# Students Academic Performance Prediction Using an Ensemble of Machine Learning Models

**Mohamed Farook Ali.N[1], Elamparithi.M[2], Anuratha.V[3]**

[1]Research Scholar, Department Of Computer Science, Kamalam College of Arts and Science, Anthiyur, Affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India, Email: anmfarooq@gmail.com
[2]Associate Professor , Department Of Computer Science, Kamalam College of Arts and Science, Anthiyur, Affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India, Email: profelamparithi@gmail.com.
[3]Associate Professor , Department Of Computer Science, Kamalam College of Arts and Science, Anthiyur, Affiliated to Bharathiar University, Coimbatore, Tamil Nadu, India, Email: profanuratha@gmail.com.

Received: 14.04.2024 Revised : 16.05.2024 Accepted: 28.05.2024

**ABSTRACT**
Instructional statistics mining (EDM) is an emerging subject because of the growth of tutorial records. While fact-mining techniques are used in training, they can discover hidden expertise and patterns that could aid in selection-making tactics to enhance the academic gadget. The techniques extracted from academic statistics Mining subjects are then used to apprehend college students, including their studying behavior and expecting their educational performance. This research used an EDM method to classify and predict overall performance using machine studying (ML) techniques. Pre-processing is achieved using data formatting but lacks records management, facts normalization, and facts filtering. After pre-processing, it is going to be shipped to the classifier for pupil academic performance prediction by the use of ML models like aid vector machine (SVM), Logistic regression (LR), Adaboost, selection tree (DT), and Random woodland (RF). With ML fashions, we use parameter optimization strategies like gradient descent to enhance accuracy. An experimental result demonstrates the efficiency of the projectedclassical in phrases of accuracy. The model Adaboost suggests satisfactory overall performance by achieving an accuracy of 70.39%.

**Keywords:** Machine Learning, Educational data mining (EDM), Data pre-processing, optimization techniques, student academic performances.

## 1. INTRODUCTION

Students' overall academic performance is a critical finding in building their future. This is because of elements: first, it's far necessary to determine which institution of students will perform well on semester-give-up exams so that you can grant scholarships, and 2d, it is vital to decide on which college students may also perform poorly so that remediation can be supplied to them [1]. A spread of things affects college students' academic achievement, together with their preceding instructional records, family records, financial standing, and performance on midterm checks [2].college students can now get entry to schooling in a ramification of ways, such as online schooling systems, web-based training structures, seminars, and workshops. Forecasting is extra tricky due to the massive amount of information stored in mastering management structures and educational databases; forecasting college students' success with schooling acquired through these systems is more challenging [3]. To cope with academic underachievement, instructional institutions must analyze gathered facts and forecast pupil achievement. This can assist in picking out students who need better instructional overall performance early in their studies [4].

Academic records mining (EDM) is one of the methods that instructional organizations can use to find patterns hidden in educational records, increase their expertise, or forecast future scholar success. At the same time, EDM is used to discover expertise from information, and device learning (ML) algorithms offer the equipment. EDM employs various set information features, metrics, and prediction strategies [5]. Predicting students' educational performance entails estimating an unknown score or grade, that's generally performed via using various category and regression techniques such as decision Tree (DT), Random forest (RF), Nave Bayes (NB), support Vector machine (SVM), and so forth. This look additionally addresses the problems related to comparing students' overall performance information. However, that allows you to perform the feature learning process in scholar instructional performance prediction; the ML version calls for human intervention. Factors, noisy information, and biased training sets can cause incorrect predictions [6]. In this paper, we use the ensemble of ML models to predict college students' overall academic performance.

## 2. LITERATURE REVIEW

**Jovana Jovićet.al [4]**It evolved the ML fashions like LR, Linear Discriminant evaluation (LDA), ok-Nearest Neighbor (KNN), DT, NB, and SVM for the prediction of scholar's performance using Belgrade Metropolitan College's educational management gadget (EMS) and studying management machine (LMS) dataset. SVM outperforms other models by achieving an accuracy of ninety-three percent because it is robust in cases where the scale is more significant than the range of samples. But, when the data set incorporates extra noise, i.e., target training overlap, the SVM version does not perform nicely.

**Amnah Saeed Alghamdiet.al [6]**It proposed ML models, which include NB, RF, and J48, for predicting pupil overall performance using an electronic questionnaire dataset. The NB method outperforms other fashions by reaching an accuracy of 99.34% because it considers each function in the samples to be independent of the others, and it has speed in addition to content with fewer schooling examples. The RF algorithm is significantly slower than other class algorithms as it uses a couple of choice timbers to make predictions, and its training time is more significant than other fashions due to its complexity.

**Opeyemi Ojajuniet.al [7]**Using the LMS database advanced the ML classifiers like random wooded area, SVM, gradient boosting, DT, LR, and intense gradient boosting (XGBoost) to predict instructional performance. Compared to different fashions, XGBoost performs better by reaching an accuracy of ninety-seven.12% due to it making minimal errors, converging extra quickly with fewer steps, and simplifying calculations to enhance pace and decrease compute charges. However, it is susceptible to overfitting, specifically when trained on small datasets or when too many bushes are used in the model.

**Girma Asefa and Fikadu Wayesa [8]**It provided ML strategies like k-NN, DT, and NB for student performance prediction using instructional information and widespread scholarly records resources of Wachemo College. Compared to other models, KNN plays better by reaching an accuracy 97.89p. For inner statistics and ninety-nine. 70% for non-educational related records because it can capture complex and nonlinear styles inside the facts. It is robust to noise in addition to outliers if a large enough k is used. However, its accuracy is determined through satisfactory statistics. The prediction stage can be gradual while handling enormous amounts of facts.

**Hussein Altabraweeet.al [9]**ML fashions like ANN, NB, DT, and LR for pupil performance prediction were evolved using the UCI database. ANN outperforms other models, attaining an accuracy of 77.04% due to the fact it's far resistant to noisy datasets and performs fine on categorizing styles that have not remained educated on, consequently it is used in conditions where there's slight expertise in the connection among the magnificence marker and the functions in the dataset. However, it needs more transparency in choice-making and is at risk of overfitting without the right regularisation.

## 3. PROPOSED METHODOLOGY

This segment discusses the proposed student educational performance prediction model. The degrees to detect students' overall instructional performance are facts series, information pre-processing like data formatting, lack of information handling, statistics normalization, statistics filtering, classification, and performance analysis. Figure 1 shows that the complete architecture of the proposed model.

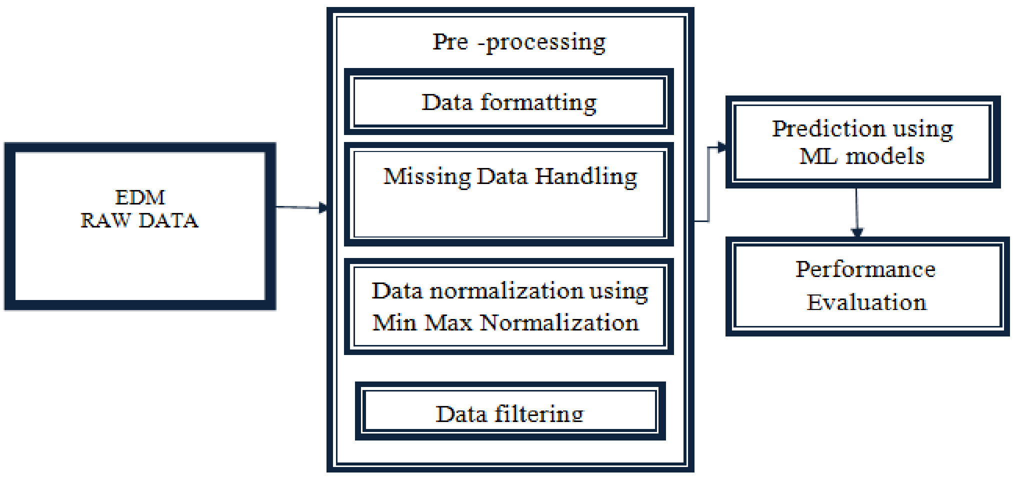


**Figure 1.** Complete architecture of the proposed model

### 4.1. Input Dataset
This work uses a modified instructional dataset from the UCI repository. This dataset incorporates a 230318 range of instances of various Attributes 13.

### 4.2 Data pre-processing
The obtained dataset will subsequently circulate into the statistics pre-processing stage. Records pre-processing is a technique that transforms the shape of obtained facts into an understandable shape. Strategies like facts formatting, missing statistics handling, data normalization, and data filtering are used for pre-processing.

#### 4.2.1 Data cleaning
This involves recognizing and modifying mistakes or irregularities in the information, such as lost values, outliers, and duplicates. Diverse strategies, such as imputation, removal, and transformation, can be used for cleaning statistics.

#### 4.2.2 Missing Data Handling
Missing values may be dealt with by removing the rows or columns worthless values. If additional than partial of the rows are valueless, the complete column may be released. Rows with one or more column values as worthless can also be released.

#### 4.2.3 Min Max Normalization
Normalization is accomplished once we remove the values lacking from the dataset. Because input academic statistics might have scale versions that cause inaccurate outcomes, normalizing the records is required to avoid these problems. This work uses the Min-max Normalization model, and the normalization system includes converting numerical values into a brand-new range with a mathematical characteristic. Minimal standardization is one of the maximummutualconducts to regularize facts. The ethics within the dataset are standardized within the assumedvariety of minimum and maximum values from the dataset.

#### 4.2.4 Data Filtering
It's miles the method of selecting a minor part of your information set and using that subset for viewing or evaluation.

### 4.3 Classification
Category is the remaining step for detecting students' educational performance. This paper uses an ensemble of ML models like NB, LR, SVM, RF, KNN, and DT for college students' instructional overall performance prediction, each of which is defined within the subsections below.

#### 4.3.1 Decision Tree
A DT model is a tree construction that looks like a flowchart. To eachinterior node on this structure characterizes a take a look at a dataset characteristic, while every tree divisioncharacterizes the test result. Furthermore, every child node represents a goal specific label, in addition the tree's uppermost node represents the foundation node, which may be binary or non-binary. DT is a popular category approach because it requires an extensive understanding of the problem domain or complex parameter settings. Furthermore, they're quickly converted to class regulations and are easily understood. The set of rules employs a characteristic or characteristic selection degree when constructing the selection tree.

#### 4.3.2 Random Forest
RF is one of the best and completely automatic machine-studying strategies. The use of DT stimulates RF. It combines the concept of "bagging" with randomly selected functions. Furthermore, it's created on class and Regression timber (CARTs) groups to make guesses. RF indicates extra efficiencyby means of extensive information regarding its capacity to deal with many variables without deleting them. It has three foremost hyperparameters that need to be set before schooling. These include node size, the range of timber, and the variety of features sampled. From there, the random woodland classifier can solve regression or type troubles. The last classification choice is primarily based on the common chances envisioned using all produced bushes. RF technique is defined as a fixed selection timber {h (x, θok), k=1,...}, wherein h (x, θokay) is a meta-classifier, notably, an unpruned choice tree created with the use of a CART set of rules; x serves because the input vector, even as {θ okay } is an unbiased and identically allotted random vector. They determine the increased manner of every choice tree.

### 4.3.3 Support vector machine

SVM is a support vector appliance algorithm used designed fortogetherorganization and reversion. The principle objective of the SVM set of rules is to locate the most helpful hyperplane in an N-dimensional area, which can separate the statistics points in one-of-a-kind training in the feature area. The hyperplane tries to ensure the margin among the closest points of different lessons is as high as possible. The measurement of the hyperplane relies upon the range of functions. If the range of enter capabilities is, then the hyperplane is just a line. If the wide variety of input capabilities is three, the hyperplane becomes a 2-D aircraft. It will become challenging to assume when the array of functions exceeds three.

### 4.3.4 Logistic Regression

LR is one of the most popular ML algorithms, and it is a thundered supervised learning method. It's widely used to predict the explicit dependent variable by means of a given set of neutral variables. Logistic regression predicts the output of an expressly established variable. Therefore, the final results must have a specific or discrete value. It could be both sure or No, zero or 1, proper or false, etc. However, as an alternative of charitabledetailed values of zero and 1, itsuggestions probabilistic standards between 0 and 1. The subsequent system characterizes the logistic model.

$$\left(A = 1 \middle| X_1, X_2, \ldots, X_k \right) = 11 + e - \left(\alpha + \sum \beta_i X_j \, \mathrm{k}1 \right)$$

Wherever$\alpha$ and $\beta$ are the representation'slimitations that container be educatedafteraestablished of labelledoccurrences in the exercise dataset and X is the input data.

### 4.3.5. AdaBoost

Adaboost set of rules, quick-time period for Adaptive Boosting, is a Boosting technique used as an Ensemble method in ML. It's called Adaptive Boosting because the weights are re-assigned to every example, with higher weights assigned to incorrectly labelled instances. Boosting is used to reduce bias and variance for supervised getting-to-recognise. It works at the precept of novices, who develop sequentially. Besides number one, each subsequent learner is grown from formerly grown beginners. In easy phrases, vulnerable, inexperienced persons are transformed into strong ones. The AdaBoost classifier operates in the following manner:

Firstly, the AdaBoost algorithm selects training samples randomly from the available dataset. This random selection ensures that the algorithm is not biased towards any specific subset of the data, allowing for a more comprehensive learning process.

- It trains the AdaBoost system to get to know the algorithm by choosing the samples primarily based on the perfect analysis of the remaining education
- It allocates the better weight to wrong labeled samples so the following repetition of classification receives the high probability for category
- It allocates the burden to the trained classifier in every repetition in step with the accuracy of the classifier. It generates an excessive-performance classifier
- This manner repeats till the entire schooling sample is suited without any mistakes

Finally, the AdaBoost algorithm performs a voting method on all the learning algorithms it has generated. This voting method associations the estimates of each individual algorithm to make a final, more correct prediction. This iterative process continues until the entire training dataset is classified without any errors.

### 4.3.6 XGBoost

Extreme Gradient Boosting, or XGBoost, is an effective gradient boosting and choice tree primarily based on an ensemble gadget mastering a set of rules. Its underlying principle includes aggregating predictions from many vulnerable choice tree models to create a more accurate and robust one. XGBoost builds upon the conventional gradient boosting version, which iteratively contains new models to correct the errors made via previous models. It presents an efficient method for mitigating the damaging results of overfitting and enhancing version generalization through integrating regularization phrases that control the complexity of the model.

### 4.3.7 Stochastic Gradient

Gradient Descent rules may be used towardsdiscovery the pleasant standards of the version's limitations throughout the schooling section. Stochastic Gradient Descent (SGD) is a gradient descent variant that optimizes machine learning models. It addresses the computational inefficiency of conventional Gradient Descent methods in gadget-studying projects dealing with massive datasets. The SGD method does now not choose the complete dataset; instead, a few samples are selected at random. A batch is a set of samples. The batch is generated from the entire dataset. When the dataset is large, the complexity increases. It makes use of an unmarried facts pattern. The sample is then randomly exchanged with the following sample, and the manner is repeated. It improves the classifier's efficiency and is simple to put into effect. Gradient descent means descending a slope to the bottom factor of the hill.

## 5. RESULTS AND DISCUSSION

This segment discusses the investigational outcomes of the proposed ML models, which are carried out in Python. Proposed ML fashions like SVM, DT, RF, Adaboost, XGBoost, LR, and SGD are compared in terms of accuracy for the educational dataset from the UCI repository.

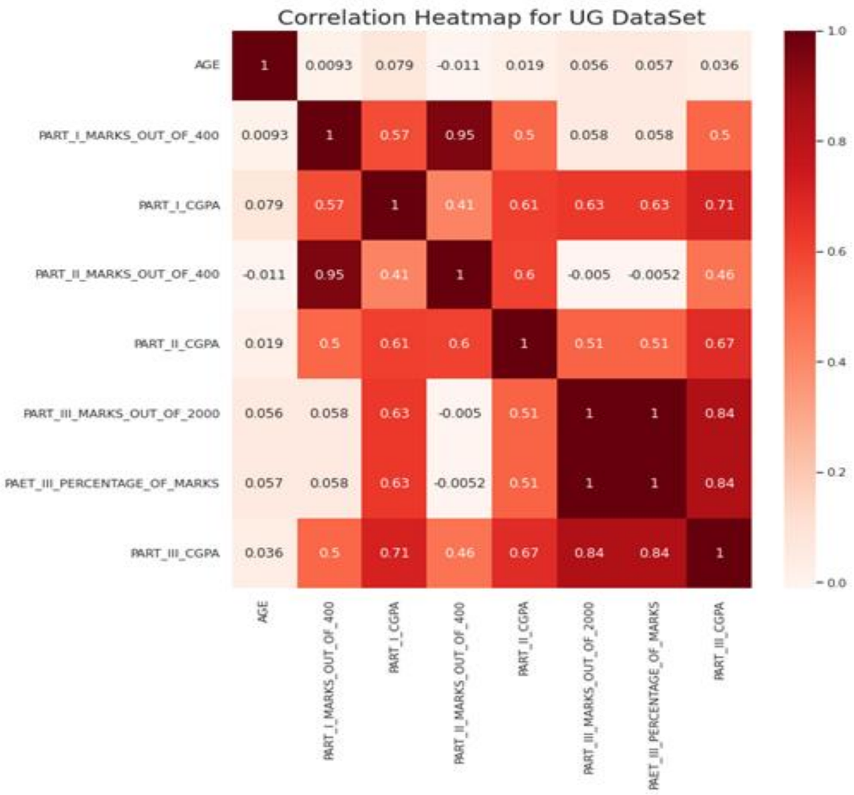


**Figure 2.** correlation heat map for UG dataset

Figure 2 Depicts the correlation heat map for the UG dataset for students' a while and grades in all parts. It represents the connection among variables. We can see if there are any patterns in the price for one or both variables by staring at how mobile colorations exchange across each axis. It converts the correlation matrix right into a coloration coding gadget.

**Table 1.** Results of the classifiers

| ML model | Model accuracy (%) | Model score (%) |
|---|---|---|
| Random forest | 60.98 | 84.32 |
| Decision tree | 68.06 | 72.36 |
| Logistic regression | 67.73 | 69.19 |
| Support vector machine | 66.29 | 67.57 |
| Adaboost | 70.39 | 69.90 |
| XGBoost | 61.42 | 75.78 |

Desk 1 shows the ML models' results in terms of accuracy and rating. In assessing different models, Adaboost achieves the best accuracy of 70.39%, while RF achieves the lowest accuracy of 60.98%; it outperforms all other models with a rating of 84.32%.

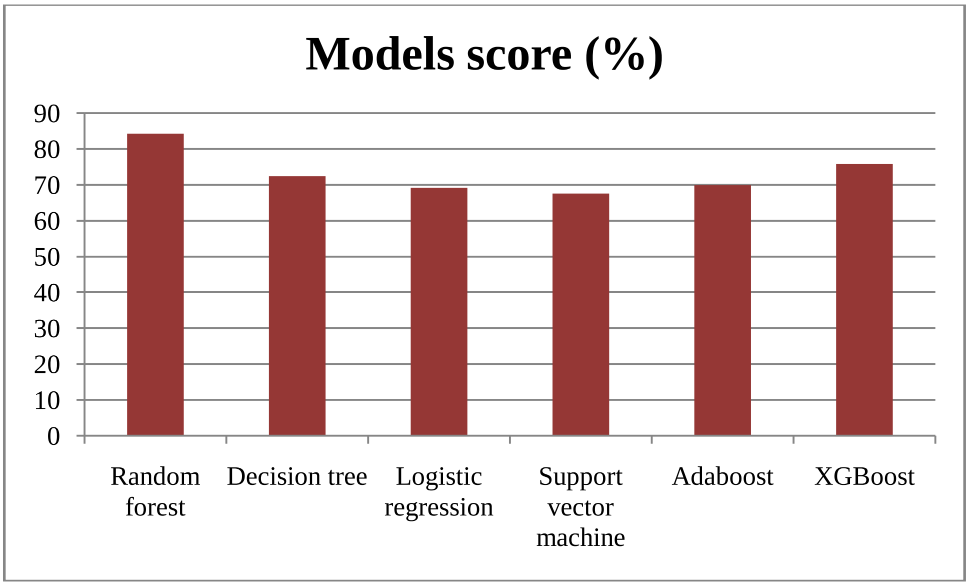


**Figure 3.** Accuracy comparisons of ML models

Figure 3 This graph depicts ML models with rectangular bars whose peak and width are proportional to their accuracy or move validation rating. In this graph, Adaboost has the best accuracy of -70.39%, while the opposite models RF, DT, LR, SVM, Adaboost, and XG raise have accuracy of 60.98%, 68.06%, 67.73%, 66.29%, and 61.42%, respectively.

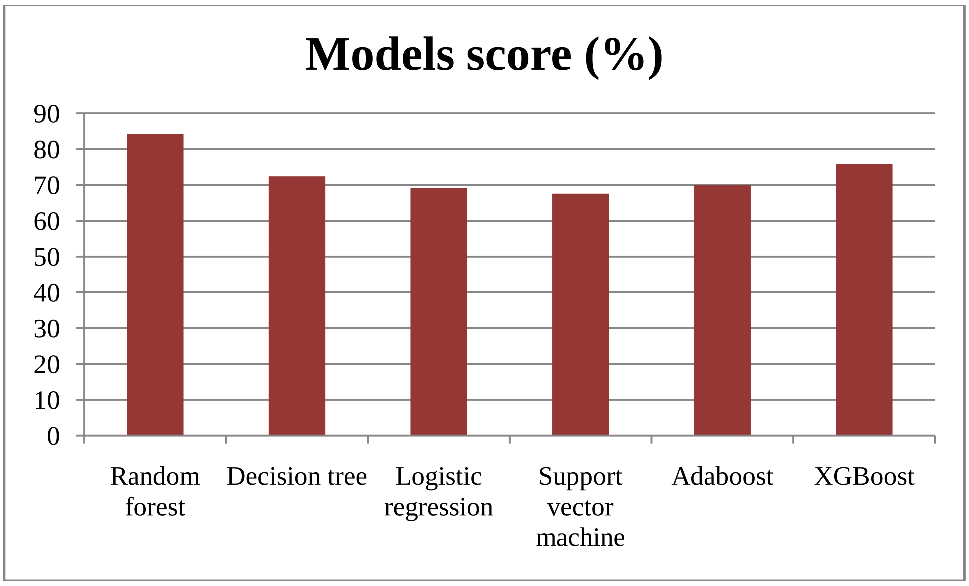


**Figure 4.** Score of ML models

Figure 4 It depicts the ML fashions with rectangular bars whose peak and width are proportional to their rankings. In this graph, RF achieves the very best rating of 84.32%, while different models, which include LR, DT, Adaboost, XG enhance, and SVM, acquire ratings of 69.19%, 72.36, and 69.90%, 75.78%, and 67.57%.

## 5. CONCLUSION

Student academic overall performance is one of the essential pleasant signs for every college. Being capable of expecting the identity of at-risk students at an early degree of pupil instructional lifestyles allows for the enhancement of getting to know and lessening dropout quotes. In this work, we've index the accuracy of seven unique gadget-mastering algorithms, which is an excellent way to expect to pass or fail the final exam. The seven ML algorithms investigated are XG increase, LR, RF, DT, Adaboost, and SVM. The outcome proved that the model Adaboost indicates outstanding performance by reaching an accuracy of 70.39% because its miles are much less vulnerable to overfitting because of its adaptive nature. In the future, it will consist of more statistics and additional functions, which will provide a more accurate version for predicting pupil educational overall performance and support the complete mastering technique.

## REFERENCES

[1] Sreenivasulu, Mr D., J. Sirisha Devi, P. Arulprakash, S. Venkataramana, and Kutubuddin Sayyadliyakat Kazi. "Implementation of latest machine learning approaches for student's grade prediction." Int. J. Early Child 14, no. 3 (2022).

[2] Sandra, Lidia, Ford Lumbangaol, and Tokuro Matsuo. "Machine Learning Algorithm to Predict Student's Performance: A Systematic Literature Review." TEM Journal 10, no. 4 (2021).

[3] Baashar, Yahia, Gamal Alkawsi, Abdulsalam Mustafa, Ammar Ahmed Alkahtani, Yazan A. Alsariera, Abdulrazzaq Qasem Ali, Wahidah Hashim, and Sieh Kiong Tiong. "Toward predicting student's academic performance using artificial neural networks (ANNs)." Applied Sciences 12, no. 3 (2022): 1289.

[4] Jovana Jović , Emilija Kisić , Miroslava RaspopovićMilić , Dragan Domazet , Kavitha Chandra "Prediction of students' academic performance using Machine Learning Techniques." In 2022 International Mobile and Embedded Technology Conference (MECON), pp. 151-156. IEEE, 2022.

[5] Verma, Utkarsh, Chetna Garg, Megha Bhushan, Piyush Samant, Ashok Kumar, and Arun Negi. "Prediction of students' academic performance using Machine Learning Techniques." In 2022 International Mobile and Embedded Technology Conference (MECON), pp. 151-156. IEEE, 2022.

[6] Alghamdi, Amnah Saeed, and Atta Rahman. "Data mining approach to predict success of secondary school students: A Saudi Arabian case study." Education Sciences 13, no. 3 (2023): 293.

[7] Ojajuni, Opeyemi, Foluso Ayeni, Olagunju Akodu, Femi Ekanoye, Samson Adewole, Timothy Ayo, Sanjay Misra, and Victor Mbarika. "Predicting student academic performance using machine learning." In Computational Science and Its Applications–ICCSA 2021: 21st International Conference, Cagliari, Italy, September 13–16, 2021, Proceedings, Part IX 21, pp. 481-491. Springer International Publishing, 2021.

[8] Asefa, Girma. "Prediction of Students' Academic Performance Using Machine Learning: The Case of Wachemo University." (2021).

[9] Altabrawee, Hussein, Osama Abdul Jaleel Ali, and Samir Qaisar Ajmi. "Predicting students' performance using machine learning techniques." Journal Of University Of Babylon for pure and applied sciences 27, no. 1 (2019): 194-205.