ESTIMATION OF STUDENTS' PERFORMANCE IN DISTANCE EDUCATION USING ENSEMBLE-BASED MACHINE LEARNING

2023
MASTER THESIS
COMPUTER ENGINEERING

Abdullah Raed Fadhil AL-SHAIKHLI

Thesis Advisor
Assist. Prof. Dr. Sait Demir

ESTIMATION OF STUDENTS’ PERFORMANCE IN DISTANCE EDUCATION USING ENSEMBLE-BASED MACHINE LEARNING

Abdullah Raed Fadhil AL-SHAIKHLI

Thesis Advisor
Assist. Prof. Dr. Sait Demir

T.C.
Karabuk University
Institute of Graduate Programs
Department of Computer Engineering
Prepared as
Master Thesis

KARABUK
January 2023

I certify that in my opinion the thesis submitted by Abdullah Raed Fadhil AL-SHAIKHLI titled "ESTIMATION OF STUDENTS' PERFORMANCE IN DISTANCE EDUCATION USING ENSEMBLE-BASED MACHINE LEARNING" is fully adequate in scope and in quality as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Sait DEMIR ....
Thesis Advisor, Department of Computer Engineering

This thesis is accepted by the examining committee with a unanimous vote in the Department of Computer Engineering as a Master of Science thesis. 31/01/2023

Examining Committee Members (Institutions)		Signature
Chairman	: Assoc. Prof Dr. Rafet DURGUT (BANÜ)	....
Member	: Assist. Prof. Dr. Sait DEMİR (KBÜ)	....
Member	: Assist. Prof. Dr. Kürşat Mustafa KARAOĞLAN (KBÜ)	....

The degree of Master of Science by the thesis submitted is approved by the Administrative Board of the Institute of Graduate Programs, Karabuk University.

Prof. Dr. Müslüm KUZU ....
Director of the Institute of Graduate Programs

“I declare that all the information within this thesis has been gathered and presented in accordance with academic regulations and ethical principles and I have according to the requirements of these regulations and principles cited all those which do not originate in this work as well.”

Abdullah Raed AL-SHAIKHLI

ABSTRACT

M. Sc. Thesis

ESTIMATION OF STUDENTS' PERFORMANCE IN DISTANCE EDUCATION USING ENSEMBLE-BASED MACHINE LEARNING

Abdullah Raed Fadhil AL-SHAIKHLI

Karabuk University
Institute of Graduate Programs
The Department of Computer Engineering

Thesis Advisor:
Assist. Prof. Dr. Sait DEMİR
January 2023, 50 pages

Machine learning techniques applied in the educational context can reveal hidden knowledge and patterns to assist decision-making processes to improve the educational system. In recent years, predicting student success in the academic sector has increased interest in improving the shortcomings of academics and providing support to future students. Machine learning techniques have been used to build prediction models using students' academic past records to assist in this task.

The performance of students in academic institutions indicates how much work such institutions need to continue to do to improve their low or even moderate performance. The importance of using machine learning techniques to utilize students' historical data to predict unknown or future performance was an important parameter that encouraged us to build the model
.

Due to its high generalization performance, ensemble learning has attracted great interest. The main challenges of building a strong ensemble are to train a variety of accurate base classifiers and combine them efficiently. The ensemble margin is calculated by taking the vote difference. The number of votes received by the correct class and the number of votes received by another class is commonly used to describe the success of ensemble learning. The classification confidence of the base classifiers is not considered in this formulation of the ensemble margin.

In this study, we applied an ensemble classifier as a classification strategy to predict the substitute achievement prediction model based on machine learning. This model uses discrete datasets to reflect the student's interaction with the teaching model. Various classifiers such as logistic regression, naive bayes tree, artificial neural network, support vector system, decision tree, random forest and k-nearest neighbor are used to evaluate the prediction model of a substitute. Furthermore, cluster processes have been used to improve the appearance of these classifiers. We have used Boosting, Bagging and Voting Algorithms, which are the most common strategies used in the research. As a result, successful results have been obtained using ensemble approaches, and the robustness of the proposed model has been demonstrated.

Keywords: Machine learning, Ensemble learning, Distance education, Performance estimation.

Science Code : 92431

ÖZET

Yüksek Lisans Tezi

TOPLULUK TABANLI MAKİNE ÖĞRENİMİ KULLANILARAK ÖĞRENCİLERİN UZAKTAN EĞİTİMDEKİ PERFORMANSININ TAHMİNİ

Abdullah Raed Fadhil AL-SHAIKHLI

Karabük Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı:
Assist. Prof. Dr. Sait DEMİR
Ocak 2023, 50 sayfa

Eğitim bağlamında uygulanan makine öğrenimi teknikleri, eğitim sistemini iyileştirmeye yönelik karar verme süreçlerine yardımcı olmak için gizli bilgileri ve kalıpları ortaya çıkarabilir. Son yıllarda akademik sektörde öğrenci başarısını tahmin etmek, akademisyenlerin eksikliklerini gidermeye ve geleceğin öğrencilerine destek sağlamaya olan ilgiyi artırmıştır. Bu göreve yardımcı olmak için öğrencilerin akademik geçmiş kayıtlarını kullanarak tahmin modelleri oluşturmak için makine öğrenimi teknikleri kullanılmıştır.

Öğrencilerin akademik kurumlardaki performansı, bu tür kurumların düşük ve hatta orta düzeydeki performanslarını iyileştirmek için ne kadar çalışmaya devam etmeleri gerektiğini gösterir. Bilinmeyen veya gelecekteki performansı tahmin etmek için öğrencilerin geçmiş verilerini kullanmak üzere makine öğrenimi tekniklerini kullanmanın önemi, bizi modeli oluşturmaya teşvik eden önemli bir parametredir.

Topluluk öğrenme, yüksek genelleme performansı nedeniyle büyük ilgi görmektedir. Güçlü bir topluluk oluşturmanın temel zorlukları, çeşitli doğru temel sınıflandırıcıları eğitmek ve bunları verimli bir şekilde birleştirmektir. Topluluk marjı, oy farkı alınarak hesaplanır. Doğru sınıfın aldığı oy sayısı ve başka bir sınıfın aldığı oy sayısı, toplu öğrenmenin başarısını tanımlamak için yaygın olarak kullanılır. Temel sınıflandırıcıların sınıflandırma güvenirliği, topluluk marjının bu formülasyonunda dikkate alınmamaktadır.

Bu çalışmada, makine öğrenimine dayalı ikame başarı tahmin modelini tahmin etmek amacıyla sınıflandırma stratejisi olarak topluluk sınıflandırıcısı yaklaşımı uygulanmıştır. Bu model, öğrencinin öğretim modeliyle etkileşimini yansıtmak için ayrık veri kümeleri kullanır. Bir ikamenin tahmin modelini değerlendirmek için lojistik regresyon, naive bayes ağacı, yapay sinir ağı, destek vektör sistemi, karar ağacı, rastgele orman ve k-en yakın komşu gibi çeşitli sınıflandırıcılar kullanılır. Ayrıca, bu sınıflandırıcıların görünümünü iyileştirmek için küme işlemleri kullanılmıştır. Araştırmada en sık kullanılan stratejiler olan Boosting, Bagging ve Voting algoritmalarını kullanılmıştır. Sonuç olarak topluluk yaklaşımları kullanılarak başarılı sonuçlar elde edilmiş ve önerilen modelin sağlamlığı gösterilmiştir.

Anahtar Kelimeler: Makine öğrenimi, Topluluk öğrenimi, Uzaktan eğitim, Performans tahmini.

Bilim Kodu : 92431

ACKNOWLEDGMENT

Firstly, I 'd like to dedicate this work to my mom for her lifelong sacrifices and to the rest of my family as well. I also thank my advisor Assist. Prof. Dr. Sait DEMİR, for his guidance.

CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

ABBREVIATIONS

ML :Machine Learning

DT :Decision Tree

SVM :Support Vector Machine

RF :Random Forest

DL :Deep Learning

NB :Naïve Bayes

KNN :K-Nearest Neighbor

AB :AdaBoost

GB :Gradient Boosting

CN2 :Cn2 Rule induction

DNN :Deep Neural Network

SGD :Stochastic Gradient Descent

AI :Artificial Intelligence

MLP :Multilayer Perceptron

ROC :Receiver Operating Characteristics

DNN :Deep Neural Network

CA :Classification Accuracy

AUC :Area Under Curve

PART 1

INTRODUCTION

Examining and enhancing performance prediction algorithms is an important part of machine learning. Based on projection results, if the student demands are supplied punctually, then the overall outcome and performance will grow year by year. Subsequently, numerous data mining techniques and classification algorithms are performed [1].

Today's educational system functions in a highly competitive and incredibly complex setting. The greatest obstacle facing today's educational institutions is the need to evaluate past results critically, pinpoint what makes them special, and formulate a plan for the future. The profile of accepted students is an important consideration for learning management systems, as is developing an understanding of the various types and features of students based on the information gathered [2]. They should also think about whether or not they have sufficient information to do an analysis of incoming students and choose the best way to structure their marketing strategy and reach out to the most promising prospective students [3].

Changes in student conduct as a result of learning experiences might be seen as a source of improved academic achievement [4]. A student's learning outcomes are the development of a latent talent or aptitude. It is possible to see how lessons affect students' actions [5] [6]. These learning outcomes manifest themselves in students' understanding of information, reasoning skills, and motor skills. Student performance can be seen in its concrete form through indicators such as comprehension of material, facility with information processing, and the maturity to use one's learned skills as the basis for one's reasoning and behavior [7]. In light of these insights, the knowledge, attitudes, and skills of students can be evaluated after a course of instruction. Because students' success is tied to the quality of their education, it makes sense to utilize learning outcomes as criteria for enhancing instruction [8]. Machine learning can be applied to many different fields. Machine learning is used to improve the connections between search terms and results. Search engines can deliver the most relevant results for a particular search query [9] by analyzing the content of websites to identify which words and phrases are most important in defining a web page. Machine learning is also used by image identification systems to distinguish objects like faces [10]. To begin, the ML algorithm ranks all the pictures that contain the target object. Given a large enough sample of images, the system can determine with high confidence whether or not an image contains the target object [11]. Furthermore, machine learning can be used to deduce the possible types of things a customer could be interested in purchasing. Taking into account the user's past purchases, the computer can recommend more items of interest [11]. All these examples have the same basic premise. The idea of "machine learning" emerged from this setting. Finding patterns and laws in digital data is a complicated task for which computers are well-suited. Machine learning is predicated on the premise that computers may, in principle, learn automatically from experience. Many different fields can benefit from machine learning, but the core concept remains the same. Massive data sets are analyzed by the computer to reveal previously unseen laws and patterns. These laws and regularities have a mathematical basis and may be defined and processed efficiently by a computer. With such criteria in hand, the computer can accurately label incoming information. A machine learning system may automatically create rules from data, and these rules get better as more data is supplied to the system [6].

To learn a set of rules from examples is one definition of machine learning [12], while building a classifier that can generalize from new examples is another. The development of a classifier requires two stages. Initially, we start from scratch and use the provided data to construct a classifier model. Training is the term for this kind of instruction. The building of classification rules occurs at this stage. Step two, called screening, involves verifying that the criteria for classification are reliable [12].

Classification and regression are the two main branches of supervised learning. Predicting the class from labeled data, which has provided a choice from a collection of options, is what classification is all about. However, regression analysis aims to extrapolate the nature of the regression connection from the available data [13].

Unsupervised learning is the second method. With unsupervised learning, not every piece of input data is labeled. Supervised learning algorithms, on the other hand, just take the labeled data as is and extract the relevant information from it [13]. Modifying datasets and combining similar examples are two typical practices in unsupervised learning. It is possible to make existing datasets more digestible by using a computer-assisted process known as dataset transformation [10].

In the clustering phase, the algorithm divides the information into groups with similar characteristics. Photos shared on social media can be organized into albums based on the people depicted in the images using a clustering algorithm. To use machine learning to a problem successfully, one must have a thorough comprehension of that problem. A complete comprehension of the issue may influence the type of data that has to be collected and the algorithm that should be used. Machine learning algorithms (supervised or unsupervised) must be tailored to the available data and the desired outcomes of any analyses [14].

In this research project, we investigate how we can utilize strategies of machine learning to foretell a student's academic achievement. The thesis focuses on contrasting the predictive performance gains made by various machine learning approaches and feature engineering methodologies. In this thesis, three distinct machine learning approaches were utilized. This group includes naive Bayes classification, decision trees, and linear regression. Predictions from these learning algorithms were enhanced by engaging in feature engineering, or the process of altering and selecting the features of a data set.

1.1. MOTIVATION

Because of the proliferation of computing devices and the accessibility of the internet, there is now a wealth of publicly available data ripe for statistical examination. New data is created daily, including online sales figures, website traffic, and user patterns. Huge data sets are both a challenge and an opportunity. Humans need help making sense of such massive amounts of data. The upside is that computers can handle this data considerably more quickly than humans because it is digitally recorded in a well-formatted way.

Machine learning is an idea that evolved in this setting. Computers can analyze digital data in ways that are far too complicated for a human to handle. The core tenet of machine learning is that data may be used to teach a computer new skills. While the specifics of where machine learning is used can differ, the underlying purpose is always the same. The computer sifts through mountains of data, looking for hints of structure and order [15]. Patterns and principles like these have a mathematical basis and are, therefore, amenable to the formal definition and computational processing. A computer system can then utilize those rules to classify new data meaningfully. Making rules from data is an automated process that improves over time as more data is supplied.

1.2. PROBLEM STATEMENT

The importance of education stems from the sector's focus on individual students and their academic progress. With the right tools, schools can use of the massive amounts of data they collect from students' academic records. This information can be used to conclude connections and ideas. It can also be used in early estimation and in developing decision support systems to assist educators and students in making informed choices.

There is a need to examine and comprehend health and these data in order to ascertain the development of the learning procedure because factors like social factors are deciding factors in many educational settings. Consider factors like average income, location, and educational attainment.

However, exploring this data necessitates using suitable techniques and resources for gleaning insights from these records. Many statistical methods have been developed in recent years for estimating students' characteristics. However, the methods still lack a reliable capacity for analyzing and forecasting student performance.

As a result, there is a growing demand for the application of powerful machine learning techniques in the classroom, where they can be used to extract useful insights from students' past records. Many issues in the classroom can be resolved if only we take the time to learn everything we can about the data and the factors that influence children's performance. The predictive power of each variable is used by machine learning techniques to help estimate performance and draw conclusions about causation. There is a technique for establishing a link between the known variants and the outcomes. This, however, necessitates an in-depth familiarity with each individual student and their circumstances.

1.3. AIMS OF THE STUDY

This thesis's focus is on using ensemble-based machine learning strategies to the problem of classifying and forecasting student performance. The study aims to accomplish the following major objectives:

- One goal is to catalog the various techniques and instruments already used to forecast student performance.
- To learn how to classify the variables involved in making predictions.
- Examine the data to see if any trends can be used to forecast how well new students will do in college based on their personal and academic backgrounds.
- Analyze student performance and recommend effective study methods to help students improve their education.

- Explores the use of machine learning algorithms to estimate a student's performance in school. The thesis primarily examines the relative efficacy of various machine learning approaches and feature engineering techniques for improving prediction performance.

1.4. THESIS STRUCTURE

- *Part 1. Introduction:* Lays out the context, provides the reasoning behind the work and identifies the issue at hand. The aims of this research are also outlined.
- *Part 2. Related Work:* This section reviews relevant literature and draws attention to some of the study and effort that has already been done on our topic. Provides some necessary context for this thesis.
- *Part 3. Theoretical Background:* Summaries background about techniques.
- *Part 4. Methodology:* Contains the approach that has been envisioned for the research of this thesis based on machine learning techniques.
- *Part 5. Results and Discussion:* Discusses the outcomes that were obtained and what those results mean.
- *Part 6. Conclusions:* We conclude this thesis by summarizing its findings and discussing its contributions. In the same part, we get a quick preview of what can be done in upcoming future.

PART 2

RELATED WORK

The intention of this study is to examine the efficacy of several algorithms using machine learning for making predictions about academic performance. Here, we take a look at the latest recent research and projects in our area.

The ability to accurately forecast how students will do in their classes has piqued the interest of a number of authors. This is due to the significant role that accurate forecasting plays in assisting teachers in correctly recognizing and supporting their students. Significant time and energy have been spent in recent years analyzing student behavior to uncover useful patterns that can be used to foretell students' future successes.

Predicting students' future success with the use of machine learning and data mining is the goal of educational data. An early projection of student performance can enable responsible parties to support low-performing kids. Student final exam performance might be affected by past assignment grades, social life, parents' jobs, and absence frequency. A comprehensive literature study on student performance prediction using the Kaggle data set will be presented in this chapter.

Thaer Thaher et al. [16] developed an accurate student performance model. Feed-forward multi-layer perceptron with stochastic training procedures is proposed. The suggested approach is evaluated using three UCI and Kaggle datasets, while SMOTE oversampling algorithm was used to handle unbalanced data. Support Vector Machine, Decision Trees, K-Nearest Neighbors, Logistic Regression, Linear Discriminant Analysis, and Random Forest were evaluated. The study found that the MLP coupled with a stochastic method known as Adam outperformed traditional and prior classifiers on most datasets with an accuracy of 0.91% [16].

In another work, Utomo Pujianto et al. [17] two classifiers, C4.5 and k-Nearest Neighbor (KNN), are used to classify students' academic achievement, and the SMOTE oversampling technique is applied to evaluate their relative efficacy. After running experiments with the Rapid Miner software, the study found that the C4.5 Decision Tree approach outperformed the K-Nearest Neighbor algorithm. The preprocessing that was presented in [17] entailed the elimination of features that had less influence on the classification results. Resulting in the remove of all three attributes from the primary dataset. Then after, the resulting dataset is prepared for classification using an oversampling algorithm known as SMOTE (Synthetic Minority Oversampling Technique) approach to equalizing the class's demographics. The results showed that the decision tree c4.5 algorithm outperformed the K-Nearest Neighbor algorithms in terms of accuracy (74.09%), recall (74.04%), and precision (75.05%) [17].

V. Vijayalakshmi et al. [18] present a deep learning-based approach for predicting the academic excellence of students. Both Keras and Tensor flow libraries were applied in the construction of the model and tested on the Kaggle dataset. The recursive feature elimination technique incorporated into the framework is a Wrapper-based greedy optimization algorithm, which aims to identify the subset of features with the highest classification accuracy by eliminating unrelated features. The Deep Learning model consists of one output layer that makes use of the softmax activation function and two hidden layers that make use of the ReLu activation function. There are a total of 10 neurons in the first hidden layer, whereas there are 55 neurons in the second hidden layer. Following the fine- tuning of 243 deep neural trainable parameters, model's accuracy increased to 85% [18].

Suad Almutairi et al. [19] argued that data mining and six different machine learning algorithms could be used to I identify the most influential behavioral and demographical factors that contribute to the prediction of students' performance, and (ii) forecast student's academic achievement. Additionally, the study aimed to show how feature selection, oversampling, ensemble learning, and parameter tweaking can boost the models' predictive ability while also addressing the issue of overfitting. Multiple machine learning strategies, including random forest, logistic regression, XGBoost, MLP, and ensemble learning via bagging and voting, have been examined. Each classifier's hyper-parameters have been adjusted for optimal efficiency during training. Variations in the values given those hyper-parameters can have a significant impact on the model's performance. Grid search is utilized to hasten the procedure of selecting a value. Overfitting can be resolved with the use of certain hyperparameters, such as the maximum depth of the trees; a bigger value implies a deeper tree, which will lead to the collection of more information on the data; other hyperparameters include the number of trees employed in the study [19]. The C parameter (the inverse of regularization strength) provides additional data for managing regularization and limiting overfitting in logistic regression. The power of regularization increases as the value drops. The highest accuracy, 77%, was achieved with selecting the 10-best characteristics using a random forest and overfitting was greatly decreased by tuning the hyper-parameters. The results showed that data mining could accurately predict the students' performance levels and highlight the most influential features [19].

In the domain of statistics and machine learning, ensemble methods are used to get superior prediction performance than could be obtained from any of the constituent learning finds by itself. This is accomplished by combining multiple learning algorithms into a single model. Common ensemble approaches such as Bagging, Boosting, and Voting Algorithms were utilized. Samuel-Soma et al. [20] made use of ensemble approaches to improve the accuracy of the evaluated results obtained from classic classifiers. These traditional classifiers included Nave Bayes (NB), Decision Tree (ID3), Support Vector Machines (SVM), and K-Nearest Neighbor (KNN). The researchers also investigated a variety of feature ranking measures, among which include the Information Gain Ratio. In this study, we employed a filter method that was based on Information Gained to evaluate feature rank. Our goal was to establish which features are the most useful when it comes to constructing a model of student performance. Each ensemble trains all four classifiers independently then combines the data obtained from those training sessions using a voting system in order to get the highest possible student prediction performance. With an accuracy of 92.3% when using behavioral features and 88.6% when without using behavioral features, KNN easily surpasses other Data Mining Techniques in a head-to-head comparison.

While the voting ensemble of KNN and ID3 worked remarkably well, with an accuracy of 96.8% [21].

Understanding and predicting the academic achievement of students based on characteristics gleaned from electronic learning management systems is the objective of the line of inquiry that Farrukh Saleem et al. [22] proposed bagging, boosting, stacking, and voting are the four ensemble techniques that are applied to the model in order to improve it further. The model that has been proposed is comprised of five classic machine learning algorithms. The experiment was carried out by the study utilizing five different ML models (DT, RF, NB, KNN, and GBT) without employing an ensemble. Within the confines of a single experiment, each model was both trained and validated. In addition, the four ensemble approaches were utilized to mix the ML models and evaluate the amount of achievement through the utilization of precision, recall, and F1 score. Experiments were carried out in order to determine whether or not a multi-method approach, known as an ensemble technique, might improve the accuracy of the model even further. The bagging (0.785%) and boosting (0.783%) methods impressively contributed to RF's level of accuracy. In addition, nearly all of the models improved their accuracy by bagging and boosting. In addition, the stacking ensemble strategy obtained above 80% accuracy in forecasting accurate cases, which is the highest percentage among other approaches. The RF model that included the information gain criterion in this research demonstrated superior performance when compared to the others (0.777%). The NB performance improved from 0.645% (single) to 0.706% (with boosting). Similarly, the integration of GBT with AdaBoost resulted in an increase in accuracy of approximately 4%.

Mahmud Ragab et al. [23] evaluated various ensemble-based classifiers to data mining techniques in different research projects to predict students' academic achievement. Logistic regression, naive bayes tree, artificial neural network, support vector machine, decision tree, random forest, and k-nearest neighbor were a few of the ensemble-based classifiers. To boost productivity, the different classifiers are gathered in one place, and the Vote method is utilized to include them in the ensemble process. According to the implementation's results, the bagging strategy was successful in significantly improving the DT model. To be more precise, the DT algorithm's accuracy with bagging increased from 90.4% to 91.4%. Recall scores improved from 0.904% to 0.914%, while precision scores followed suit from 0.905% to 0.915% [23].

Classifiers like the naive bayesian (NB), decision tree (DT), k-nearest neighbor (KNN), discriminant analysis (Disc), and pairwise coupling (Pairwise) can be improved with the help of ensemble algorithms like AdaBoost, Bag, and RUSBoost, as recommended by Samuel-Soma et al. [21]. (PWC). Each ensemble must train all five classifiers and then integrate those findings using a voting mechanism in which the majority vote is cast in order to obtain the best potential prediction performance from the students. The boosting methods outperform the other ensemble methods in terms of NB and DT performance. Through the use of boosting, NB's accuracy goes up from 0.83% to 0.85%, while Precision results go up from 0.75% to 0.77%, and Recall results go up from 0.76% to 0.77%. Accuracy for DT has gone up from 0.77% to 0.78%, precision from 0.75% to 0.78%, and recall from 0.74% to 0.86%. In addition, the results show that there is a substantial and significant link between students' actions and their performance in the classroom. The prediction model's accuracy increases to 84.2% when behavioral elements are included but drops to 72.6% when they aren't. With the addition of ensemble methods, precision rose to 94.1%, which is a significant improvement over previous efforts [21].

Feature selection is the process of narrowing down a big dataset to a manageable collection of useful features (or variables) for modeling purposes. Feature selection goes by a few different names, including variable/attribute/variable subset selection. In order to foretell students' learning rates and actions, Rasheed Mansoor et al. [24] ran tests employing a Linear Discriminant Analysis (LDA) and Convolutional Neural Network (CNN) model method. LDA is a well-known method for minimizing the number of dimensions, and it is utilized extensively in the fields of machine learning and pattern recognition. The retrieved features from CNN analysis are filtered using the mRMR approach, which stands for the Minimum Redundancy Maximum Relevance method. Stochastic gradient descent (SGD) is used to measure the feature weights, and these weights are then modified so that to improve CNN learning. Experiments demonstrated that the proposed model produces higher levels of accuracy (96.5%), precision (94.0%), recall (92.0%), and F-score (95.0%) than the approaches that are currently in use, all while requiring less time to compute. In comparison to those other approaches in [9], the execution of the suggested method took just 3.2 seconds when it was applied to 500 records and only 6.5 seconds for 2500 records [24].

Sana et al. [25] were particularly concerned with the timeliness of pupils as well as the impact that parental involvement had on their children's education. This group of functions centers on the communication between the student and the LMS. The filter approach is applied to determine the relevant subset of features while avoiding the other ones. These methods rate the features by utilizing variable ranking algorithms in order to allow for the selection and application of highly ranked features to the learning algorithm. The proposed model was able to attain an accuracy of up to 10% to 15%, which is a significant improvement in comparison to the results obtained when similar elements were eliminated. It has come to our attention that ANN performs better than other classification methods. The Artificial Neural Network achieves an accuracy of 78.1% when it uses highly rated features, whereas it only achieves 59.1% when it does not use ranked features. Similarly, Md. Hasibur Rahman et al. [26] model demonstrates that ANN outperformed other models using ensemble-based features filtering with the accuracy of 84.3%.

The categorization and forecasting procedures benefit tremendously from the utilization of data mining techniques because these procedures are among the most efficient and potent technologies available. Results from educational data mining can provide administrators with useful insights that can be used to raise the quality and efficiency of their institutions. Resul Butuner et al. [27] examined Random Forest, AI, Naive Bayes, SVM, LR, and DL. Deep Learning, Random Forest, and Support Vector Machines produce better prediction results than others. Rapid Miner Studio and Orange are used to examine data, perform data mining techniques, and assess models by generating new ones. Deep Learning (DL), Support Vector Machines (SVM), LR, and Random Forest (RF) algorithms obtained above 96% success, while LR scored 0.998% in validation.

Neural Networks are a sort of data mining that Mussa S. Abubakari et al. [28] utilized to learn about relationships between 480 students and 16 factors. In order to partition the dataset for the training and testing phase, a cross-validation with a 10-fold was performed. The method continued with fitting the model over the course of 200 iterations (epochs) using 10 batch sizes of inputs, which was then followed by the outcomes evaluation for the purpose of producing a knowledge representation. When the Adam model optimizer was utilized, the resultant accuracy was lower than 60% percent. However, after utilizing the dropout technique and the stochastic gradient descent optimizer, the accuracy was improved to be greater than 75%. The overall stable accuracy that was acquired was 76.9%, which is a figure that may be considered satisfactory. This suggests that the NN model that was suggested can be reliable when used for prediction, particularly in the field of social science studies.

The finding of the reviewed literature can be summarized in the following Table 2.1

Table 2.1. Related studies.

Reference	Year	Datasets	Method	Best Results/Accuracy
Thaer Thaher et al. [16]	2020	Different dataset	(Svm, DT, K-NN, Logistic Regression, Linear Discriminant, RF, MLP) with Synthetic Minority Oversampling Technique SMOTE	MLP - 0.91%
Utomo Pujianto et al.[17]	2020	Same dataset	(C4.5 and KNN) with Synthetic Minority Oversampling Technique SMOTE.	C4.5 - 74.09%
V. Vijayalakshmi et al. [18]	2019	Same dataset	Deep Learning with ReLu activation function	85%
Suad Almutairi et al. [19]	2019	Same dataset	(RF, LR, XGBoost, MLP) with ensemble methods	Random Forest – 77% (with ensemble and hyper-parameters tuning)
Samuel-Soma et al. [20]	2022	Same dataset	(NB, DT, Svm KNN) with ensemble and Information Gain features selection	(KNN and ID3) – 96.8%
Farrukh Saleem et al.[22]	2021	Same dataset	(DT, NB, RF, KNN, GBT) with ensemble (Bagging, Boosting, Voting, and Stacking)	Random Forest – 0.785% (with Bagging ensemble)

Mahmoud Ragab et al.[23]	2021	Different dataset	(NB, ANN, Svm, DT, RF, and KNN) with ensemble learning	Decision Tree – 91.4% (with bagging ensemble)
Samuel-Soma et al. [21]	2020	Same dataset	(NB , DT, K-NN, Discriminant Analysis, and Pairwise Coupling) with FS and ensemble	Decision Tree – 94% (with ensemble and feature selection)
Rasheed Mansoor et al.[24]	2022	Different dataset	Hybrid model of Linear Discriminant Analysis (LDA) and the CNN	96.5%
Sana et al. [38]	2019	Same dataset	Hybrid model ANN and filter-based feature selection	78.1%
Md. Hasibur Md. Hasibur Rahman et al. [26]	2017	Same dataset	(NB, ANN, DT, NB,ANN,DT, KNN) with feature selection and ensemble	84.3%
Resul Butuner et al. [27]	2021	Different datasets	(RF, AI, NB, Svm, LR, and DT)	LR – 99.6%
Mussa S. Abubakari et al. [28]	2020	Same dataset	Neural Network with Adam model optimizer	76.9%

PART 3

THEORITICAL BACKGROUND

Each artificial neural network (ANN) comprises two distinct sorts of computations. It is possible to perform two passes, one in the forward direction, from input to output, and another in the reverse direction, from output to input [7]. Using the network's inputs and the forward pass computes the network's output by feeding the results of the computations performed in each layer into the next. The weights are then updated using gradient descent in the backward pass. Derivatives of the output to the weights can be computed by comparing the values produced by the ANN in the forward-pass to the target values in the dataset. If you use gradient descent, you may determine the inverse of the gradient descent at a given site, which is the value of the location weights that needs to be adjusted in order to reduce the error. In this manner, a neural network can be updated to perform a task by providing the appropriate output in response to the specified input values. By decreasing the deviation (or "loss") between the forward pass output and the expected output, back propagation can boost a neural network's performance [29][30].

The applications of ensemble methods have broadened to encompass a wide range of industries, from healthcare and finance to insurance and transportation to manufacturing and even bioinformatics and aerospace [31]. Joining and evaluating the various methods in some way results in an ensemble of classifiers that may be used to classify new test data. When it comes to supervised learning, ensemble learning has emerged as a major subject of study for machine learning experts [32].

3.1. MACHINE LEARNING CLASSIFIERS

Machine learning (ML) is an important subfield of AI that focuses on developing methods for computers to teach themselves new information. Predefining every rule for ML approaches to make a choice or extract a pattern is unnecessary. Moreover that's possible because of the extensive data sets used throughout its training, which illuminate its architecture and conceptualization. This means that the algorithms are self-taught [33]. The term "machine learning" (ML) refers to a computer's capacity to make reliable forecasts based on available data. Recent years have witnessed tremendous advancements in ML as a result of the exponential growth in computer storage space and processing capacity. There are many benefits to using ML, one of which is the capacity to analyze massive data sets and discover patterns. The ability to quickly and easily interpret data based on images to aid in the making of complex judgments by specialists. Additionally, it allows for the rapid processing of massive volumes of data, something the human brain just cannot do [34].

Learning is only one of several industries where ML methods have found widespread application. As a result of the high price tag and extensive effort required to analyze educational data, ML methods have been adapted for use in the healthcare industry [35]. ML can be a good alternative to conventional approaches when time and money spent on development are of paramount whenever the topic is of paramount importance or appears excessively sophisticated to be examined in its entirety [36]. The three primary forms of ML are depicted in Figure 3.1

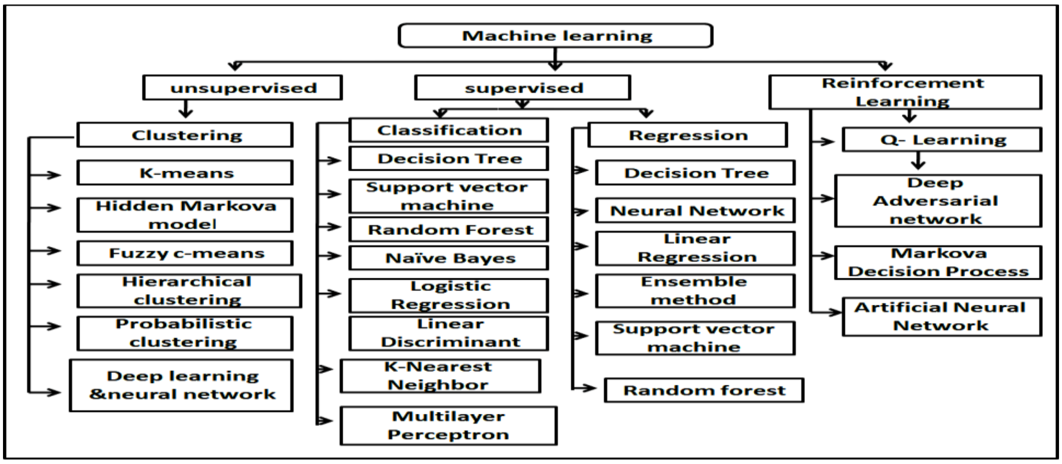

Figure 3.1. Illustration of ML techniques.

3.1.1. The Decision Tree Technique

Systems that generate classifiers are a popular data science technique [37]. Classification algorithms can process massive amounts of data, making them useful in data mining. Assumptions about categorical class names, knowledge classification using training sets and class labels, and the classification of newly obtained data are all possible uses [38]. Several different classification methods exist in the field of machine learning.

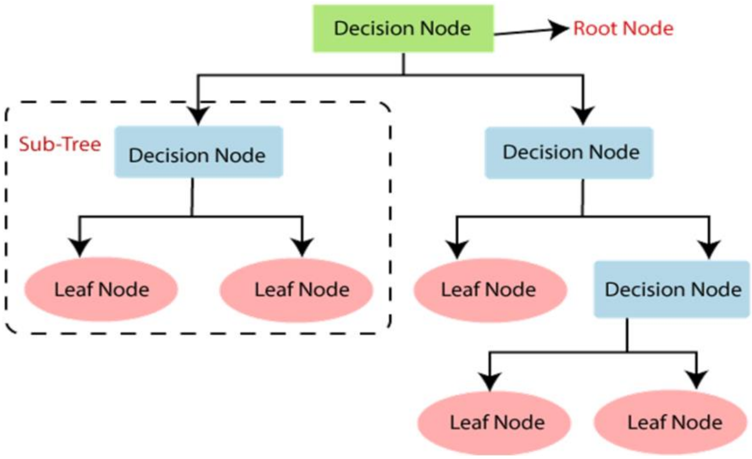

Figure 3.2. The structure of the technique [31].

3.1.1.1. Benefits of DT Method

- It is clear and uncomplicated.
- To build a tree, the DT method does neither require massive amounts of data preparation nor does it have a disproportionately high cost [39].
- It can be used for both numerical and category information.
- Both single and multiple outcome predictions are supported, in addition to binary ones.
- The (DT) method is a type of "white box" methodology, which means that its inner workings are completely transparent.
- There are quantitative measures that can be used to evaluate the algorithm's efficacy.

3.1.1.2. Limitation of with DT

- Overfitting is a typical issue when employing the DT method. Pruning, figuring out how few specimens are needed per leaf node, and measuring the tree's depth are some of the techniques utilized to lessen this issue.
- One possible result of outliers in a decision tree is increased volatility. Decision trees used as part of an ensemble approach are useful here.
- Predictions based on a decision tree are not continuous but rather are piecewise constant approximations.
- XOR and equivalency difficulties are two examples of ideas that are difficult for DT to convey.
- Biased trees may be produced if the data set's categories aren't spread out uniformly.

3.1.2. The Support Vector Machine (Svm) Technique

It was developed in the middle of 1990 as a powerful supervised machine learning method based on statistical learning theory. Svm is an efficient technique for making predictions and classifying data [40]. Even so, it is among the top methods for classifying data using machine learning [41]; hence it is mostly employed for that purpose. Classification is performed by segmenting the input space of the dataset in either a linear or non-linear fashion [42]. As can be seen in Figure 3.3, this is accomplished by defining the hyperplane in an N-dimensional vector space by establishing a hierarchy between two types of items. If there is more than one possible hyperplane dividing the classes, then the one with the highest margin distance between them is selected. The decision border can be moved with the help of support vectors, which are the points nearest to the boundary [43].

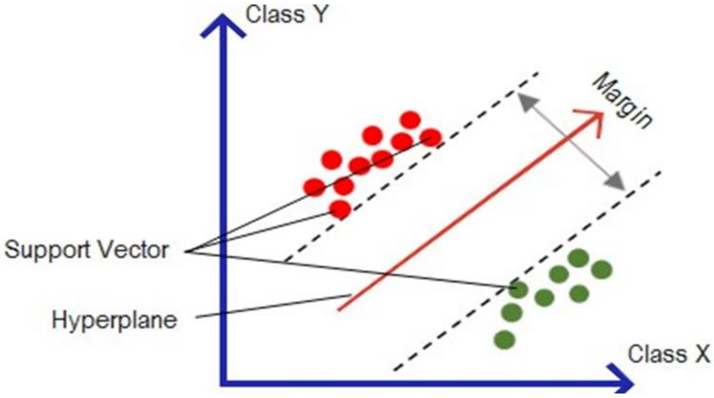

Figure 3.3. Sum margin [31].

3.1.2.1. Benefits of Svm

- When the total amount of dimensions is larger more so than the total quantity of samples, the approach is still usable, and it shows good performance in high-dimensional spaces.
- The decision function's utility is increased by the flexibility with which its definition use a large variety of kernel functions as its foundation.

3.1.2.2. Limitation of Svm

- Over-fitting must be avoided costs when selecting Kernel functions, and regularization terms. Suppose the number of characteristics is much beyond the range of the samples taken.
- Since SVM employs five-fold cross-validation to derive probability estimates, sometimes it might be a lengthy process.

3.1.3. Random Forest Technique

Like other supervised learning techniques, random forest (RF) applies to this situation to accomplish classification and regression-related tasks. The system forms several trees and uses a hive mind technique to make judgments (i.e., a forest). One advantage of this group method is that it trains numerous models simultaneously and then merges them into a single decision tree. RF is an ensemble method that combines and associates several decision trees into a single basic learner model. This algorithm often employs the bagging or voting strategy with random trees when combining them [42]. Bagging, boosting, voting, and stacking are only a few methods used in ensemble learning. Having a higher level of prediction and accuracy is the primary goal of developing and integrating several decision trees. The method reduces error and improves precision by constructing numerous trees [44].

This strategy was employed multiple times on a learning dataset addressing a variety of issues. Because of this, it is simple to comprehend when several characteristics are split using splitting rules and helpful for making forecasts. The resulting categorization tree incorporates a wide variety of techniques. You can get to the desired node with the use of characteristics and data values. Specifically, the first part of this research involves the model used to apply the RM tool on its own. In a second point, the experiment was carried out many times using different ensemble techniques [45].

3.1.4. Gradient Boost Technique

To classify and predict data, this model uses a decision tree. This algorithm, which creates new predictions by learning from old ones, goes under another name, the forward learning ensemble technique. This model is called "boosted" because it uses additional work to learn from errors and increase the decision model's accuracy and precision. Using a forward learning method, the model offers a predicted score that may be used analytically. It is a collaborative tree-structure model that combines multiple weak trees into a single robust one [46].

The widespread application of GBT on educational datasets attests to its practicality and efficacy in forecasting a wide range of pedagogical characteristics. In the Brazilian public education system, for instance, the concept of using GBT to analyze and forecast student performance has lately been utilized. The research demonstrated that internal grades and absences are two of the most important factors in producing a reliable predictive model [31]. In addition, the gradient boosted decision tree model has been used multiple times with proven efficacy in anticipating the dropout ratio of enrolled students for both online and on-campus education [38]. Due to its potential for improving model accuracy, this classifier was also chosen for further testing in this study, where it was used both with and without the ensemble method. At each iteration, gradient descent (GD) learns the optimal values for the model's parameters in order to decrease the cost function (CF) [47]. SGD is a stochastic variation of GD in which a single sample is picked randomly for model training at each iteration [48]. The training time needed to obtain local minima is drastically reduced if only one training sample *xi* is used to find CF on each iteration. It accomplishes this by recalculating the model's parameters in light of new information about the relationship between *xi* and *yi* in each iteration.

$$\theta_j = \theta_j - \alpha(y^{i\,\mathrm{r}} - y^i)x^i \qquad (3.1)$$

Where *j* represents a parameter and is the model's learning rate. The effectiveness of SGD on the data at hand is supported by a number of hyperparameters.

3.1.5. Naive Bayes (Nb) Technique

In addition to traditional classification methods, this research uses Naive Bayes (NB). It can be applied to high-dimensional data for analysis and prediction, it is user-friendly, and it works with huge datasets [31].

This program uses the Bayesian theorem to determine the probabilities associated with each class given a set of inputs. Its wide and robust use in a variety of contexts, including spam classification [48], weather forecasting systems, and sentiment analysis [38], has led to its description as a potent ML algorithm. The NB model was implemented in this research to forecast a student's performance based on a variety of inputs. In the first step, we examined the model's prediction accuracy was applied as a single learner with a variety of settings. It was also employed with the help of several ensemble techniques like boosting, bagging, stacking, and voting. The implementation portion covers every conceivable case while a thorough evaluation of the final model's efficacy is conducted.

3.1.6. K-Nearest Neighbor (Knn) Technique

It is well known that the supervised classification algorithm k-nearest neighbor (KNN) works well with the ML method. It uses the well-known k-nearest neighbor technique, in which unknown samples are compared to k-training examples. An instance is assigned to the training example that is closest to it, depending on the distance between the two [49]. Since the dataset included in this analysis is of a composite nature, the distance was determined using the "Mixed Euclidean Distance" method. The dataset was subjected to KNN with and without the use of ensemble methods for prediction. In the results section, we delve deeper into the classifier's effectiveness.

3.1.7. Artificial Neural Network (Ann) Technique

An ANN is a network of inputs and outputs where each connection has a certain weight. It has three layers: an input layer, a layer in the middle, and an output layer [50]. Changing the weight of a connection is how a neural network learns. The network's efficiency rises with each iterative update of the weight. There are two distinct types of ANN, known as feed-forward networks and recurrent networks, depending on the type of connections they use. Unlike recurrent neural networks, in which connections form a cycle, feed-forward networks' connections between units do not repeat themselves. How a neural network act depends on its learning rule, architecture, and transfer function [51]. The neural network's neurons are stimulated by the input's weighted sum. The transfer function takes the activation signal and generates a single output from the neuron. This transfer function causes the network to be nonlinear. The network's accuracy is improved during training by adjusting the weights of its interconnections. Its many benefits include parallelism, resistance to noise, and a strong learning capacity [52].

3.1.8. AdaBoost Technique

By re-weighting each example the learner has recognized, adaptive boosting (AC) is an ensemble method that gives extra weight to incorrectly categorized instances [38].

The sequential method through which AC constructs an ensemble of learners directly contributes to the decreased variance and bias in the models. Boosting can transform a weak learner into a strong one by focusing on examples that the weak learner misclassified in the past. The weak learners *k1*, *k2,..., km* are merged in the following way for a given dataset *(x1, y1),..., (xn, yn)* where each instance *xi* has a matching target variable *yi1,+1*.

$$C(t1)(xn) = w1k1(xn)+,,,+w(t1)k(t1)(xn), \tag{3.2}$$

where, w1,,,w(t1) is the weight assigned to each Calculating Ct (xn) is as simple as:

$$Ct\ (xn) = C(t1)(xn) + wt\ Ct\ (xn). \tag{3.3}$$

3.1.9. Logistic Regression Technique

Like multivariate linear regression, logistic regression models the effect of several variables at once. The supervised machine learning method logistic regression was designed to help with learning categorization difficulties. When the goal variable is categorical variable, we have a classification learning problem. Assigning a probability to a new example based on the likelihood that it belongs to one of the target classes is the purpose of logistic regression, which is achieved by mapping a function from the dataset's attribute to the targets [53].

PART4

METHODOLOGY

Proposed method as in Figure 4.1 clarifying all steps from starting to the end including preparing data and testing then applying ensemble method and get the final model.

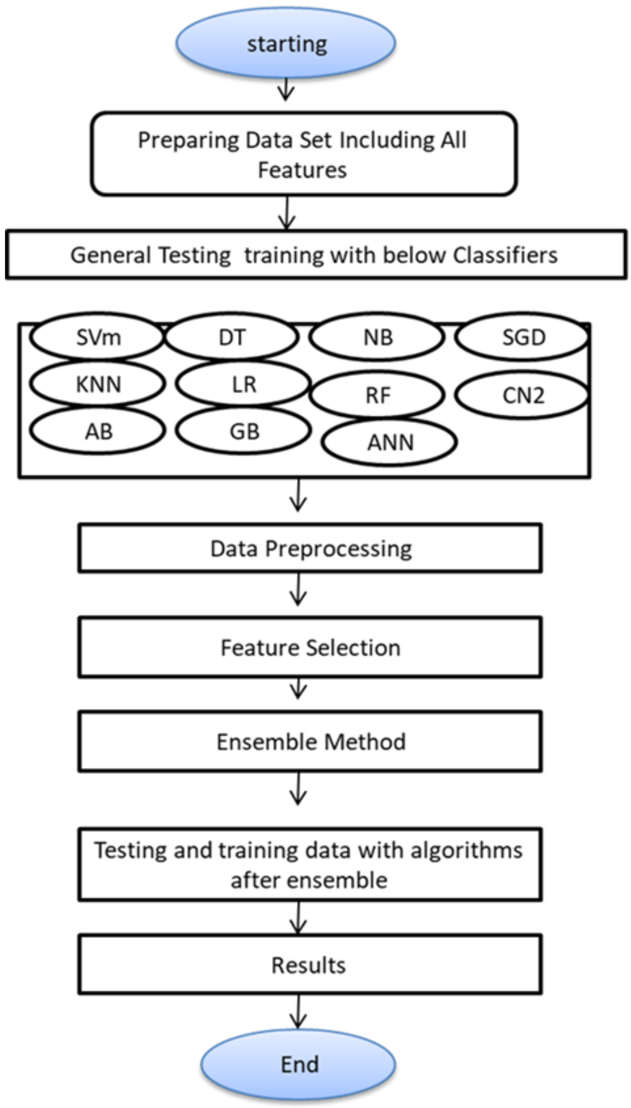

Figure 4.1. Flow chart of the methodology.

4.1. GENERAL TESTING

The original Student Performance Dataset consists of 480 samples with 16 features [54]. The original class contain three classes H (High), L (Low) and M (Medium). Our first step was testing different machine learning classifiers on original datasets for this implementation. We used orange 3.32 and tested all the following classifiers in Table 4.1

Table 4.1 Classifiers names in general testing.

Nu.	Classifiers Names
1	Logistics Regression
2	KNN
3	Neural Network
4	Gradient boosting
5	Ada boost
6	Random Forest
7	Naïve Bayes
8	Cn2 Rule Induction
9	Svm
10	Stochastic Gradient Descent
11	Decision Tree

4.2. DATASET EXPLAINATION

These datasets will be used as a reference point to get insight into the students' attitudes, motivations, and associated traits, with the ultimate goal of determining what causes a student to enroll and what kinds of outcomes can be expected. We utilized these datasets because they contain important characteristics for our purposes, such as student age, gender, and enrolment status. This research will investigate the significance of location data for comprehending cultural identifiers. With the use of analysis investigation, we construct a model to foretell how my students would act or perform. This data used in [7] [55] and taken from Kaggle [54]. The information, which spans 16 columns, is compiled from sources covering 480 students.

Table 4.2. Description of dataset features.

Nu.	Attributes	Description
1	**Gender**	Nominative: "Male" or "Female".
2	**Nationality**	Student's country of origin) nominal: 'Kuwait,' 'Lebanon,' 'Egypt,' 'Saudi Arabia,' 'USA,' 'Jordan,' 'Venezuela,' 'Iran,' 'Tunis,' 'Morocco,' 'Syria,' 'Palestine,' 'Iraq',
3	**Place of birth**	Student's country of origin (nominal): Kuwait; Lebanon; Egypt; Saudi Arabia; the United States of America; Jordan; Venezuela; Iran; Tunis; Morocco; Syria; Palestine; Iraq; and Libya
4	**Educational Stages**	Student's educational stage (nominal: "lower level," "middle school," or "high school).
5	**Grade Levels**	Grade Levels: the level of education to which the student currently belongs (nominal: 'G-01,' 'G-02,' 'G-03,' 'G-04,' 'G-05,' 'G-06,' 'G-07,' 'G-08,' 'G-09,' 'G-10,' 'G-11,' 'G-1').
6	**Section ID**	Class Section ID ('A', 'B', or 'C'): Where each student is placed in the classroom
7	**Topic**	Topic - the focus of a given course (often "English," "Spanish," "French," "Arabic," "Information Technology," "Math," "Chemistry," "Biology," "Science," "The Holy Quran," and "Geology).
8	**Semester**	Each academic year is divided into two semesters (usually referred to as "First" and "Second).
9	**Parent responsible for student**	A parent (mother or father) of student.
10	**Raised hand-**	How often a student raises his or her hand in class is referred to as the "raised hand" statistic (numeric:0-100).
11	**Visited resources**	How often a student accesses specific materials in a course (numeric: 0-100).
12	**Viewing announcements**	The number of times which a student views the latest announcements (numeric: 0-100).
13	**Discussion groups**	Student participation in discussion groups (numeric: 0-100)
14	**Parent Answering Survey**	Parent whom Answering Survey-parent replied the surveys which are offered from school or not (nominal: “Yes, No")
15	**Parent School Satisfaction**	Level of the parental contentment with their child's education (binary: "Yes" or "No")
16	**Student Absence Days**	How many days each student was absent from school (nominal: above-7, under-7)

4.3. DATAPRE-PROCESSING

Pre-processing plays a vital role in data science, from data mining to machine learning. Given the inherent in consistency, noise, and potential absence/redundancy and irrelevance of real-world data. It can lead to erroneously learned information and a decrease in algorithm performance. Pre-processing is used to prepare the data to be processed by the algorithms by making any necessary corrections, such as resizing it [56]. Additionally, feature selection is used to pick the top features, as shown in Figure 4.2

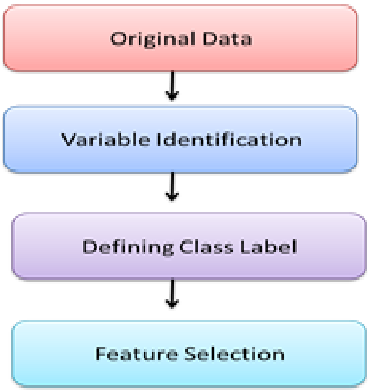

Figure 4.2. Data pre-processing phases.

4.4. FEATURE SELECTION

Using feature selection methods in machine learning is to locate the optimal collection of characteristics from which to build effective models. To discover which input variables are most strongly correlated with the outcome variable, we must first assess the strength of the correlation between each input variable and the outcome variable according to a set of criteria. Decision accuracy, dataset size, and training time can all be improved through feature selection. The four most common ways for choosing which features to use are the filter method, the wrapper method, the embedding method, and the hybrid method [57]. There are a plethora of methods for ranking features, metrics used for evaluating features, like information gain and gain ratio. To determine which features are most relevant when developing a model of students' performance, utilizing a filter-based approach to analysis. We used selection algorithms based on the gain ratio. Figure 4.3 illustrates the feature selection process of data.

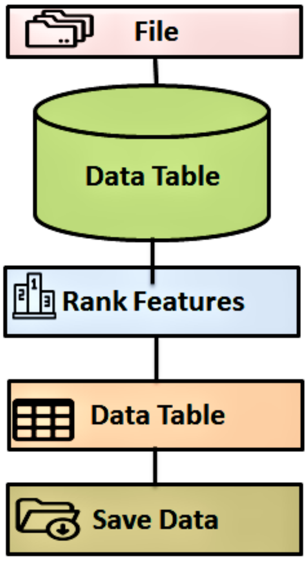

Figure 4.3. Feature selection of data.

4.5. ENSEMBLE METHOD

We used Orange stacking learning and Sklearn's python ensemble voting classifiers to see if we could get better results. Stacking is a method that allows for the use of a number of distinct models to be combined in order to boost prediction accuracy. Rather than picking one model from many, stacking learns from all of them. This method of combining different types of classifiers into one run is referred to as a "stacked generalization" [42]. Stacking offers a novel idea of ensemble learning in comparison to bagging and boosting. Using numerous different classifiers during training and then creating a meta-learner from the results [58].

Learning classifiers vote using the majority rule (for classification) or average rule (for regression). Ultimately, we can estimate the most votes any one category will receive on average. Using a classification technique for prediction, the "Vote" operator reads a sample dataset from the input node and produces a classification model. Each classifier used in the "Vote" operator contributes a vote, and the results are averaged to form the final forecast [59].

4.6. BOOSTING

In machine learning, "boosting" is a popular ensemble strategy. Boosting is created by training multiple learning models [60]. Boosting methods can be used in with other machine algorithms [61]. The RM tool's implementation of AdaBoost is a meta-algorithm because it allows for the execution of the process by adding another algorithm as a sub-process. Multiple models are run and trained in order to produce a single robust learner by integrating weak learners; this process requires additional processing power and time to complete [62]. AdaBoost use on existing educational datasets from earlier research demonstrates the value of boosting methods.

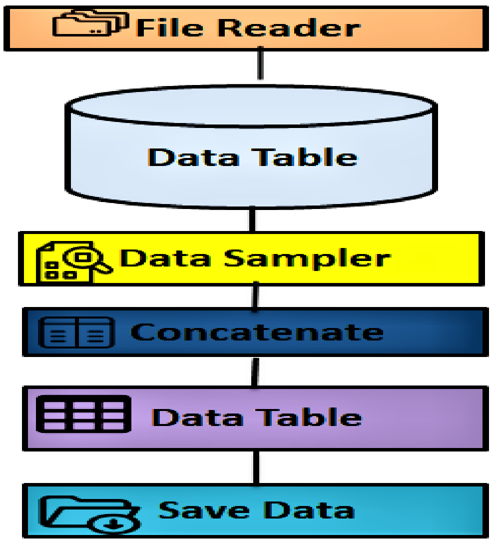

Figure 4.4. Data bootstrapping mechanism.

AdaBoost, an ensemble method that integrates separate classification algorithms, is utilized to hone the study's final model of classification accuracy. AdaBoost was chosen primarily to show how improved over non-boosted models the results of the decision-making process are. In the analysis and discussion, we discuss the model's overall performance and analysis.

The following are the most fundamental criteria used to assess these metrics:

- The number of cases that were correctly detected, also known as "true positives" (TP).
- Incorrectly recognized cases are referred to as "false positives" (FP).
- The number of cases that were properly rejected; also known as "True Negatives" (TN).
- Incorrectly rejected cases, often known as "false negatives" (FN).

The performance measures are detailed in equations 4.1 to 4.3 as below:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \tag{4.1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{4.2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{4.3}$$

PART 5

RESULTS & DISCUSSION

5.1. RESULTS

The result of the initial classification is shown with classifiers in Table 5.1 The classifiers ranked from best to worst classifiers to find out which classifiers is the has high accuracy before building the models in Table 5.1 Explaining the accuracy of classification or describing the corrected categorization of objects (dataset), is a critical part of every investigation (Confusion matrix, accuracy, precision, recall, f1-score, ...etc.). After the general testing of classifiers, we checked the data to expose all details of data like roc and confusion matrix as in Figure 5.1.

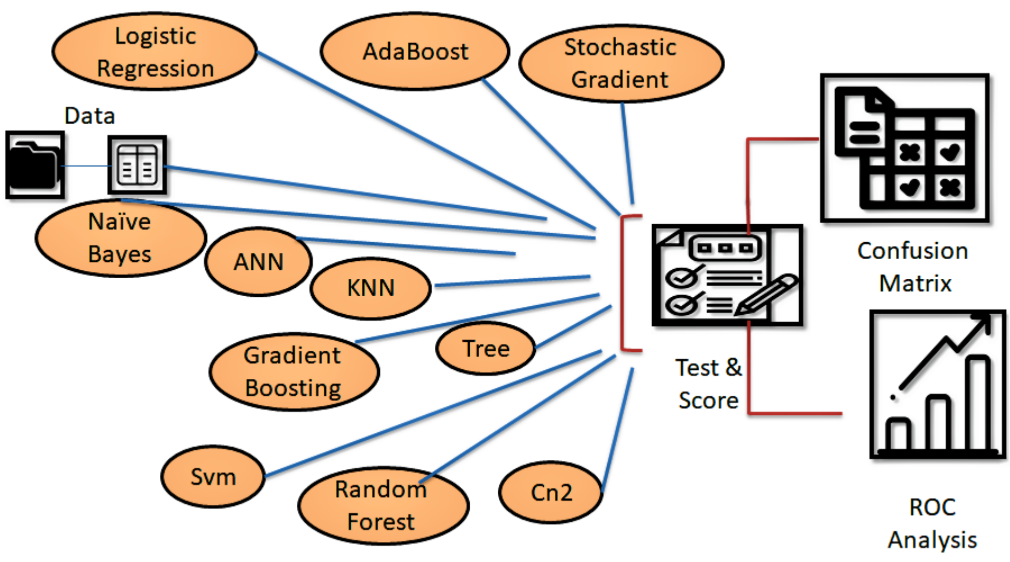

Figure 5.1. General testing of data with classifiers.

Table 5.1. Auc and CA percentage before feature selection and bootstrap.

Model Name/Measure	AUC	CA	F1	Precision	Recall
Random Forest	0.923	0.806	0.807	0.808	0.806
Neural Network	0.902	0.779	0.779	0.780	0.779
Gradient Boosting	0.895	0.779	0.779	0.779	0.779
AdaBoost	0.814	0.758	0.758	0.759	0.758
Decision Tree	0.803	0.721	0.721	0.723	0.721
Naïve Bayes	0.871	0.710	0.706	0.711	0.710
SGD	0.772	0.700	0.698	0.698	0.700
CN2 Rule Inducer	0.752	0.650	0.650	0.650	0.650
Logistic Regression	0.819	0.637	0.635	0.635	0.637
KNN	0.787	0.625	0.620	0.620	0.625
SVM	0.792	0.604	0.604	0.609	0.604

It is possible to visualize how well a classification model performs across a range of cutoff points using a receiver operating characteristic (ROC) curve. TPR and FPR are plotted against classification thresholds on a ROC curve. By decreasing the threshold for positive categorization, more data will be labeled as positive, increasing the number of False Positives and True Positives. A typical ROC curve is depicted in the next image. The ROC of the data is shown below. Each classifier has different color as in Figure 5.2.

Figure 5.2. The colors of each classifier.

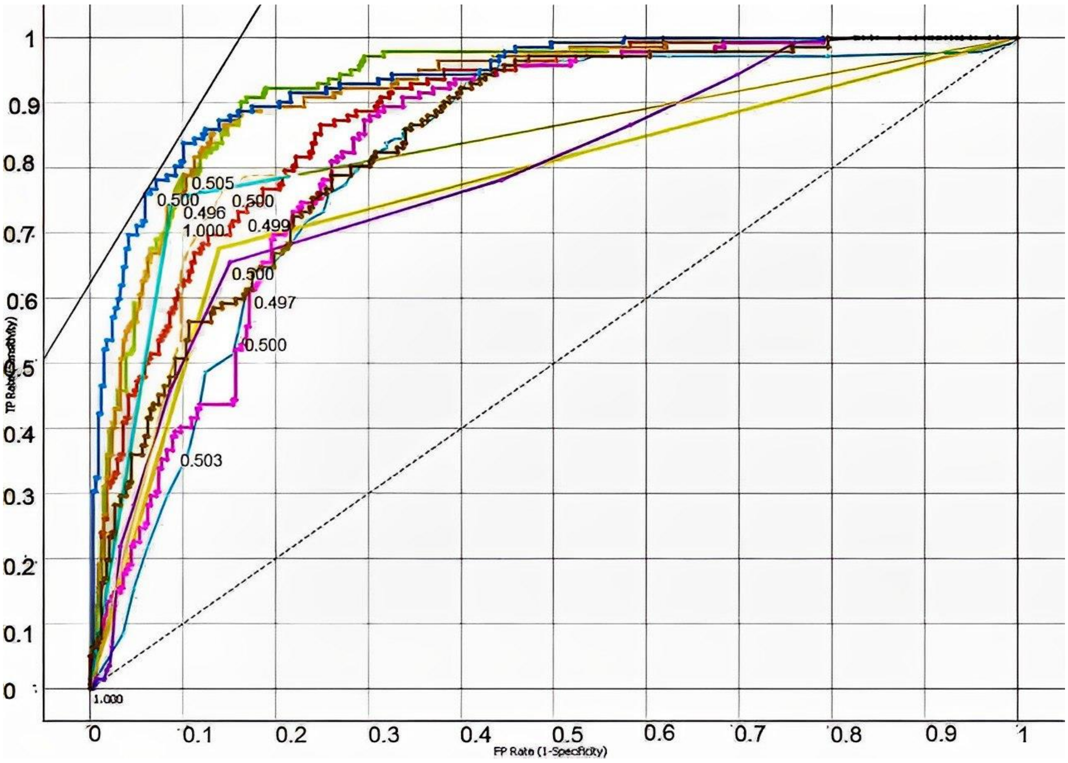

Figure 5.3. Roc analysis of high class before feature selection and bootstrap.

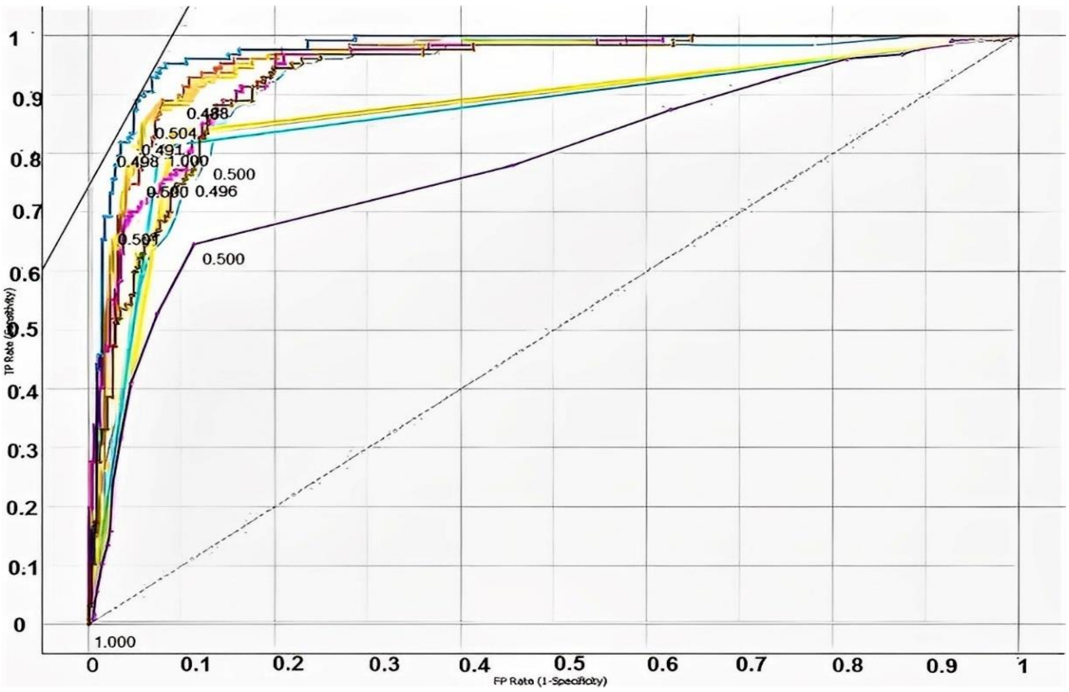

Figure 5.4. Roc analysis of low class before feature selection and bootstrap.

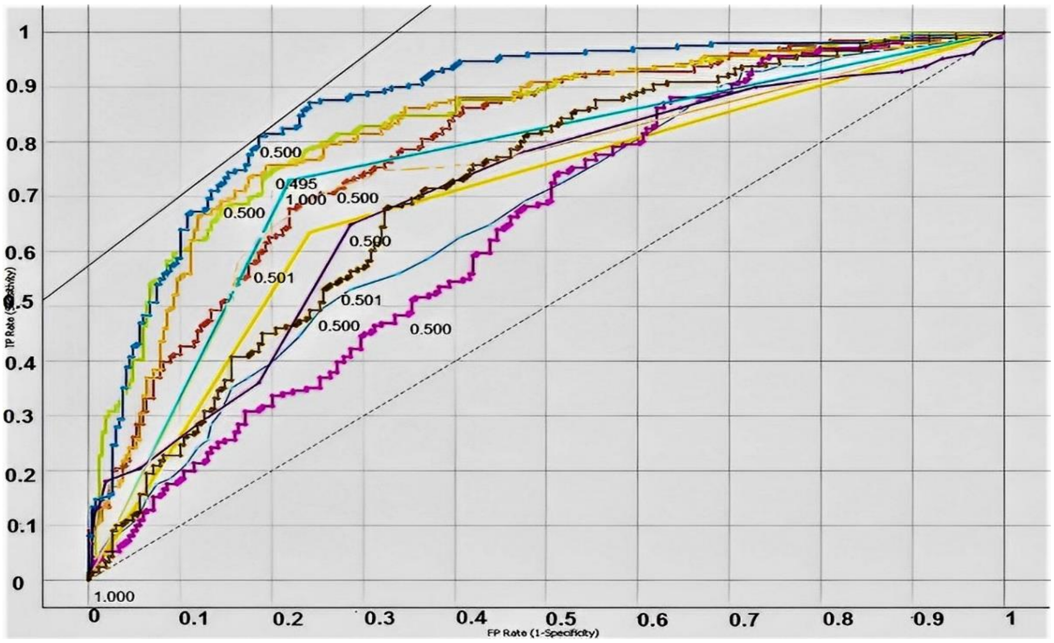

Figure 5.5. Roc analysis of medium class before feature selection and bootstrap.

As we see in Figure 5.4, the Roc plot shows the performance of each within classifiers ranges between (60%-80%) and the (AUC) area under the curve between (72%-92%).

Table 5.2. Confusion matrix before feature selection and bootstrap.

Classifier Name	Summation	Actual H Class	Actual L Class	Actual M Class	Total Predicted H Class	Total Predicted L Class	Total Predicted M Class
AdaBoost	480	142	127	211	137	130	213
KNN	480	142	127	211	142	151	187
ANN	480	142	127	211	144	124	212
Cn2	480	142	127	211	144	122	214
Logistic. R	480	142	127	211	147	139	194
Random. f	480	142	127	211	127	123	230
SVM	480	142	127	211	172	121	187
G.B	480	142	127	211	139	131	210
NB	480	142	127	211	161	145	174
SGD	480	142	127	211	139	136	205
TREE	480	142	127	211	158	120	202

Before feature selection, we measured and examined the dispersion of target labels in the output distribution, Which labels appear and how often. According to the data, goal M ranks highest. Thus, this is an unbalanced data collection.

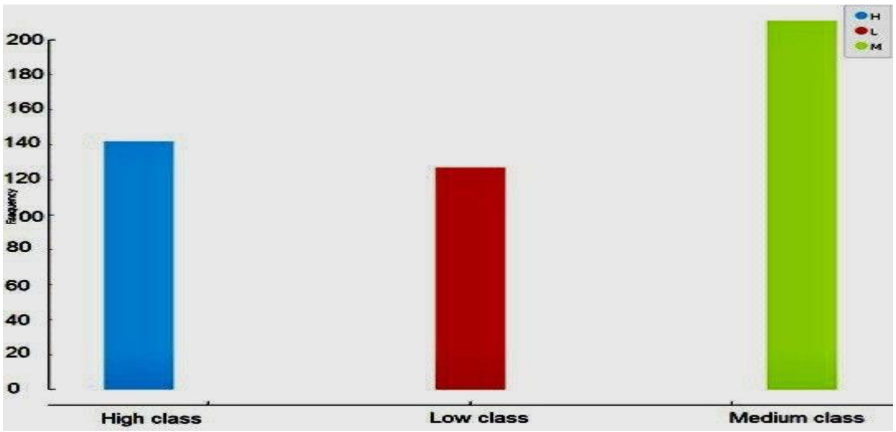

Figure 5.6. Distribution of data classes.

The aim of feature selection and bootstrap is to enhance and improve data correlation in the original dataset to remark weaknesses and increase accuracy. We can see correlations and properties after feature selection and bootstrap. We can see the original features in Table 5.3 and selected features in Table 5.4.

Table 5.3. Correlation between features before selection.

Feature	Info. Gain	Gain Ratio	Gini	Relief
Student Absence Days	0.397	0.410	0.131	0.312
Resources Visited by a Student	0.391	0.195	0.145	0.148
Raising Hands	0.362	0.181	0.139	0.135
Assignments Viewed by Students	0.253	0.127	0.098	0.076
Parent Answering Survey	0.150	0.152	0.055	0.138
Nationality	0.128	0.052	0.045	0.053
Relation	0.126	0.129	0.049	0.080
Place of Birth	0.123	0.051	0.046	0.048
Satisfaction of Parents	0.107	0.111	0.040	0.091
Group Discussion	0.088	0.044	0.038	0.058
Topic / Course	0.076	0.023	0.030	0.060
Gender	0.052	0.055	0.019	0.090

Grade ID	0.047	0.019	0.019	0.045
Semester	0.012	0.012	0.005	-0.016
Stage ID	0.011	0.008	0.005	0.006
Section ID	0.007	0.006	0.003	-0.004

Table 5.4. Features after selection.

Feature	Info. Gain	Gain Ratio	Gini	Relief
Student Absence Days	0.397	0.410	0.131	0.326
Resources visited by a student	0.391	0.195	0.145	0.179
Raising Hands	0.362	0.181	0.139	0.108
Assignments Viewed by Students	0.253	0.127	0.098	0.061
Parent Answering Survey	0.150	0.152	0.055	0.083
Relation	0.126	0.129	0.049	0.111
Place of Birth	0.123	0.051	0.046	0.054
Satisfaction of Parents	0.107	0.111	0.040	0.068
Group Discussion	0.088	0.044	0.038	0.054
Topic / Course	0.076	0.023	0.030	0.129
Gender	0.052	0.055	0.019	0.085
Grade ID	0.047	0.019	0.019	0.156

Table 5.5. Auc and CA percentage after feature selection and bootstrap.

Model Name/Measure	AUC	CA	F1	Precision	Recall
Random Forest	0.993	0.958	0.958	0.958	0.958
Neural Network	0.989	0.949	0.949	0.950	0.949
AdaBoost	0.949	0.934	0.934	0.935	0.934
Gradient Boosting	0.984	0.931	0931	0.932	0.931
CN2 Rule Inducer	0.935	0.903	0.903	0.904	0.903
Tree	0.940	0.908	0.907	0.908	0.908
SGD	0.863	0.821	0.820	0.820	0.821
Naïve Bayes	0.912	0.765	0.764	0.766	0.765
SVM	0.877	0.697	0.697	0.704	0.697
Logistic Regression	0.844	0.684	0.681	0.680	0.684
KNN	0.822	0.659	0.649	0.655	0.659

We can see clearly that percentages were increased by bootstrapping and became better than before if we compare it with before bootstrapping, especially AUC and CA. After that, we performed the confusion matrix plot as Table 5.6.

Table 5.6. Confusion matrix after feature selection and bootstrap.

Classifier name	**Summation**	**Actual H class**	**Actual L class**	**Actual M Class**	**Total Predicted H class**	**Total Predicted L Class**	**Total Predicted M Class**
AdaBoost	671	179	199	293	174	203	294
KNN	671	179	199	293	117	218	336
ANN	671	179	199	293	187	203	281
Cn2	671	179	199	293	173	196	302
Logistic. R	671	179	199	293	176	218	277
Random. F	671	179	199	293	179	199	293
SVM	671	179	199	293	220	199	252
G.B	671	179	199	293	171	196	304
NB	671	179	199	293	198	203	270
SGD	671	179	199	293	173	211	287
TREE	671	179	199	293	179	213	279

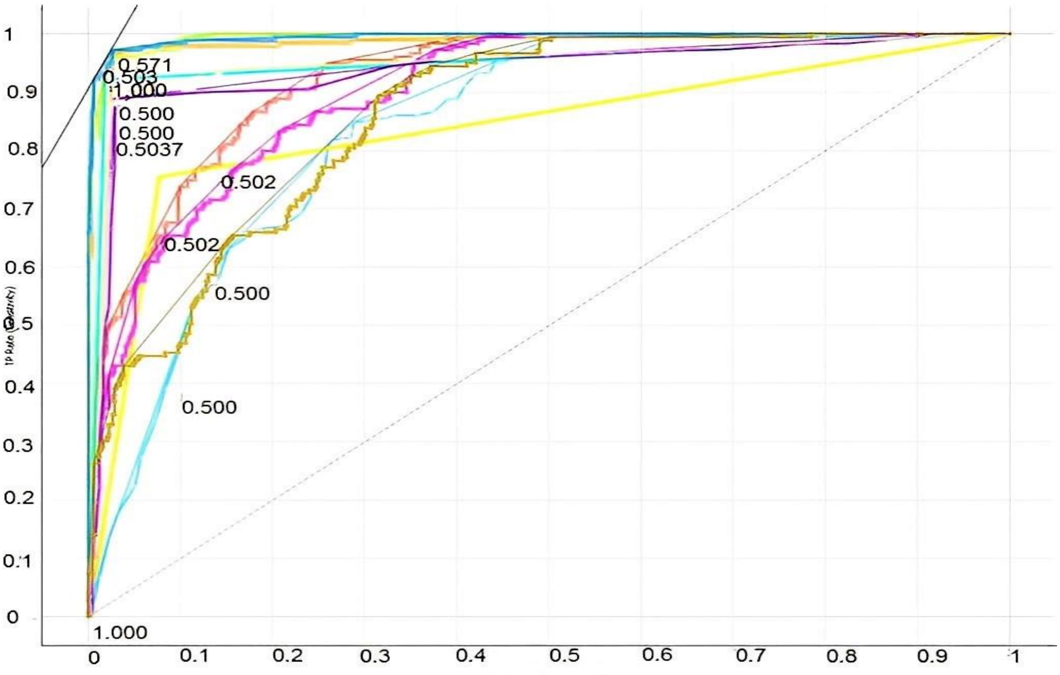

Figure 5.7. Roc of High class after feature selection and bootstrap.

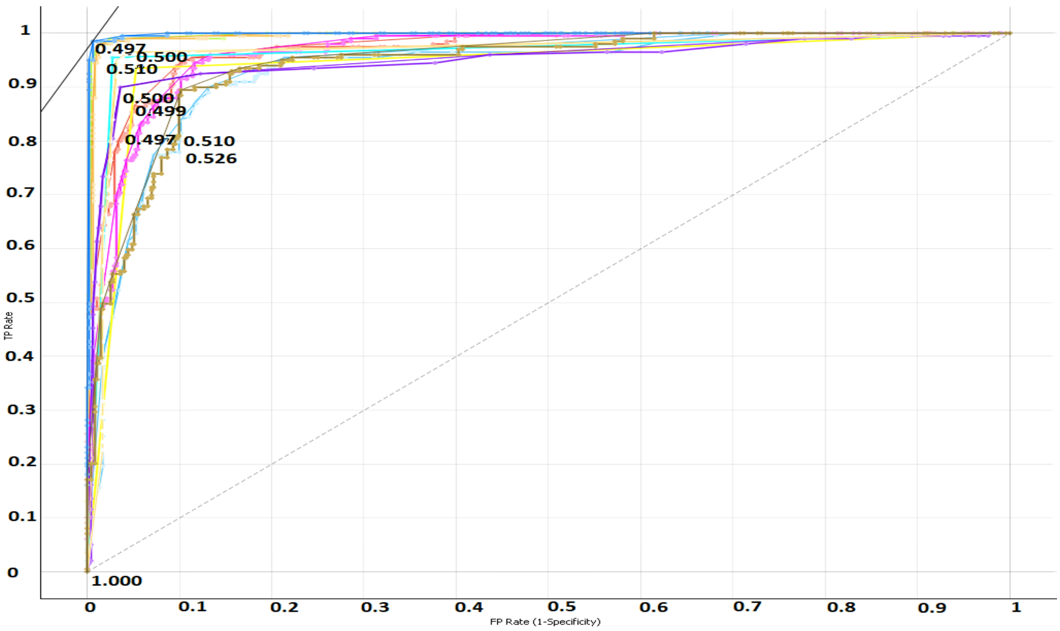

Figure 5.8. Roc analysis of low class after feature selection and bootstrap.

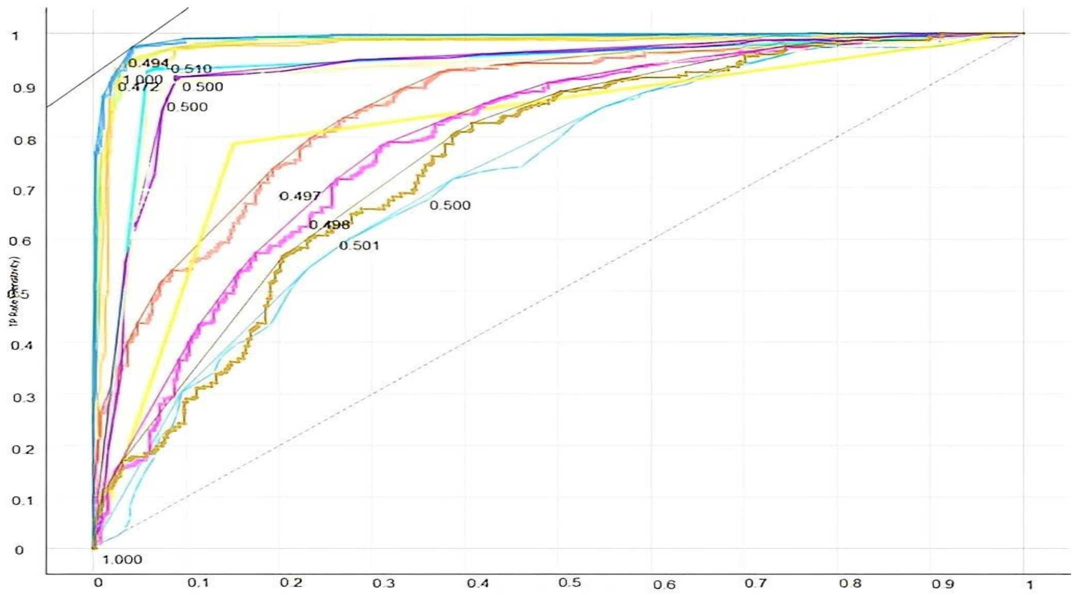

Figure 5.9. Roc analysis of medium class after feature selection and bootstrap.

As we see, the feature selection and bootstrap helped us refine data and get better results than data before enhancement, as in Table 5.5 It is clear that the Auc percentage has increased. For example, random forest has reached 99% in table of classifiers and, AdaBoost has reached 94%; then the Roc figures showed that the curve is near one value which is a very good result.

After the general testing of classifiers and feature selection and bootstrap, the final step was to implement the final model by taking the two higher classifiers after feature selection and bootstrap, which are AdaBoost and random forest (AdaBoost accuracy 93% and the area under curve 94%, random, forest classifier accuracy is 95% and area and curve is 99%) by applying the voting to get the best result of prediction the accuracy of final model after voting is 98% as shown in Figure 5.10

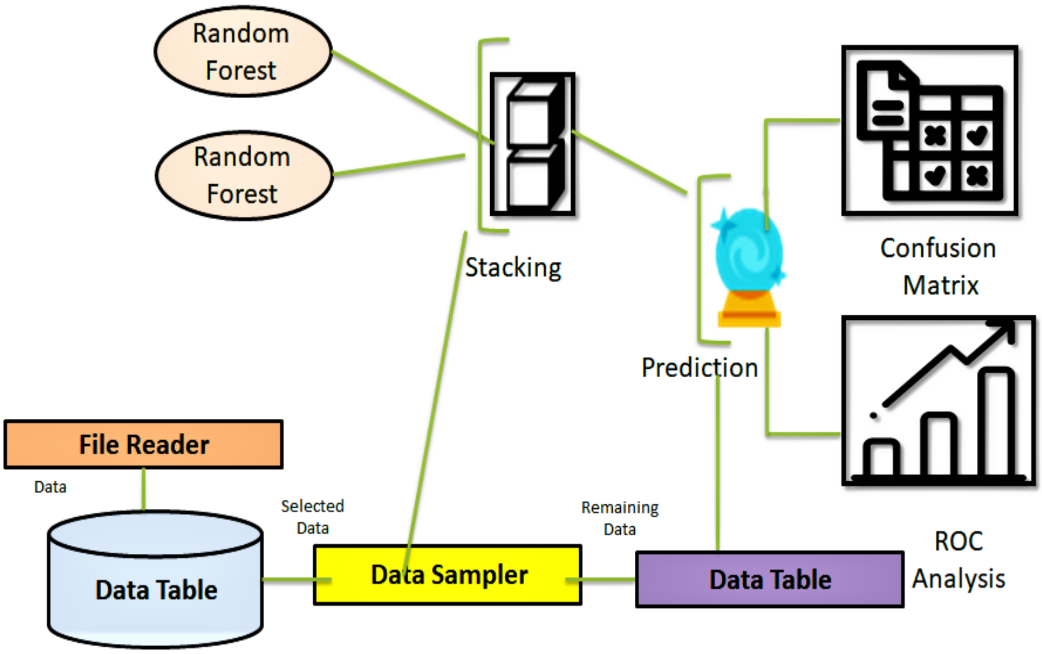

Figure 5.10. Final model.

Table 5.7. Confusion matrix of final model.

Summation	Predicted H class	Predicted L Class	Predicted M Class
Actual H Class =45	45	0	0
Actual L Class =51	0	51	0
Actual M Class =65	3	0	62
Total=161	Total=48	Total=51	Total=62

As the confusion matrix demonstrates, the model can distinguish and observe the students based on the classes very well to detect and estimate students' performance.

Table 5.8. Accuracy of final model.

Class	Gender	Place of Birth	Grade Id	Topic
H	F	Egypt	G-07	Quran
H	F	Egypt	G-07	Quran
H	F	Egypt	G-07	Quran
H	F	Egypt	G-07	Quran
M	M	Egypt	G-04	Math
L	M	Egypt	G-02	French
M	M	Iran	G-09	IT
M	M	Iran	G-09	IT
H	F	Iraq	G-07	Biology
H	F	Iraq	G-02	Arabic
H	F	Iraq	G-02	Arabic
M	M	Iraq	G-08	Geology
M	M	Iraq	G-08	Geology
M	M	Iraq	G-08	History
H	M	Iraq	G-07	Biology

Performance Scores					
Model	**AUC**	**CA**	**F1**	**Precision**	**Recall**
Stack	0.999	0.981	0.981	0.983	0.981

As we can see, the model's accuracy reached 98% percent, which is very good by using the model with two classifiers, AdaBoost and Random Forest, with a precision of 98.3% and recall of 98.1%.

5.2. DISCUSSION

In the bagging set of classifiers, we used (KNN, NB, SVM, NN, Cn2, SGD, LR, RF, DT, B, and GB) to rank each classifier before the classifier without enhancement. In a single experiment, we could train and verify all models. We also used the stacking and voting procedure to combine ML models and evaluate the success of the analysis. In the preceding paragraph, we looked individually at each performance indicator. In conclusion, this study's findings highlight the potential efficacy of the ensemble approach in enhancing prediction accuracy and suggest more investigation into this avenue.

Both the fitting and generalization problems that the suggested stacking model has solved plague ML models. The best classifier, which was employed in this work, was discovered by integrating ML models. To increase performance, the authors of this work advise using ML classifiers, or "stacking," rather than a single classifier. Overall, the study's findings demonstrated that the proposed stacking model's reliability and implementation could help predict the student's success in the DL system.

PART 6

CONCLUSION

6.1. CONCLUSION

In most countries, the highest priority for secondary education is improving students' academic performance. Learner systems produce massive amounts of data. These records contain previously unknown information that could significantly improve pupils' performance in the classroom. A model for forecasting student achievement is proposed in this work that relies solely on ensemble techniques. The methods (bagging and boosting) focus on improving the predictive model's classifiers (such as an artificial neural network, decision tree, or naive Bayesian).

The findings revealed to show that these models are superior to traditional classifiers. The proposed method then utilizes bagging or boosting to merge two distinct classifiers. Compared to other approaches, this one proved more effective in raising students' and schools' levels of achievement. Using advanced data mining techniques, we will compile data from a wide range of students from various educational institutions to provide significant results. Learning management systems, pedagogical foundations, students, and teachers can all benefit from this initiative's focus on enhancing performance.

Additional data sets will be used in subsequent projects with these models. These findings validate the authenticity of the predictive models, especially when compared to the many well-established good classifiers. Last but not least, these models can aid educators in grasping students, pinpointing areas for improvement, fostering diverse pedagogical approaches, and reducing the prevalence of academic attrition. Improvements in instructional strategies can also help principals.

6.2. FUTURE WORK

An issue of great importance is the ability to anticipate a student's future performance. Deep research led us to the conclusion that different student datasets yield varied findings with distinct features. As a result, different evaluation metrics, such as accuracy, precision, and geometric mean, yield varying findings. As a result of these investigations, we have concluded that the outcomes of each approach and algorithm depend on the dataset and variable attribute utilized to make the prediction.

However, if we utilize machine learning algorithms, we can get more detailed findings for future forecasts and contribute to improving the educational system. This way, we can increase our educational system's prediction methodologies and performance.

REFERENCES

1. Ashraf, A., Anwer, S., and Khan, M. G., "A Comparative study of predicting student's performance by use of data mining techniques", ***American Scientific Research Journal For Engineering, Technology, And Sciences (ASRJETS)***, 44 (1): 122–136 (2018).

2. Evangelista, E., "A hybrid machine learning framework for predicting students' performance in virtual learning environment", ***International Journal Of Emerging Technologies In Learning (IJET)***, 16 (24): 255–272 (2021).

3. Kabakchieva, D., "Predicting student performance by using data mining methods for classification", ***Cybernetics And Information Technologies***, 13 (1): 61–72 (2013).

4. Parveen, Z. and Quadri, M. P., "Classification and Prediction of Student Academic Performance using Machine Learning A Review", ***International Journal Of Computer Sciences And Engineering***, 7 (3): 607–614 (2019).

5. Ahmad, N., Hassan, N., Jaafar, H., and Enzai, N. I. M., "Students' Performance Prediction using Artificial Neural Network", ***IOP Conference Series: Materials Science And Engineering***, 1176 (1): 012020 (2021).

6. Bedregal-alpaca, N., Cornejo-aparicio, V., Zárate-valderrama, J., and Yanque-churo, P., "Classification-models-for-determining-types-of-academic-risk-and-predicting-dropout-in-university-students2020International-Journal-of-Advanced-Computer-Science-and-Applications.pdf", 11 (1): 266–272 (2020).

7. Amrieh, E. A., Hamtini, T., and Aljarah, I., "Mining educational data to predict student's academic performance using ensemble methods", ***International Journal Of Database Theory And Application***, 9 (8): 119–136 (2016).

8. Dhilipan, J., Vijayalakshmi, N., Suriya, S., and Christopher, A., "Prediction of Students Performance using Machine learning", ***IOP Conference Series: Materials Science And Engineering***, 1055 (1): 012122 (2021).

9. Sagi, O. and Rokach, L., "Ensemble learning: A survey", ***Wiley Interdisciplinary Reviews: Data Mining And Knowledge Discovery***, 8 (4): 1–18 (2018).

10. Mohamad, S. K. and Tasir, Z., "Educational data mining: A review", ***Procedia-Social And Behavioral Sciences***, 97: 320–324 (2013).

11. Witten, I. H. and Frank, E., "Data mining: practical machine learning tools and techniques with Java implementations", ***Acm Sigmod Record***, 31 (1): 76–77

(2002).

12. Agarwal, S., "Data mining: Data mining concepts and techniques", (2013).

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

14. Müller, A. C. and Guido, S., "Introduction to Machine Learning with Python: A Guide for Data Scientists", ***" O'Reilly Media, Inc.,"*** (2016).

15. Ghayyur, S., Ghosh, D., He, X., and Mehrotra, S., "MIDE: accuracy aware minimally invasive data exploration for decision support", ***Proceedings Of The VLDB Endowment***, 15 (11): 2653–2665 (2022).

16. Thaher, T. and Jayousi, R., "Prediction of Student’s Academic Performance using Feedforward Neural Network Augmented with Stochastic Trainers", ***14th IEEE International Conference On Application Of Information And Communication Technologies, AICT 2020 - Proceedings***, (2020).

17. Pujianto, U., Agung Prasetyo, W., and Rakhmat Taufani, A., "Students Academic Performance Prediction with k-Nearest Neighbor and C4.5 on SMOTE-balanced data", ***2020 3rd International Seminar On Research Of Information Technology And Intelligent Systems, ISRITI 2020***, 348–353 (2020).

18. Vijayalakshmi, V. and Venkatachalapathy, K., "Deep neural network for multi-class prediction of student performance in educational data", ***International Journal Of Recent Technology And Engineering***, 8 (2): 5073–5081 (2019).

19. Almutairi, S., Shaiba, H., and Bezbradica, M., "Predicting Students’ Academic Performance and Main Behavioral Features Using Data Mining Techniques", ***Communications In Computer And Information Science***, 1097 CCIS: 245–259 (2019).

20. Ajibade, S. S. M., Dayupay, J., Ngo-Hoang, D. L., and ..., "Utilization of Ensemble Techniques for Prediction of the Academic Performance of Students", ***Journal Of …***, (June): (2022).

21. Ajibade, S.-S. M., Ahmad, N. B., and Shamsuddin, S. M., "A data mining approach to predict academic performance of students using ensemble techniques", (2020).

22. Saleem, F., Ullah, Z., Fakieh, B., and Kateb, F., "Intelligent decision support system for predicting student’s e-learning performance using ensemble machine learning", ***Mathematics***, 9 (17): 1–22 (2021).

23. Ragab, M., Abdel Aal, A. M. K., Jifri, A. O., and Omran, N. F., "Enhancement of Predicting Students Performance Model Using Ensemble Approaches and Educational Data Mining Techniques", ***Wireless Communications And Mobile Computing***, 2021: (2021).

24. Mansoor Ali, R. S. and Perumal, S., "Multi-class LDA classifier and CNN feature extraction for student performance analysis during Covid-19 pandemic", ***Int. J. Nonlinear Anal. Appl***, 13 (August 2021): 2008–6822 (2022).

25. Tuaha, S., Siddiqui, I. F., and Ali Arain, Q., "Analyzing Students’ Academic Performance through Educational Data Mining", ***3C Tecnología_Glosas De Innovación Aplicadas A La Pyme***, (May): 402–421 (2019).

26. Hasibur Rahman, M. and Rabiul Islam, M., "Predict Student’s Academic Performance and Evaluate the Impact of Different Attributes on the

Performance Using Data Mining Techniques", ***2nd International Conference On Electrical And Electronic Engineering, ICEEE 2017***, (September 2020): 1–4 (2018).

27. BÜTÜNER, R. and CALP, M. H., "Estimation of the Academic Performance of Students in Distance Education Using Data Mining Methods", ***International Journal Of Assessment Tools In Education***, 9 (2): 410–429 (2022).

28. S. Abubakari, M., Arifin, F., and G. Hungilo, G., "Predicting Students' Academic Performance in Educational Data Mining Based on Deep Learning Using TensorFlow", ***International Journal Of Education And Management Engineering***, 10 (6): 27–33 (2020).

29. Yıldız, T. K., Yurtay, N., and Öneç, B., "Classifying anemia types using artificial learning methods", ***Engineering Science And Technology, An International Journal***, 24 (1): 50–70 (2021).

30. Ritonga, M., Ihsan, M. A. Al, Anjar, A., and Rambe, F. H., "Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm", (2021).

31. Cui, S., Yin, Y., Wang, D., Li, Z., and Wang, Y., "A stacking-based ensemble learning method for earthquake casualty prediction", ***Applied Soft Computing***, 101: 107038 (2021).

32. Nagy, M. and Molontay, R., "Predicting dropout in higher education based on secondary school performance", (2018).

33. Charbuty, B. and Abdulazeez, A., "Classification based on decision tree algorithm for machine learning", ***Journal Of Applied Science And Technology Trends***, 2 (01): 20–28 (2021).

34. Jordan, M. I. and Mitchell, T. M., "Machine learning: Trends, perspectives, and prospects", ***Science***, 349 (6245): 255–260 (2015).

35. Devasia, T., Vinushree, T. P., and Hegde, V., "Prediction of students performance using Educational Data Mining", (2016).

36. Simeone, O., "A brief introduction to machine learning for engineers", ***Foundations And Trends® In Signal Processing***, 12 (3–4): 200–431 (2018).

37. Milliken, J. A., Wartak, J., Lywood, D. W., Fay, J. E., Kraus, A., Orme, W., and Mehta, S., "Validity of computer interpretation of the electrocardiogram", ***Canadian Medical Association Journal***, 105 (11): 1147 (1971).

38. Liang, J., Yang, J., Wu, Y., Li, C., and Zheng, L., "Big data application in education: dropout prediction in edx MOOCs", (2016).

39. SAIHOOD, Q. L., .

40. Delen, D., "A comparative analysis of machine learning techniques for student

retention management", ***Decision Support Systems***, 49 (4): 498–506 (2010).

41. Khan, M. I., Khan, Z. A., Imran, A., Khan, A. H., and Ahmed, S., "Student Performance Prediction in Secondary School Education Using Machine Learning", (2022).

42. Kiangala, S. K. and Wang, Z., "An effective adaptive customization framework for small manufacturing plants using extreme gradient boosting-XGBoost and random forest ensemble learning algorithms in an Industry 4.0 environment", ***Machine Learning With Applications***, 4: 100024 (2021).

43. Autonomous, S. C., "Misbehavior Detection in C-ITS", ***Springer International Publishing***, 641–652 (2019).

44. Almasri, A., Celebi, E., and Alkhawaldeh, R. S., "EMT: Ensemble meta-based tree model for predicting student performance", ***Scientific Programming***, 2019: (2019).

45. Verma, A., "A Comparative Study of Ensemble Learning", ***2017 7th International Conference On Cloud Computing, Data Science & Engineering - Confluence***, 155–158 (2017).

46. Chen, J., "WORKSHOP 2-INTRODUCTORY MACHINE LEARNING: DECISION TREES, RANDOM FORESTS, AND OTHER DENDROLOGICAL METHODS", (2020).

47. Rustam, F., Ashraf, I., Mehmood, A., Ullah, S., and Choi, G. S., "Tweets classification on the base of sentiments for US airline companies", ***Entropy***, 21 (11): 1078 (2019).

48. Tan, Z., Chen, J., Kang, Q., Zhou, M., Abusorrah, A., and Sedraoui, K., "Dynamic embedding projection-gated convolutional neural networks for text classification", ***IEEE Transactions On Neural Networks And Learning Systems***, 33 (3): 973–982 (2021).

49. Pangastuti, S. S., Fithriasari, K., Iriawan, N., and Suryaningtyas, W., "Data mining approach for educational decision support", ***EKSAKTA: Journal Of Sciences And Data Analysis***, 33–44 (2021).

50. Bala, R. and Kumar, D., "Classification Using ANN: A Review", ***International Journal Of Computational Intelligence Research***, 13 (7): 1811–1820 (2017).

51. Li, L., Hu, Q., Wu, X., and Yu, D., "Exploration of classification confidence in ensemble learning", ***Pattern Recognition***, 47 (9): 3120–3131 (2014).

52. Basheer, I. A. and Hajmeer, M., "Artificial neural networks: fundamentals, computing, design, and application", ***Journal Of Microbiological Methods***, 43 (1): 3–31 (2000).

53. Singh, P., "Deploy Machine Learning Models to Production", ***Springer***,

(2021).

54. Internet: Aljarah, I., "No Title", **https://www.kaggle.com/aljarah/xAPI-Edu-Data** .

55. Amrieh, E. A., Hamtini, T., and Aljarah, I., "Preprocessing and analyzing educational data set using X-API for improving student's performance", ***2015 IEEE Jordan Conference On Applied Electrical Engineering And Computing Technologies, AEECT 2015***, (2015).

56. GE, O., "A Machine Learning Based Framework for Predicting Student's Academic Performance", ***Physical Science & Biophysics Journal***, 4 (2): (2020).

57. Sáez, J. A. and Romero-Béjar, J. L., "On the Suitability of Bagging-Based Ensembles with Borderline Label Noise", ***Mathematics***, 10 (11): 1–14 (2022).

58. Borkar, S. and Rajeswari, K., "Predicting students academic performance using education data mining", ***International Journal Of Computer Science And Mobile Computing***, 2 (7): 273–279 (2013).

59. Enughwure, A. A. and Ogbise, M. E., "Application of Machine Learning Methods to Predict Student Performance: A Systematic Literature Review", ***International Research Journal Of Engineering And Technology***, (May): 3405–3415 (2020).

60. Abdulkareem, N. M. and Abdulazeez, A. M., "Machine learning classification based on Radom Forest Algorithm: A review", ***International Journal Of Science And Business***, 5 (2): 128–142 (2021).

61. Kumari, P., Jain, P. K., and Pamula, R., "An efficient use of ensemble methods to predict students academic performance", ***Proceedings Of The 4th IEEE International Conference On Recent Advances In Information Technology, RAIT 2018***, 1–6 (2018).

62. Vultureanu-Albişi, A. and Bădică, C., "Improving students' performance by interpretable explanations using ensemble tree-based approaches", (2021).

RESUME

Abdullah Raed AL-SHAIKHLI he graduated with a bachelor's degree in computer engineering from Al-Iraqi university in 2019. In 2020 he began studying for his master's degree in computer engineering department at Karabuk University in Turkey.