4.2 shows the confusion matrix for Relief F getting high a score and for SVM algorithm that has the best score in Roc curve 0.540.

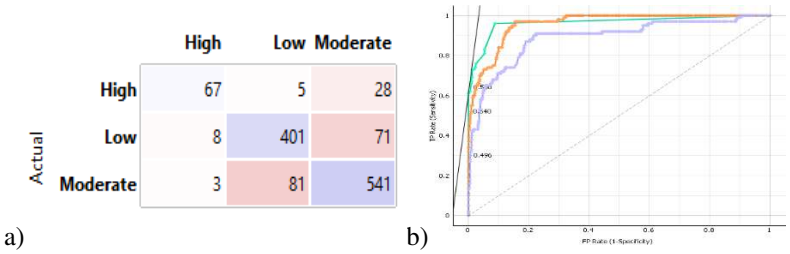


Figure 4.2 a) Confusion matrix for Relief F and b) ROC curve for SVM algorithm

4.4 Results of Model 3 Bagging Algorithms of the Student Adaptation

Two famous algorithms were employed from bagging in model 3, it is DT and RF. Table 4.15 includes results of this model by using all features. The RF algorithm is more accurate than DT, as it reached 0.716. Also, the high scores of each F1, Precision, and Recall for RF were as follows: 0.708, 0.723, and 0.716.

Table 4.15 Results of model 3 with all features

MODEL3/BAGGING ALL FEATURES	AUC	K=10F	F1	PRECISION	RECALL
DT	0.769	0.694	0.675	0.693	0.687
RF	0.838	0.716	0.708	0.723	0.716

Table 4.16 contains the results of model 3 when use Relief F technique with 10 features. RF Algorithm performed better than DT, as it achieved an accuracy of 0.852. While DT accuracy was 0.838.

Table 4.16 Results of model 3 by Relief F with a choice of 10 features

MODEL3/BAGGING RELIEF F10 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
DT	0.947	0.838	0.838	0.838	0.838
RF	0.954	0.852	0.851	0.852	0.852

The lowest results in this study are using ANOVA technique with 8 features in Model 3. RF algorithm achieved 0.616 accuracy and RF achieved only 0.583 accuracy. Table 4.17 showed all results for RF and DT algorithms.

Table 4.17 Results of model 3 by ANOVA with a choice of 8 features

MODEL3/BAGGING ANOVA 8 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
DT	0.612	0.583	0.536	0.602	0.583
RF	0.697	0.616	0.595	0.616	0.616

Chi-square technique with 10 features achieved good scores in RF algorithm. The accuracy score was 0.866. AUC for this algorithm was 0.963 and the scores for F1, Precision, and Recall was 0.874, 0.876, and 0.874. Table 4.18 contains all the results.

Table 4.18 Results of model 3 by chi-square with a choice of 10 features

MODEL3/BAGGING CHI SQUARE/10 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
DT	0.956	0.857	0.857	0.857	0.857
RF	0.963	0.866	0.874	0.876	0.874

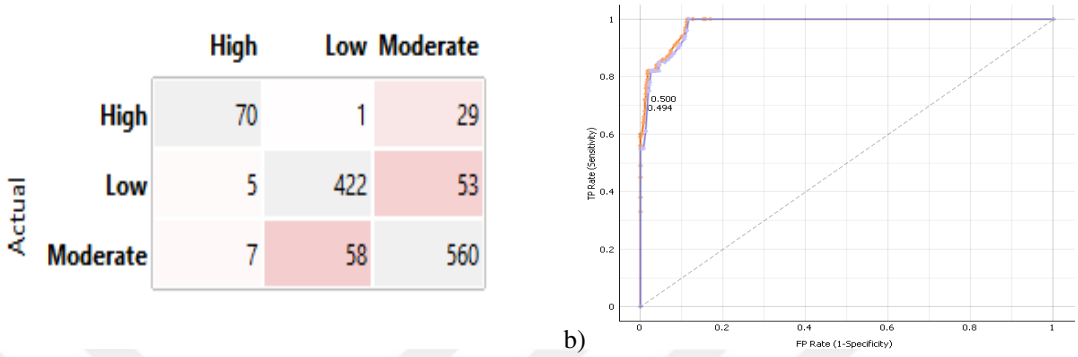
Last technique used in this model is Information Gain with 7 features. The best score was 0.759 accuracy for RF algorithm with a slight difference from the DT algorithm which achieved 0.755 accuracy. All results for these algorithms are in Table 4.19.

Table 4.19 Results of model 3 by information gain with a choice of 8 features

MODEL3/BAGGING INFORMATION GAIN/7 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
DT	0.885	0.755	0.755	0.755	0.755
RF	0.887	0.759	0.757	0.758	0.759

After reviewing all the results of model 3 and all the used feature selection techniques, it is noted that the highest accuracy is for RF algorithm using the Chi-square technique, as its accuracy reached 0.868. The higher score for AUC was 0.963 by RF. The higher scores for each of F1, Precision, and Recall were equal to 0.857.

Figure 4.3 contains a confusion matrix for the high score of this model which is achieved by chi-square features selection. Also, shows the Roc curve where DT has a 0.500 score, which is the highest score in this model.



a) b)

Figure 4.3 a) Confusion matrix for chi-square and b) ROC curve for random forest

4.5 Results of Model 4 Boosting Algorithms of the Student Adaptation

Two boosting algorithms have been used to develop this model. The boosting algorithms used in this model are: Gradient boosting and Catboost. Table 4.20 contains the results of these models by using all features. 0.851 was the best accuracy score achieved by Gradient Boosting and also, and it achieved a high score for AUC reaching 0.953. The rest of the results were shown in the same Table.

Table 4.20 Results of model 4 with all features

MODEL4/BOOSTING ALL FEATURES	AUC	K=10F	F1	PRECISION	RECALL
Gradient boosting	0.953	0.851	0.848	0.859	0.851
Catboost	0.934	0.815	0.813	0.815	0.815

With an accuracy reaching 0.853, Gradient Boosting outperformed to Catboost which achieved 0.823 accuracies in model 4 by using Relief F with 10 Features. The scores of F1, Precision, and Recall were: 0.851, 0.857, and 0.856 for Gradient Boosting. Table 4.21 shows the results of model 4 by Relief F with a choice of 10 features. The best accuracy was 0.771 for Gradient Boosting, while Catboost achieved 0.741 accuracies.

Table 4.21 Results of model 4 by Relief F with a choice of 10 features

MODEL4/BOOSTING ALL Relief F FEATURES	AUC	K=10F	F1	PRECISION	RECALL
Gradient boosting	0.905	0.771	0.767	0.772	0.771
Catboost	0.872	0.741	0.738	0.738	0.741

Table 4.22 contains the results of model 4 by ANOVA with 8 features. The best accuracy was 0.856 for Gradient Boosting, while Catboost achieved 0.823 accuracies.

Table 4.22 Results of model 4 by ANOVA with a choice 8 features

MODEL4/BOOSTING ANOVA /8 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
Gradient boosting	0.955	0.853	0.851	0.857	0.856
Catboost	0.922	0.823	0.821	0.825	0.823

The highest score in this model was achieved by Chi-square with 10 features when using Gradient Boosting which was 0.866 accuracy. The accuracy for Catboost was 0.813. AUC for Gradient Boosting was 0.960. All results for this model are shown in Table 4.23.

Table 4.23 Results of model 4 by chi-square with 10 features

MODEL4/BOOSTING CHI-SQUARE/10 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
Gradient boosting	0.960	0.866	0.864	0.868	0.866
Catboost	0.926	0.813	0.811	0.813	0.813

Last technique used in this model is Information Gain with 7 features. The best score was 0.682 accuracy for Gradient Boosting algorithm. The Catboost algorithm was achieved 0.665 accuracy. All results for these algorithms are in Table 4.24.

Table 4.24 Results of model 4 by information gain with 7 features

MODEL4/BOOSTING INFORMATION GAIN/7 FEATURES	AUC	K=10F	F1	PRECISION	RECALL
Gradient boosting	0.805	0.682	0.671	0.686	0.682
Catboost	0.768	0.665	0.633	0.706	0.665

Among the four feature selection techniques used with this model, Chi-square technique excelled with Gradient Boosting algorithm as it gave an accuracy of 0.866. The least accuracy of this model was achieved by Catboost algorithm with the information Gain technique, where its accuracy was 0.665.

Figure 4.4 shows confusion matrix for chi-square technique which has the best score and Gradient Boosting algorithm which has a high score in Roc curve which is 0.515.

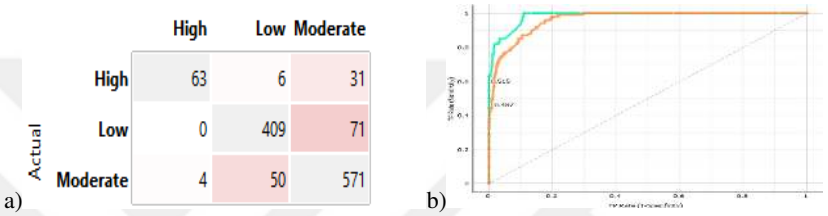


Figure 4.4 a) Confusion matrix for Chi-square and b) ROC curve for gradient boosting

4.6 Results of Stacked Ensemble Algorithms of the Student Adaptation

The results of the staked ensemble were arranged in Tables 4.26, 4.27, 4.28, and 29 as a type of algorithms in models. At the bottom of each Table, there is a confusion matrix and a Rock curve. Algorithms that were used to create a based model (level-0) were: SVM, KNN, RF, DT, and XGboost. Catboost was the Meta classifier (level-1) for model 1 with all features selection techniques.

Table 4.25 shows the results of stacked ensemble learning for model 1. Relief F with 10 features gets a high score for accuracy that = 0.870, and a high score for Roc Curve equaled 0.508. Also, chi-square feature selection achieved the best score for the AUC with value = 0.963.

Table 4.25 Results of stacked ensemble learning for model 1

TECHNIQUE AND NO. OF FEATURES	LEVEL 0	LEVEL 1	AUC	K=10F	F1	PRECISION	RECALL
All Features	SVM, KNN, RF, DT, XGBoost	Catboost Classifier	0.941	0.832	0.829	0.828	0.829
Relief F-10 Features	SVM, KNN, RF, DT, XGBoost	Catboost Classifier	0.958	0.870	0.869	0.871	0.870
Anova-8 Features	SVM, KNN, RF, DT, XGBoost	Catboost Classifier	0.888	0.758	0.767	0.771	0.769
Chi-square 10 Features	SVM, KNN, RF, DT, XGBoost	Catboost Classifier	0.963	0.865	0.864	0.865	0.865
Information Gain-7 Features	SVM, KNN, RF, DT, XGBoost	Catboost Classifier	0.891	0.765	0.765	0.765	0.765

Figure 4.5 shows the performance metrics of the stacked ensemble model, Model 1, as displayed through a confusion matrix and Receiver Operating Characteristic (ROC) curve.

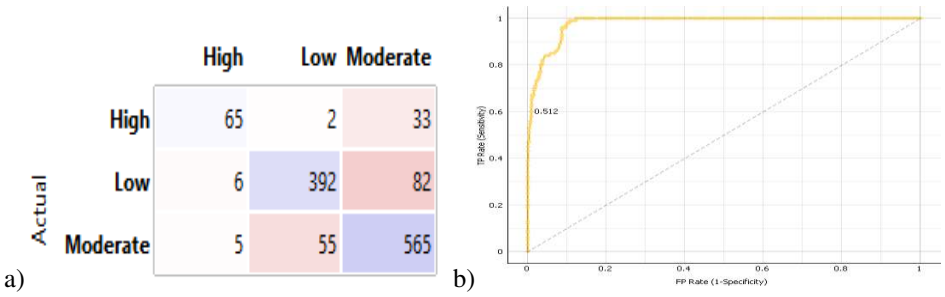


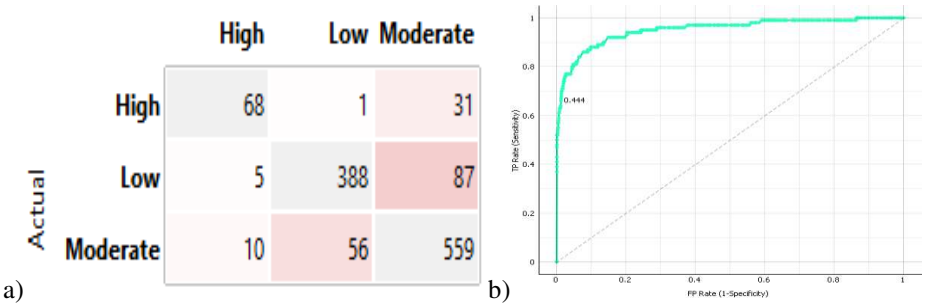
Figure 4.5 a) Confusion matrix and b) ROC curve for stack of model 1

In Table 4.26 the results of stacked learning for model 2 were attached. All level-0 algorithms used to employ this model are single based, they are: SVM, KNN, and Neural network. Gradient Boosting was the Meta Classifier used as a level-1. The

score of 0.846 was the high score of accuracy in Chi-square technique. The high score for Roc Curve in this model was 0.491 as shown in Figure 4.6 F1, Precision, and Recall were: 0.845, 0.846 and 0.846.

Table 4.26 Results of stacked ensemble learning for model 2

TECHNIQUE AND NO. OF FEATURES	LEVEL-0	LEVEL-1	AUC	K=10F	F1	PRECISION	RECALL
All Features	SVM. KNN, Neural Network	Gradient Boosting Classifier	0.917	0.820	0.819	0.820	0.820
Relief F-10 Features	SVM. KNN, Neural Network	Gradient Boosting Classifier	0.926	0.831	0.830	0.830	0.831
Anova-8 Features	SVM. KNN, Neural Network	Gradient Boosting Classifier	0.860	0.739	0.737	0.737	0.739
Chi-square 10 Features	SVM. KNN, Neural Network	Gradient Boosting Classifier	0.934	0.846	0.845	0.846	0.846
Information Gain-7 Features	SVM. KNN, Neural Network	Gradient Boosting Classifier	0.872	0.762	0.761	0.760	0.762



a) b)

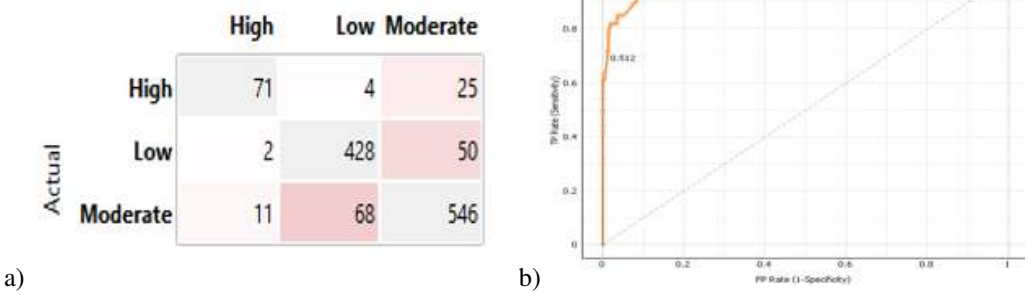
Figure 4.6 a) Confusion matrix and b) ROC curve for stack of model 2

About model 3. The two algorithms that employ a based model (level-0) were bagging, they are: RF and DT. XGBoost is the Meta classifier used for this model to get stack. When the features selection techniques were used with this stack ensemble

model, the best score for accuracy was 0.873 in Chi-square technique and the high point for Roc curve was 0.524. All these results are shown in Table 4.27 and Figure 4.7.

Table 4.27 Results of stacked ensemble learning for model 3

TECHNIQUE AND NO. OF FEATURES	LEVEL-0	LEVEL-1	AUC	K=10F	F1	PRECISION	RECALL
All Features	RF and DT	XGBoost Classifier	0.833	0.731	0.729	0.730	0.731
Relief F-10 Features	RF and DT	XGBoost Classifier	0.957	0.865	0.864	0.866	0.865
Anova-8 Features	RF and DT	XGBoost Classifier	0.710	0.634	0.629	0.632	0.634
Chi-square 10 Features	RF and DT	XGBoost Classifier	0.963	0.873	0.871	0.871	0.871
Information Gain-7 Features	RF and DT	XGBoost Classifier	0.892	0.761	0.760	0.761	0.761



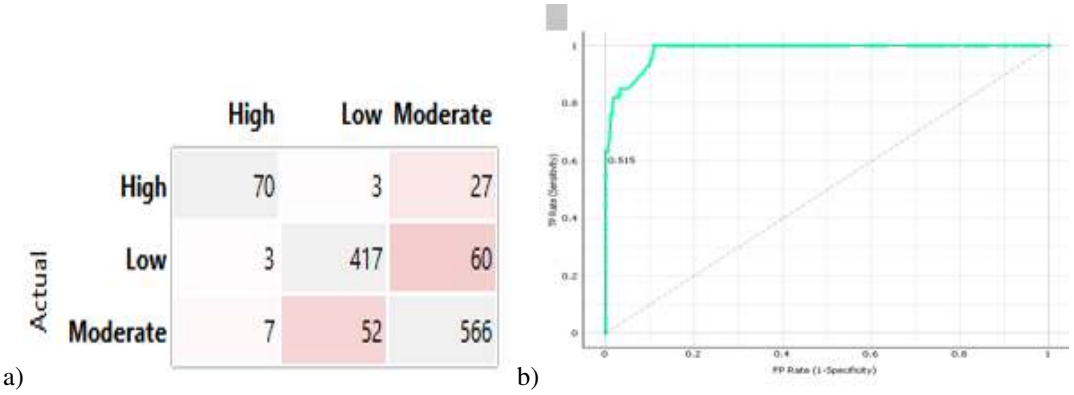
a) b)

Figure 4.7 a) Confusion matrix and b) ROC curve for stack model 3

For model 4 boosting algorithms are employed for a based model: Gradient Boosting and Catboost. The Meta classifier for this model is logistic regression. Also, Chi-square outperformed the other algorithms by achieving an accuracy score of 0.874 and a high score for Roc curve which is 0.515 as shown in Table 4.28 and Figure 4.8.

Table 4.28 Results of stacked ensemble learning for model 4

TECHNIQUE AND NO. OF FEATURES	LEVEL-0	LEVEL-1	AUC	K=10F	F1	PRECISION	RECALL
All Features	Gradient Boosting and Catboost	Logistic Regression	0.956	0.854	0.852	0.856	0.854
Relief F-10 Features	Gradient Boosting and Catboost	Logistic Regression	0.955	0.831	0.824	0.825	0.826
Anova-8 Features	Gradient Boosting and Catboost	Logistic Regression	0.901	0.783	0.780	0.783	0.783
Chi-square 10 Features	Gradient Boosting and Catboost	Logistic Regression	0.960	0.874	0.873	0.874	0.874
Information Gain-7 Features	Gradient Boosting and Catboost	Logistic Regression	0.810	0.710	0.704	0.707	0.710



a) b)

Figure 4.8 a) Confusion matrix and b) ROC curve for stack of model 4

When comparing the results of both the algorithms alone and the stack, it is clear that the results of the stack are superior. Among the types of stack models, the model whose base algorithms are boosting and the Meta classifier is Logistic Regression, using Chi-square technique for feature selection, excels. The accuracy of this model was 0.874.

5. CONCLUSION AND RECOMMENDATION

In this study, the researcher conducted a comparison between the types of machine learning in categorizing students' adaptation to online learning and trying to prove the superiority of stacked learning in improving results to the greatest extent. Four methods of feature selection were used to improve the output. The study aims to find out the extent to which students from different levels of education adapt to online learning. Four different models were developed, each model contained algorithms for a type of machine learning (Bagging, boosting, and Stacking), in addition to a model that contained a variety of algorithms (single and ensemble learning). These four models used four methods of selecting features, which are: Relief F, ANOVA, Information Gain, and Chi-square. Cross validation with k = fold is employed to prevent overfitting. If a low k number is chosen, there will be a high deviation and low variance; if a high k value is determined, there will be a low deviation and large variation. K= 10 was used with classifiers in all models in this study.

When reviewing the results of using feature selection techniques, the superiority of both Chi-square and relief f techniques was clear in most of the models in this study. Despite the difference in the mechanics and mechanism of action of each of them, they gave somewhat similar results in the developed models. From the researcher's point of view, it may be related to the number of features chosen for each technique, as ten were determined by ten for each.

The Gradient Boosting algorithm achieved the highest accuracy rate among the first-level (level-0) algorithms, using the Chi-square technique, as its accuracy reached 0.866 % with a slight superiority over the KNN algorithm, which had an accuracy of 0.863 by using Relief F technique. For the logistic regression classifier (level-1). Stack algorithms (level-1) outperformed first-level classifiers in the accuracy of results in all developed models. The few studies cited in the literature review also demonstrated the superiority of the stacked model. This study is the first of its kind (as far as the researcher knows), which used stacked ensemble learning to classify the level of students' adaptation to online education. The highest value of accuracy

obtained by the stacked learning model 4 Boosting classifiers in level-0 and Logistic Regression in level-1 by using the Chi-square technique was = 0.874%.

On the confusion matrix plot, the rows indicate the actual class, and the columns refer to the predictive class (target class) of the student's adaptivity dataset. The light blue cells represent the correct classification (true prediction). The pink cells represent what is incorrectly classified. To analyze the confusion matrix for the model with the highest value in the results, the first row contained 70 correct predictions out of 100 students who were highly adapted to online education. Incorrect expectations were 3 of them classified as low and 27 of them as moderate. The second row contains 417 correct predictions out of 480 students that have low adaptivity to online education. 566 correct predictions in the third row out of 625.

ROC curve analysis is not affected by decision criterion also, it is independent since it is based on sensitivity and specificity. Figure.4.8 shows the ROC curve for the highest score. The area under the ROC curve (AUC) is used to evaluate the effectiveness of stacked learning accuracy in classification. An AUC of 1 represents a perfect test. Based on the result Table, the highest AUC value which was 0.960 is achieved. This shows the success of the stack of boosting learning (model 3) in predicting and classifying students' adaptation.

5.1 Outline of the Contributions

- A group of models was trained to classify the level of students' adaptation to online education. The highest-resulting model is the fourth stack model, where the based learner (level-0) was composed of two Boosting algorithms: Gradient Boosting and Catboost. The meta-learner algorithm (level-1) was Logistic Regression. The highest accuracy of this model was 0.874 % in Chi-square technique.
- Four methods to feature selection have been tried to get the best results for students' classification to their adaptation. These methods are: Relief F, ANOVA,

Chi-square, and Information Gain. Convergence of performance of both Relief F and Chi-square but the best score was for chi-square.

- Data mining techniques analyze student conditions and their environment, and also study and understand the correlation with their adaptivity to online education.
- The results of this study may help to develop and employ the best online learning solutions.
- When Predict student's adaptation to online education that help them to have different knowledge and backgrounds to achieve their learning goals and develop them effectively in online learning.
- This presented approach proposes more comprehensive and intensified cooperation for institutions concerned with developing curricula in order to improve online learning outcomes.
- This study recommends that developers pay attention to the appropriate technical design of online education tools to be useful as distance learning tools and available to the largest number of students and researchers.
- This study identified the adaptability of students that may support students' online teaching experience and achievement.
- Focusing on trying to increase students' adaptation to online learning and understanding and overcoming the difficulties they face for many reasons, the most important of which is the tendency of many educational institutions to include e-learning in their study plans.

5.2 Future Work

It is important for future studies to consider the importance of internal and external factors that affect students' tendencies and motivations for online learning. Further studies of online learning could contribute depending on different machine learning models to improve the results of classification and predict students' adaptation to this type of education. It is better to be there a future vision for teaching education via the Internet and how to integrate technology with classical education methods in order to obtain an integrated education.

Due to the fact that this study came out with a result of the superiority of stacked learning in classification over the rest of the types of machine learning, so this study recommends the necessity of emphasizing the interest in developing this field and trying other techniques for selecting features with the data set used in order to improve the results.

5.3 Conclusion

Improving the online learning process is through knowing the specific priorities for the level of students' adaptation to online learning. That is knowing the environmental causes and priorities that contribute to increasing student adaptation makes it easier for educational institutions at all levels to develop tools, programs, and curricula. All that in order to ensure that students and learners obtain the best possible educational outcomes. Therefore, this study intended to interpret the students' data set and analyze their classification results by using data mining techniques to help specialists understand and improve the educational system via the Internet.

REFERENCES

Ahmed, N. S. and Sadiq, M. H. 2018. Clarify of the random forest algorithm in an educational field. In 2018 international conference on advanced science and engineering, pp. 179-184, IEEE

Alamri, H. L. Almuslim S. R. Alotibi S. M. Alkadi K., Ullah Khan, D. I. and Aslam, N. 2020. Predicting student academic performance using support vector machine and random forest, In: 2020 3rd International Conference on Education Technology Management, pp. 100-107

Amir, L. R., Tanti, I., Maharani, D. A., Wimardhani, Y. S., Julia, V., Sulijaya, B. and Puspitawati, R. 2020. Student perspective of classroom and distance learning during COVID-19 pandemic in the undergraduate dental study program Universitas Indonesia. BMC medical education, 20(1): 1-8.

Baashar, Y., Alkawsi, G., Ali, N. A., Alhussian, H. and Bahbouh, H. T. 2021. Predicting student's performance using machine learning methods: A systematic literature review. In 2021 International Conference on Computer & Information Sciences (ICCOINS), pp. 357-362, IEEE.

Bahassine, S., Madani, A., Al-Sarem, M. And Kissi, M. 2020. Feature selection using an improved Chi-square for Arabic text classification. Journal of King Saud University-Computer and Information Sciences, 32(2): 225-231.

Bhusal, A. 2021. Predicting student's performance through data mining. arXiv preprint arXiv:2112.01247.

Binkhonain, M. and Zhao, L. 2019. A review of machine learning algorithms for identification and classification of non-functional requirements. Expert Systems with Applications: X, 1, 100001.

Buyrukoğlu, S. and Savaş, S. 2022. Stacked-based ensemble machine learning model for positioning footballer. Arabian Journal for Science and Engineering, 1-13.

Chakraborty, P., Mittal, P., Gupta, M. S., Yadav, S. and Arora, A. 2021. Opinion of students on online education during the COVID-19 pandemic. Human Behavior and Emerging Technologies, 3(3): 357-365.

Charbuty, B. and Abdulazeez, A. 2021. Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2(01): 20-28.

Chen, M., Liu, Q., Chen, S., Liu, Y., Zhang, C. H. and Liu, R. 2019. XGBoost-based algorithm interpretation and application on post-fault transient stability status prediction of power system. IEEE Access, 7: 13149-13158.

Coman, C., Țîru, L. G., Meseșan-Schmitz, L., Stanciu, C. and Bularca, M. C. 2020). Online teaching and learning in higher education during the coronavirus pandemic: Students' perspective. Sustainability, 12(24): 10367.

Dhankhar, A. and Solanki, K. 2021. Comparative analysis of various techniques used for predicting student's performance, In Proceedings of the workshop on technological innovations in education and knowledge dissemination, CEUR workshop proceedings, pp. 10-24.

Elden, A. S., Moustafa, M. A., Harb, H. M. and Emara, A. H. 2013. AdaBoost ensemble with simple genetic algorithm for student prediction model. International Journal of Computer Science & Information Technology, 5(2): 73.

Fahd, K., Miah, S. J. and Ahmed, K. 2021. Predicting student performance in a blended learning environment using learning management system interaction data. Applied Computing and Informatics, (ahead-of-print).

Farhana, S. 2021. Classification of academic performance for university research evaluation by implementing modified naive bayes algorithm. Procedia Computer Science, 194:224-228.

Ghosh, S. K. and Janan, F. 2021. Prediction of student's performance using random forest classifier, In: Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management, pp. 7-11, Singapore

González, S., García, S., Del Ser, J., Rokach, L. and Herrera, F. 2020. A practical tutorial on bagging and boosting based ensembles for machine learning: algorithms, software tools, performance study, practical perspectives and opportunities. Information Fusion, 64: 205-237.

Guang-yu, L. and Geng, H. 2019. The behavior analysis and achievement prediction research of college students based on XGBoost gradient lifting decision tree

algorithm, In: Proceedings of the 2019 7th International Conference on Information and Education Technology, pp. 289-294.

Hasan, R., Palaniappan, S., Raziff, A. R. A., Mahmood, S. and Sarker, K. U. 2018. Student academic performance prediction by using decision tree algorithm, In: 2018 4th international conference on computer and information sciences (ICCOINS), pp. 1-5. IEEE.

Hassan, H., Ahmad, N. B. and Anuar, S. 2020. Improved students' performance prediction for multi-class imbalanced problems using hybrid and ensemble approach in educational data mining. Journal of Physics: Conference Series 1529(5): 052041.

Hu, T. and Song, T. 2019. Research on XGboost academic forecasting and analysis modelling. In Journal of Physics: Conference Series 1324(1):012091.

Hussain, S. and Khan, M. Q. 2021. Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. Annals of data science, 1-19.

Jawthari, M. and Stoffová, V. 2021. Predicting students' academic performance using a modified kNN algorithm. Pollack Periodica, 16(3): 20-26.

Jiao, Y. and Du, P. 2016. Performance measures in evaluating machine learning based bioinformatics predictors for classifications. Quantitative Biology, 4: 320-330.

Karo, I. M. K., Fajari, M. Y., Fadhilah, N. U. And Wardani, W. Y. 2022. Benchmarking Naïve Bayes and ID3 algorithm for prediction student scholarship, In: IOP Conference Series: Materials Science and Engineering, 1232(1): 012002.

Katarya, R. 2019. A review: predicting the performance of students using machine learning classification techniques, In: 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), pp. 36-41. IEEE.

Kawade, D. R., Oza, K. S. and Naik, P. G. 2020. Student performance classification: A data mining approach. JIMS8I-International Journal of Information Communication and Computing Technology, 8(2): 462-466.

Ketui, N., Wisomka, W. and Homjun, K. 2019. Using classification data mining techniques for students performance prediction, In: 2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering, pp. 359-363. IEEE.

Khaire, U. M. and Dhanalakshmi, R. 2019. Optimizing feature selection parameters using statistically equivalent signature (SES) algorithm, In: 2019 4th International Conference on Information Systems and Computer Networks (ISCON), pp. 625-629. IEEE.

Krishnan, R., Nair, S., Saamuel, B. S., Justin, S., Iwendi, C., Biamba, C. and Ibeke, E. 2022. Smart analysis of learners performance using learning analytics for improving academic progression: a case study model. Sustainability, 14(6): 3378.

Kumbhar, P. and Mali, M. 2016. A survey on feature selection techniques and classification algorithms for efficient text classification. International Journal of Science and Research, 5(5): 9.

Lagman, A. C., Calleja, J. Q., Fernando, C. G., Gonzales, J. G., Legaspi, J. B., Ortega, J. H. J. C. And Santos, R. C. 2019. Embedding naïve Bayes algorithm data model in predicting student graduation, In: Proceedings of the 3rd international conference on telecommunications and communication engineering, pp. 51-56.

Lau, E. T., Sun, L. and Yang, Q. 2019. Modelling, prediction and classification of student academic performance using artificial neural networks. SN Applied Sciences, 1:1-10.

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J. and Liu, H. 2017. Feature selection: A data perspective. ACM computing surveys (CSUR), 50(6): 1-45.

Manu, G. P. S. 2016. Classifying educational data using support vector machines: A supervised data mining technique. Indian Journal of Science and Technology, 9: 34.

Martin, A. J., Collie, R. J. and Nagy, R. P. 2021. Adaptability and high school students' online learning during COVID-19: A job demands-resources perspective. Frontiers in Psychology, 12: 702163.

Mulyati, S. and Setiani, N. 2018. Identifying students' academic achievement and personality types with naive bayes classification. Sebatik, 22(2): 64-68.

Muñoz-Carpio, J. C., Jan, Z. and Saavedra, A. 2021. Machine Learning for Learning Personalization to Enhance Student Academic Performance, In: LALA, pp. 88-99.

Mythili, M. S. and Shanavas, A. M. 2014. An Analysis of students' performance using classification algorithms. IOSR Journal of Computer Engineering, 16(1): 63-69.

Nair, S., Saamuel, B. S., Iwendi, C., Biamba, C. and Ibeke, E. 2022. Smart analysis of learners performance using learning analytics for improving academic progression: A case study model. Sustainability (2022)14: 3378.

Narassiguin, A. 2018. Ensemble Learning, Comparative Analysis and Further Improvements with Dynamic Ensemble Selection (Doctoral dissertation, Université de Lyon).

Nawai, S. N. M., Saharan, S. and Hamzah, N. A. 2021. An analysis of students' performance using CART approach. In AIP Conference Proceedings, 2355(1): 060009.

Ng, H., bin Mohd Azha, A. A., Yap, T. T. V. and Goh, V. T. 2021. A machine learning approach to predictive Modelling of student performance. F1000Research, 10.

Niyogisubizo, J., Liao, L., Nziyumva, E., Murwanashyaka, E. and Nshimyumukiza, P. C. 2022. Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization. Computers and Education: Artificial Intelligence, 3: 100066.

Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., and Akinjobi, J. 2017. Supervised machine learning algorithms: classification and comparison. International Journal of Computer Trends and Technology (IJCTT), 48(3): 128-138.

Pal, A. K. and Pal, S. 2013. Classification model of prediction for placement of students. International Journal of Modern Education and Computer Science, 5(11): 49.

Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J. and Phasinam, K. 2021. Classification and prediction of student performance data using various machine learning algorithms. Materials today: proceedings.

Patro, V. M. and Patra, M. R. 2015. Classification of web services using fuzzy classifiers with feature selection and weighted average accuracy. Transactions on Networks and Communications, 3(2): 107.

Perez, J. G. and Perez, E. S. 2021. Predicting Student Program Completion Using Naïve Bayes Classification Algorithm. International Journal of Modern Education & Computer Science, 13(3).

Pintelas, P. E. and Livieris, I. E.. 2020. Ensemble algorithms and their applications. MDPI-Multidisciplinary Digital Publishing Institute.

Qiu, F., Zhang, G., Sheng, X., Jiang, L., Zhu, L., Xiang, Q. and Chen, P. K. 2022. Predicting students' performance in e-learning using learning process and behaviour data. Scientific Reports, 12(1): 453.

Ragab, M., Abdel Aal, A. M., Jifri, A. O. and Omran, N. F. 2021. Enhancement of predicting students performance model using ensemble approaches and educational data mining techniques. Wireless Communications and Mobile Computing, 2021: 1-9.

Rizaldy, A. and Santoso, H. A. 2017. Performance improvement of Support Vector Machine (SVM) With information gain on categorization of Indonesian news documents, In: 2017 International Seminar on Application for Technology of Information and Communication (iSemantic), pp. 227-232, IEEE.

Sandra, L., Lumbangaol, F. and Matsuo, T. 2021. Machine learning algorithm to predict student's performance: a systematic literature review. TEM Journal, 10(4): 1919-1927.

Sekeroglu, B., Dimililer, K. and Tuncal, K. 2019. Student performance prediction and classification using machine learning algorithms, In: Proceedings of the 2019 8th International Conference on Educational and Information Technology pp. 7-11.

Shakeela, S., Shankar, N. S., Reddy, P. M., Tulasi, T. K. and Koneru, M. M. 2021. Optimal ensemble learning based on distinctive feature selection by univariate ANOVA-F statistics for IDS. International Journal of Electronics and Telecommunications, 267-275.

Shardlow, M. 2016. An analysis of feature selection techniques. The University of Manchester, 1(2016): 1-7.

Siddique, A., Jan, A., Majeed, F., Qahmash, A. I., Quadri, N. N. and Wahab, M. O. A. 2021. Predicting academic performance using an efficient model based on fusion of classifiers. Applied Sciences, 11(24): 11845.

Smirani, L. K. and Boulahia, J. A. 2022. An Algorithm based on convolutional neural networks to manage online exams via learning management system without using a webcam. International Journal of Advanced Computer Science and Applications, 13(3).

Smirani, L. K., Yamani, H. A., Menzli, L. J. and Boulahia, J. A. 2022. Using ensemble learning algorithms to predict student failure and enabling customized educational paths. Scientific Programming, 2022: 1-15.

Soni, A., Kumar, V., Kaur, R. and Hemavathi, D. 2018. Predicting student performance using data mining techniques. International Journal of Pure and Applied Mathematics, 119(12): 221-227.

Sridhar, S., Mootha, S. and Kolagati, S. 2020. A university admission prediction system using stacked ensemble learning, In: 2020 Advanced Computing and Communication Technologies for High Performance Applications, pp. 162-167, IEEE.

Sudais, M. and Asad, D. 2022. Student's Academic Performance Prediction–A Review.

Sudrajat, R., Irianingsih, I. and Krisnawan, D. 2017. Analysis of data mining classification by comparison of C4. 5 and ID algorithms. In IOP Conference Series: Materials Science and Engineering, 166 (1): 012031.

Tarik, A., Aissa, H. and Yousef, F. 2021. Artificial intelligence and machine learning to predict student performance during the COVID-19. Procedia Computer Science, 184: 835-840.

Tripathi, A., Yadav, S. and Rajan, R. 2019. Naive bayes classification model for the student performance prediction, In: 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies, pp. 1548-1553, IEEE.

Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S. and Moore, J. H. 2018. Relief-based feature selection: Introduction and review. Journal of biomedical informatics, 85: 189-203.

Vikström, A. 2021. A comparison of different machine learning algorithms applied to hyperspectral data analysis.

Xiao, W., Ji, P. and Hu, J. 2022. A survey on educational data mining methods used for predicting students' performance. Engineering Reports, 4(5): 12482.

Yağcı, M. 2022. Educational data mining: prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, 9(1): 11.

Yuvalı, M., Yaman, B. and Tosun, Ö. 2022. Classification comparison of machine learning algorithms using two independent CAD datasets. Mathematics, 10(3): 311.

Zhang, Y., Ni, M., Zhang, C., Liang, S., Fang, S., Li, R. and Tan, Z. 2019. Research and application of AdaBoost algorithm based on SVM, In: 2019 IEEE 8th joint international information technology and artificial intelligence conference, pp. 662-666, IEEE.

Zhang, Y., Yun, Y., An, R., Cui, J., Dai, H. And Shang, X. 2021. Educational data mining techniques for student performance prediction: method review and comparison analysis. Frontiers in psychology, 12: 698490.