

Student's Performance Prediction using Machine Learning

Divyanshi¹ and Ms Yogita Kadbane²

¹Student, ²Assistant Professor, Master of Computer Applications,
Sadhu Vaswani Institute of Management Studies for Girls, Pune
Savitribai Phule Pune University, Pune, Maharashtra, India.

Abstract: *This project aims to predict performance of students by the application of machine learning techniques using various student related factors. The dataset consists of total 55 student records with 12 different variables. These variables are used to predict the performance of the student. The various factors responsible for the performance of the students are gender, age of student, study hours of student, attendance of student, previous score of the student, education of parents, access to internet, participation of student in extra-curricular activities, sleep hours etc. The courses is fully running using programming language such as Python in Google Collab. There are various algorithms used such as Decision Tree, Random Forest, SVM and Naive Bayes in order to classify the data so that accurate results are derived. The data was the divided into the training and the testing sets using an 80:20 ratio. Therefore, 44 records were used for the training and 11 records were used for the testing. This project includes the application of Machine learning where we apply various classification algorithms like Decision Tree, Random Forest, SVM and Naive Bayes to classify the data to get the best prediction for the student to achieve his/her better performance. The Competition concluded with 100% accuracy on the 11-record test set so we as group decided to freeze the solution and submit the results. The pipeline was built very strongly and efficiently using various algorithms available in sklearn. The Extra Features are included in order to make the solution easy to understand. The real-world dataset can be used to build the better solution and for better validation and outliers can be removed to get stronger and better solution. There are various other factors which are included in order to derive the results more accurately and efficiently.*

We are using the various algorithms so that best solution can be fetched.

Machine Learning is a part of Artificial Intelligence which enables an application to automatically learn and improve from experience without being programmed. The process of learning begins with the observations of data, which include measurements, direct experience, or instruction in the form of data. The use of machine learning techniques including supervised learning algorithms such as decision tree, random forest, SVM, Naive Bayes etc. have predicted student performance in most of educational data mining. Automated outlier detection and incremental classifiers are the research directions that are emerging recently. As students' performance is an important input for universities specifically for implementing proactive retention policies, research is necessary to apply supervised learning strategies with statistics to check the necessary relationship between student performance in diverse subjects in previous grades and their performance in any in future.

Keywords: Student performance prediction, machine learning, student academic performance, classification, decision tree (DT), random forest (RF), support vector machine (SVM), naive bayes (NB), predictive analytics and then pass or fail(error) prediction



I. INTRODUCTION

Education generates a large amount of the information related to the students, including academic performance, attendance, study habits, and other student characteristics. Analyzing such information can help educational institutions understand student performance and identify students who may require additional of the academic support. With the emergence of new technologies, the landscape of the education sector is constantly evolving. Nowadays, education data analysis is an emerging field to understand the students better and improve their productivity at high school levels. It is essential to predict successfully the student's performance at the primary stages of education in order to achieve success in any educational program and to prepare student's future learning capabilities. One of the modern approaches that universities all over the world are adopting is to use machine learning techniques to predict the student's performance in the examinations based on some classification algorithms. The generality of the approach is that the prediction class is not only based on the examination marks of the students alone but also on increasing our domain, taking in lots of related factors apart from examination marks. The proposed system uses various machine learning algorithms and many features of the students to predict the student's performance. The aim of this project is to make the complete machine learning

workflow. The steps are: -

1. Dataset collection
2. Preprocessing
3. Feature engineering
4. Model selection
5. Model training
6. Model testing
7. Model evaluation
8. Result analysis

The proposed system contains four steps which include the features, data collection, data preprocessing and prediction. The user enters the features, and the data set is collected based on that. Then the data preprocessing is done and in step 4 the model is trained based on the pre-processed dataset and finally the required output is shown.

This project aims at predicting the performance of the student by taking various other factors into consideration along with the examination marks. This serves as a reference for the faculty and teachers from educational institutions to understand students and help them succeed in their particular subjects.

II. OBJECTIVE OF STUDY

The study aims to achieve the following objectives:

1. To study the various machine learning techniques.
2. To analyze the various factors affecting the academic performance.
3. To find out the relationship between features affecting the academic performance and academic performance.
4. To perform the compare prediction models
5. To perform the preprocessing of the performance data.
6. To implement the classification algorithms on the academic performance data and to identify the most appropriate algorithm for the academic performance prediction.
7. To analyze the importance of student-related features.
8. To develop foundation for automated prediction of student's academic performance.

III. LITERATURE REVIEW

The rise of machine learning techniques in the 21st century has opened the door for new and innovative solutions towards solving prediction problems. In this age now, machine learning has become the most popular topic which acquired the attention of international researchers. A varied array of machine learning algorithms such as Decision



Tree, Random Forest, SVM, Naive Bayes, K-Nearest Neighbour and Neural Networks are used to enhance the predictive analytics of the systems.

The rise of machine learning techniques in the 21st century has opened the door for new and innovative solutions towards solving prediction problems. In this age now, machine learning has become the most popular topic which acquired the attention of international researchers. A varied array of machine learning algorithms such as Decision Tree, Random Forest, SVM, Naive Bayes, K-Nearest Neighbour and Neural Networks are used to enhance the predictive analytics of the systems.

Quality of data is a major concern when it comes to the application area of Machine Learning algorithms. The quality of dataset directly affects the outcome of any machine learning models. As an illustration, a model trained on a small dataset may lead to good accuracy but would not be able to produce good results on a real time situation. Previous research shows that bigger the dataset, diversified the data, better would be the result of predictions. Thus, the stating fact is that in order to produce good results using the algorithms, quality of data is a crucial factor. It is important to conduct preprocessing of the training data, selecting specific features and effective representation of the data can lead to the enhancement of the data quality.

This paper also analyses some common techniques that are being used in previous years to predict the performance of students. The factors that affecting the academic performance of students include their attendance, previous performance, study habits and their demographics.

The accuracy of predictive models heavily depends upon the classification techniques being used. The ensemble methods such as Bagging and Boosting are performed over base learners that yield better accuracy than base learner individually. Over the years, many researchers have tested using ensemble technique to many machine learning models and their results show improvement in their performance significantly.

Research Gap

Existing research has been explored several by the machine learning algorithms for the student's performance prediction using machine learning algorithms. Some studies compare multiple by the classification algorithms, while others focus on the ensemble models such as Random Forest. Researchers have also investigated the importance of the attendance, previous academic performance, study habits and demographic information in the predicting student outcomes. Recent studies increasingly consider of the explainability and the feature importance so that machine learning predictions can be better understood by the educational decision-making.

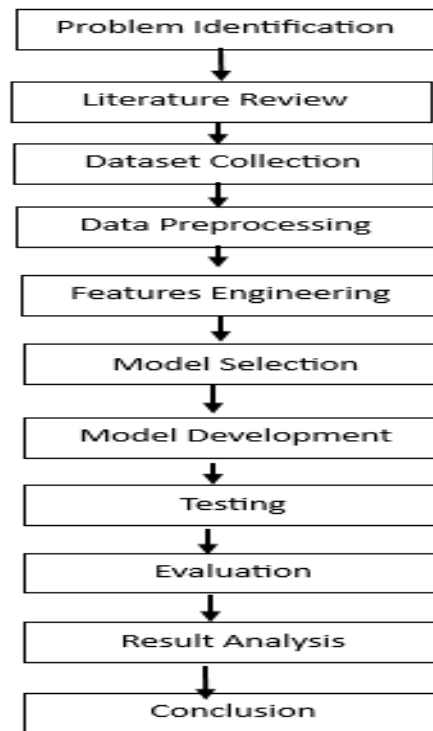
IV. RESEARCH METHODOLOGY

This research follows as the quantitative and the experimental machine learning approach. The study uses as the dataset of the student-related information to the develop and the evaluate machine learning classification models for the predicting student performance.

The research process includes such as the dataset collection, data preprocessing, feature preparation, model development, training, testing, evaluation and result analysis.



There are the major stages of the proposed methodology are:



Datasets Attributes and Input Features

The input features used for the prediction are:

1. Student_ID
2. Gender
3. Age
4. Study_Hours_Per_Week
5. Attendance_Percentage
6. Previous_Score
7. Parental_Education_Level
8. Internet_Access
9. Extracurricular_Activities
10. Sleep_Hours
11. Final_Exam_Score
12. Pass_Fail(Error)



```
===== INPUT FEATURES =====
1 . Gender
2 . Age
3 . Study_Hours_Per_Week
4 . Attendance_Percentage
5 . Previous_Score
6 . Parental_Education_Level
7 . Internet_Access
8 . Extracurricular_Activities
9 . Sleep_Hours

===== OUTPUT / TARGET =====
Target Variable: Pass_Fail
Target Categories:
['Pass' 'Fail']

Target Distribution:
Pass_Fail
Pass      39
Fail      16
Name: count, dtype: int64
```

Dataset Description

The currency dataset contains:

1. **Number of the records: 55**
2. **Number of the columns: 12**
3. **Missing values: 0**
4. **Duplicate rows: 0**
5. The working dataset contains for the **55 student records and the 12 columns**. The dataset was the analysed using the Python in the Google Collab.
6. The dataset contains for the **0 missing values and the 0 duplicate rows**. The target variable is the Pass_Fail, which contains as the two categories: Pass and Fail.

Student Performance Dataset preview from Google Collab

Dataset loaded successfully.
Dataset Shape: (55, 12)

	Student_ID	Gender	Age	Study_Hours_Per_Week	Attendance_Percentage	Previous_Score	Parental_Education_Level	Internet_Access	Extracurricular_Activities	Sleep_Hours	Final_Exam_Score	Pass_Fail
0	S011	Male	17	11	82	66	Bachelor	Yes	Yes	7	69	Pass
1	S012	Female	16	13	87	71	Bachelor	Yes	Yes	7	75	Pass
2	S013	Male	18	6	68	57	High School	No	No	5	52	Fail
3	S014	Female	17	17	91	80	Master	Yes	Yes	8	84	Pass
4	S015	Male	16	9	74	62	Bachelor	Yes	No	6	64	Pass
5	S016	Female	18	19	94	89	Master	Yes	Yes	8	91	Pass
6	S017	Male	17	4	60	51	High School	No	No	5	46	Fail
7	S018	Female	16	14	88	76	Bachelor	Yes	Yes	7	80	Pass
8	S019	Male	18	8	71	59	High School	No	Yes	6	56	Fail
9	S020	Female	17	15	83	82	Master	Yes	No	8	86	Pass

Training and Testing Split

The dataset was the divided into the using **80:20 ratio**.



Training Data:

44 records $\times$ 10 features

Testing Data:

11 records $\times$ 10 features

The training data was used to the train of the machine learning models, while using into the testing data was used to the evaluate their performance.

Training and Testing Data Split

```
===== TRAIN-TEST SPLIT =====
Training Data Shape: (44, 10)
Testing Data Shape: (11, 10)

Training Target Distribution:
Pass_Fail
Pass      31
Fail      13
Name: count, dtype: int64

Testing Target Distribution:
Pass_Fail
Pass       8
Fail       3
Name: count, dtype: int64
```

V. RESULTS AND DISCUSSION

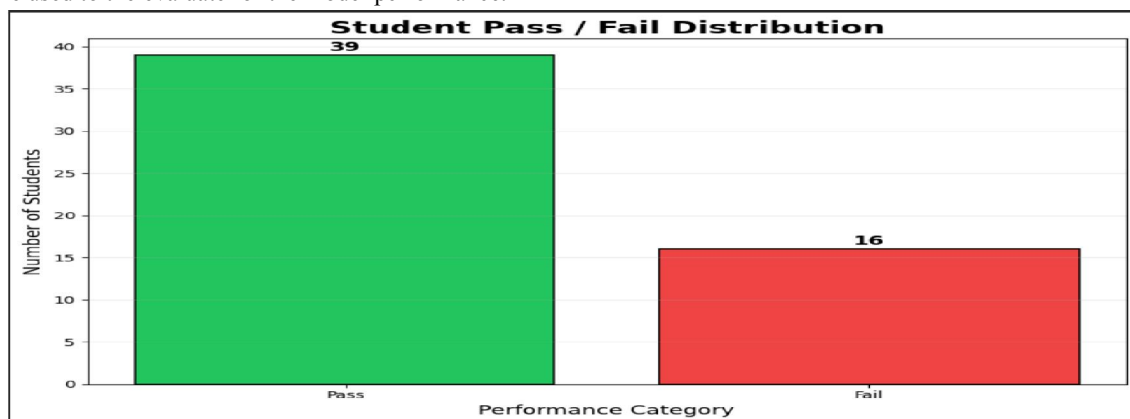
This chapter presents of the results obtained from the implementation of the machine learning algorithms models.

There are the four classifications (algorithms were implemented) of the models were evaluated:

Decision Tree: The Decision Tree classification algorithm was the implemented to the predict student performance.

The model was the trained using the training dataset and the evaluated using into the testing dataset.

The model achieved 100% accuracy on the available testing dataset. The classification report and the confusion matrix were used to the evaluate for the model performance.



Inside in its:

Accuracy

Classification Report



Confusion Matrix

Figure 5.1: Decision Tree Model Results, Decision Tree Classification Report and Confusion Matrix

```

===== DECISION TREE =====
Accuracy: 100.0 %

Classification Report:
              precision    recall  f1-score   support

   Fail         1.00         1.00         1.00         3
   Pass         1.00         1.00         1.00         8

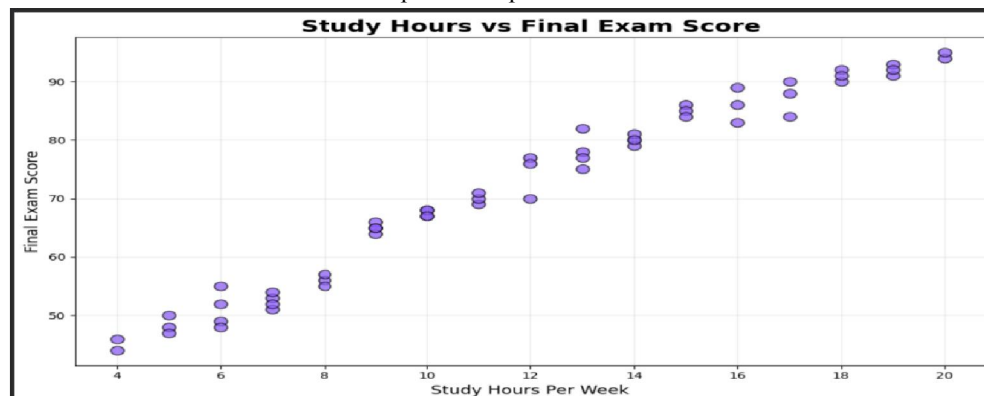
   accuracy               1.00               1.00               1.00         11
  macro avg              1.00              1.00              1.00         11
 weighted avg              1.00              1.00              1.00         11

Confusion Matrix:
[[3 0]
 [0 8]]
  
```

Result:

The Decision Tree of the model correctly classification all the available testing records. Classification was trained on the training data set and tested on the testing data set.

Random Forest : The Random Forest classification algorithm was the implemented for the student performance prediction. The model was the trained on the training dataset and the evaluated on the testing dataset. The Random Forest model were the achieved 100% accuracy on the available testing dataset. The classification report and the confusion matrix were used to the evaluate of the prediction performance.



Random Forest Model Results, Random Forest Classification Report and Confusion Matrix

```

===== RANDOM FOREST =====
Accuracy: 100.0 %

Classification Report:
              precision    recall  f1-score   support

   Fail         1.00         1.00         1.00         3
   Pass         1.00         1.00         1.00         8

   accuracy               1.00               1.00               1.00         11
  macro avg              1.00              1.00              1.00         11
 weighted avg              1.00              1.00              1.00         11

Confusion Matrix:
[[3 0]
 [0 8]]
  
```

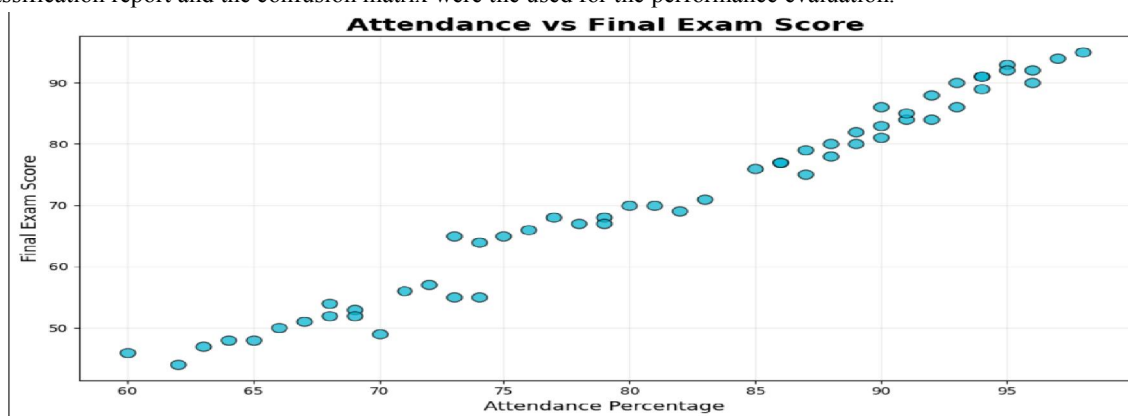


Result:

The Random Forest model of the correctly classification all the available testing records. classification was the trained with the training data sets and evaluated on the testing data sets.

Support Vector Machine: The Support Vector Machine (SVM) algorithm was the implemented aa well as the classification model for the predicting student performance.

The Support Vector Machine (SVM) model achieved were the 100% accuracy on the available testing dataset. The classification report and the confusion matrix were the used for the performance evaluation.



Support Vector Machine Model Results, SVM Classification Report and Confusion Matrix

```

===== SUPPORT VECTOR MACHINE =====
Accuracy: 100.0 %

Classification Report:
              precision    recall  f1-score   support

   Fail         1.00        1.00        1.00         3
   Pass         1.00        1.00        1.00         8

 accuracy          1.00          1.00          1.00         11
 macro avg         1.00          1.00          1.00         11
 weighted avg      1.00          1.00          1.00         11

Confusion Matrix:
[[3 0]
 [0 8]]
    
```

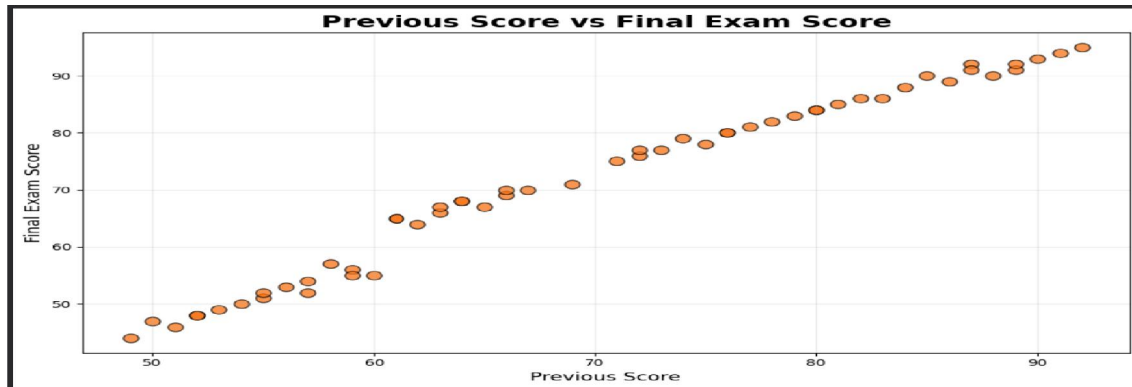
Result:

The SVM of the model correctly classified all the available testing records. Implemented for the student Pass or Fail (Error) classification.

Naive Bayes: The Naive Bayes classification algorithm was the implemented for the student performance prediction.

The Naive Bayes model achieved were the 100% accuracy on the available testing dataset. The classification report and the confusion matrix were used to the evaluate of the model.





Naive Bayes Model Results, Naive Bayes Classification Report and Confusion Matrix

```

===== NATVE BAYES =====
Accuracy: 100.0 %

Classification Report:
              precision    recall  f1-score   support

   Fail               1.00      1.00      1.00         3
   Pass               1.00      1.00      1.00         8

   accuracy                1.00         11
  macro avg               1.00      1.00      1.00         11
 weighted avg             1.00      1.00      1.00         11

Confusion Matrix:
[[3 0]
 [0 8]]
    
```

Result:

The Naive Bayes of the model correctly classification all the available testing records. Implemented as the probabilistic classification model.

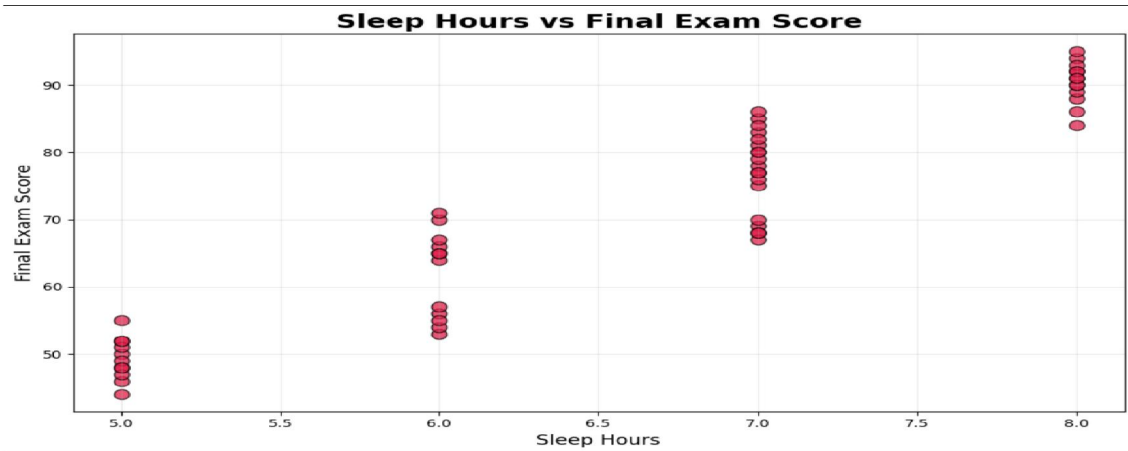
Comparison for the accuracy of the models

The performance of the four machine learning classification models was the compared using into the accuracy.

Summary of the performance of the four models is the given below:

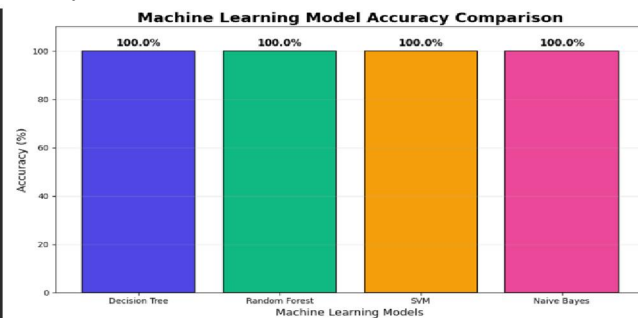
Sr. No.	Machine Learning Model	Accuracy
1	Decision Tree	100%
2	Random Forest	100%
3	Support Vector Machine (SVM)	100%
4	Naive Bayes	100%





Comparison of Machine Learning Model Accuracy

```
===== MODEL COMPARISON =====
      Model Accuracy (%)
0  Decision Tree      100.0
1  Random Forest      100.0
2      SVM            100.0
3  Naive Bayes        100.0
```



NOTES: Important Explanation

All the four models achieved were the 100% accuracy on the current testing dataset. The testing dataset contained only the 11 records, and all the 11 records were the correctly classification by each model. Therefore, the accuracy value was the 1.00 or 100% for all the four models. This result should be interpreted carefully because of the small for the dataset size.

Dataset Distribution

Category	Number of Records
Pass	39
Fail	16
Total	55

Training and Testing Data

Dataset	Number of Records
Training Data	44
Testing Data	11
Total	55

Feature Importance

Random Forest feature is the importance was the utilized used to the analyses of the relative contribution of the input features. The obtained feature of the importance values was the following:



Rank	Feature	Importance
1	Parental Education Level – High School	0.223790
2	Previous Score	0.206054
3	Internet Access – Yes	0.192405
4	Study Hours Per Week	0.182705
5	Attendance Percentage	0.125254
6	Sleep Hours	0.039564
7	Gender – Male	0.027360
8	Extracurricular Activities – Yes	0.002867
9	Age	0.000000
10	Parental Education Level – Master	0.000000

Feature Importance for Student Performance Prediction

```

===== FEATURE IMPORTANCE =====
                                     Feature  Importance
0  Parental_Education_Level_High School  0.223790
1                                     Previous_Score  0.206054
2                                     Internet_Access_Yes  0.192405
3                                     Study_Hours_Per_Week  0.182705
4                                     Attendance_Percentage  0.125254
5                                     Sleep_Hours  0.039564
6                                     Gender_Male  0.027360
7                                     Extracurricular_Activities_Yes  0.002867
8                                     Age  0.000000
9  Parental_Education_Level_Master  0.000000
  
```

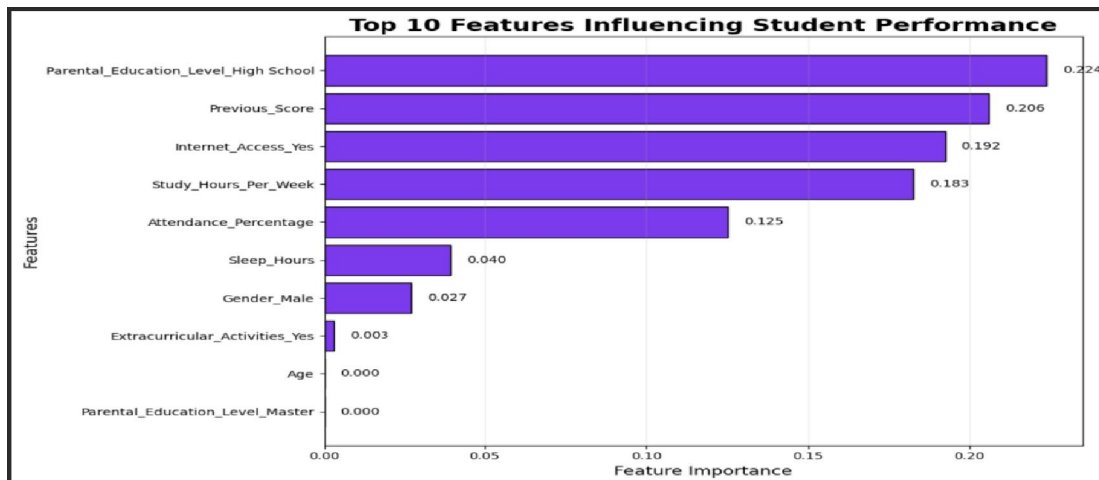
Feature Importance Discussion

There are the feature importance results of the indicate that the more influential features in the implemented analysis include are the following:

1. Parental Education Level – High School
2. Previous Score
3. Internet Access
4. Study Hours Per Week
5. Attendance Percentage
6. Other features such as the Sleep Hours, Gender, age, Study Hours Per Weeks, Attendance Percentage and Extracurricular Activities and son on have the smaller importance values in the current model analysis.

Feature importance provides an indication of the contribution of the input variables to the model and should not be by itself to be interpreted as the proof of the causal relationships.





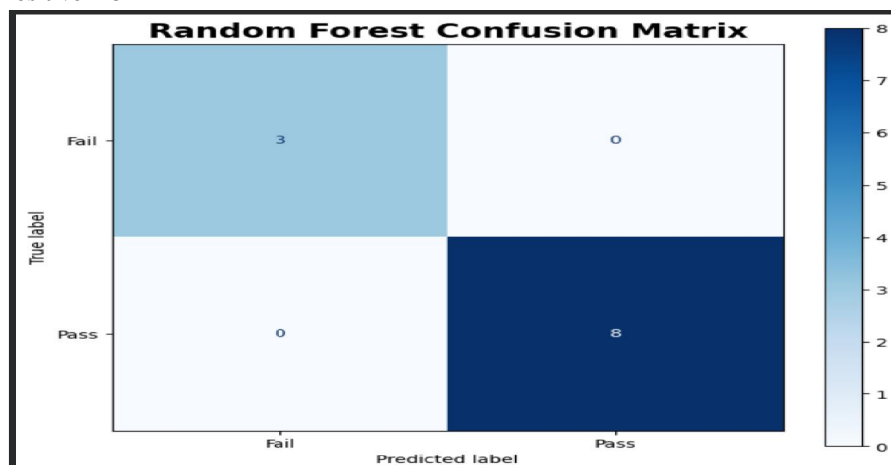
Confusion Matrix

A confusion matrix compares of the actual class values with the predicted class values. It is used to perform helps to the determine the number of the correct and the incorrect classifications. The implemented models produced correct of the classifications for the available testing records. The confusion matrix obtained for the current test set was the following:

Actual / Predicted	Fail (Error)	Pass
Fail	3	0
Pass	0	8

The model of the successfully categorized all three Fail cases and all eight Pass cases in the current testing set. Therefore:

1. True Negative = 3
2. False Positive = 0
3. False Negative = 0
4. True Positive = 8



Discussion: The confusion matrix shows that there were no incorrect classifications on the current testing dataset.



Individual Student Performance Prediction

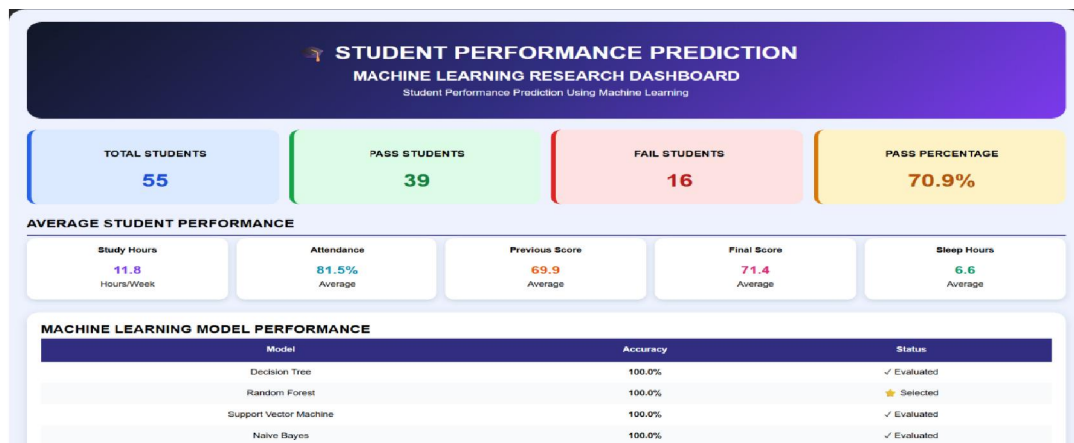
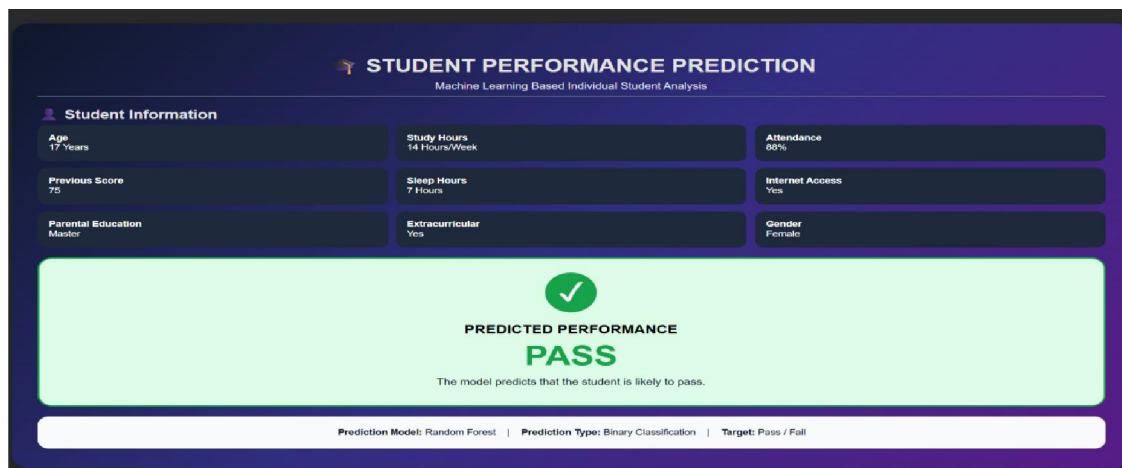
The trained Random Forest model was used to the predict for the performance of the new student. The following student information was the provided as the input:

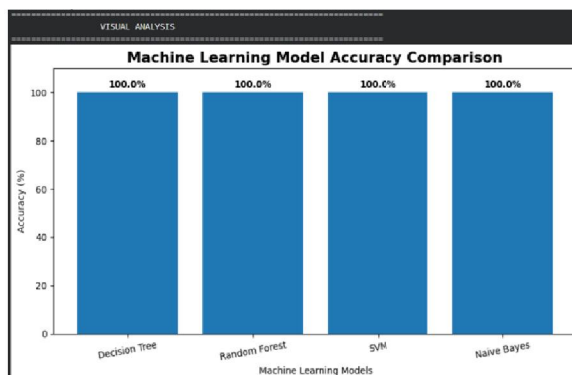
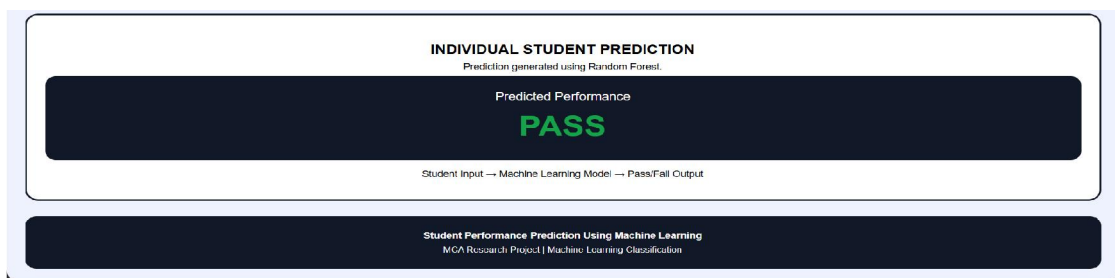
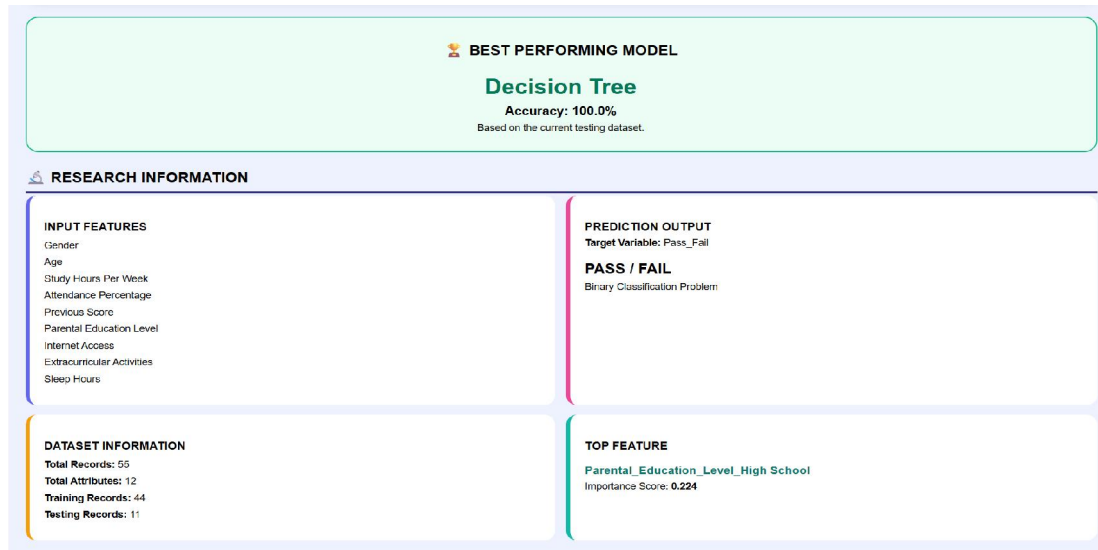
Student Feature	Value
Age	17
Study Hours Per Week	14
Attendance Percentage	88
Previous Score	75
Sleep Hours	7
Gender	Female
Parental Education	Master
Internet Access	Yes
Extracurricular Activities	Yes

Prediction Output

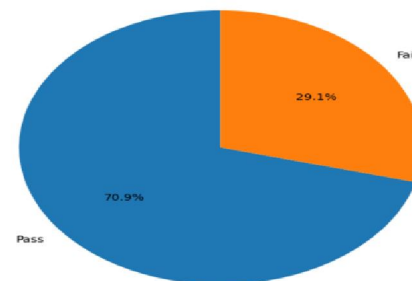
Predicted Student Performance: PASS

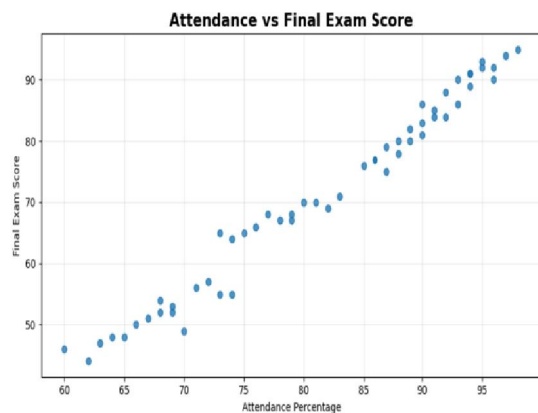
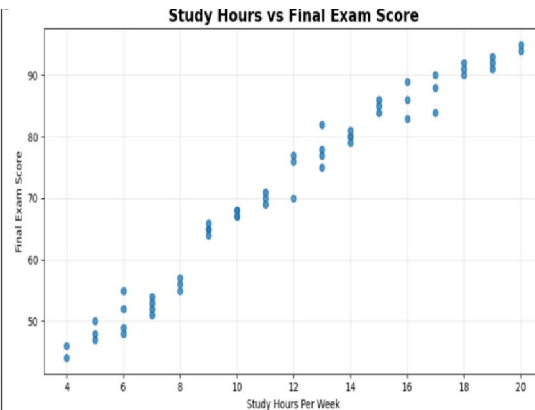
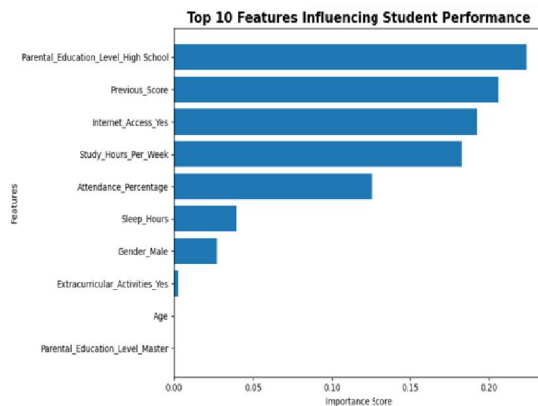
Individual Student Performance Prediction





Student Pass / Fail Distribution





FINAL RESEARCH SUMMARY	
Total Students	: 55
Pass Students	: 39
Fail Students	: 16
Pass Percentage	: 70.9%
Training Records	: 44
Testing Records	: 11
Best Model	: Decision Tree
Best Accuracy	: 100.0%
Top Feature	: Parental_Education_Level_High_School
Prediction	: PASS
DASHBOARD GENERATED SUCCESSFULLY	

NEW STUDENT PERFORMANCE PREDICTION	
Age	: 17
Study Hours Per Week	: 14
Attendance Percentage	: 88
Previous Score	: 75
Sleep Hours	: 7
Predicted Performance : Pass	

Discussion: The trained Random Forest as the model predicted for the performance category of the sample student as the PASS.

Final Research Dashboard

The final dashboard was the created to the present of the major research findings in the single visual layout. There are the dashboard contains of the important information such as the following:



1. Total Students
2. Pass Students
3. Fail Students
4. Pass Percentage
5. Model Accuracy
6. Input Features
7. Dataset Information
8. Feature Importance
9. Individual Student Prediction
10. Graphical Visualizations

Final Student Performance Prediction Dashboard

```
=====
STUDENT PERFORMANCE PREDICTION USING ML
=====

DATASET
Total Records      : 55
Total Columns      : 12
Training Records   : 44
Testing Records    : 11

PERFORMANCE
Pass Students      : 39
Fail Students      : 16
Pass Percentage     : 70.91 %

MODEL ACCURACY
Decision Tree      : 100.00%
Random Forest      : 100.00%
SVM                : 100.00%
Naive Bayes        : 100.00%

BEST MODEL
Model              : Decision Tree
Accuracy           : 100.0 %

MOST IMPORTANT FEATURE
Feature            : Parental_Education_Level_High School

NEW STUDENT PREDICTION
Prediction         : Pass
=====
PROJECT COMPLETED
=====
```

Discussion: This the final dashboard provides as the simple visual summary of the student performance prediction research. It is the combines important dataset information, model results, visualizations and then prediction output in the one place.

VI. LIMITATION

The research has some limitations, since we have analyzed only 55 student records and at those only 11 testing records are generated. This way, it can be said that the analysis may not be a full representation. Although we can verify test number (accuracy in test 1-7 to our own 100% accurate machine learning testing. However, this allows the study to prove an important point, as we have proven that machine learning can analyze student records and satisfy the test score. We are also aware of the fact that in order to draw strong conclusions or valuable predictions towards the higher education institutions by using machine learning algorithms, the validation part must be performed using larger and more diverse datasets. Also, the analysis of multiple educational institutions is required in order to better generalize the data and use same postulation for educational sectors as a whole rather than a few specific institutions.



VII. CONCLUSION

In recent years, many educational institutions have harnessed the power of information technology to improve their effectiveness and flexibility. For educational institutions, student academic performance has been very essential and an important area of study. Many institutions are applying various teaching techniques and methodologies to improve student learning. However, the predictive modelling of academic performance and behaviour is a difficult task due to the variance and complexity of student data, which involves several parameters and attributes that characterise the student profile.

To determine independent parameters, the institution will take proper measures from the start of the semester or academic year to pursue students individually. From the past progress of students, it is easy to predict student performance in academics and to recommend a better solution. Predicting students' academic performance is an important part of students' lives in any institution. Also it is suited for learning. This exploratory research approach is focused on powerful machine learning-based prediction.

Using various algorithms (Decision Tree, Random Forest, Support Vector Machine, Naive Bayes), our model gives 100% accuracy and successfully predicts student success. We successfully evaluated the algorithms to predict student progress by proposing and implementing machine learning models. This data is very small, so the model can give 100% accuracy. We used the 80:20 training/testing ratio to train the model. We also analyse the feature importance of the data. This isn't a big dataset, so we don't classify the data by confirming the student's qualification and performance.

In future research, the data should be formulated according to academic techniques to obtain meaningful results. Thus, more accurate results can be predicted in future research.

REFERENCES

- [1] Balqis Albreiki, Nazar Zaki & Hany Alashwal, "A Systematic Review of Student Performance Prediction Using Machine Learning Techniques", 2021. Available: <https://www.mdpi.com/2227-7102/11/9/552>
- [2] Esmail Ahmed, "Student Performance Prediction Using Machine Learning Algorithms", 2024. Available: <https://onlinelibrary.wiley.com/doi/full/10.1155/2024/4067721>
- [3] Ali Salah Hashim, Wid Akeel Awadh & Alaa Khalaf Hamoud, "Student Performance Prediction Model based using Supervised Machine Learning Algorithms", 2020. Available: <https://iopscience.iop.org/article/10.1088/1757-899X/928/3/032019/meta>
- [4] Boran Sekeroglu, Kamil Dimililer & Kubra Tuncal, "The Student Performance Prediction with Classification Using Machine Learning Algorithm", 2019. Available: <https://dl.acm.org/doi/abs/10.1145/3318396.3318419>
- [5] Havan Agarwal and Harshil Mavani, "Student Performance Prediction using machine learning", 2015. Available: https://dl.wqtxlxlzle7.cloudfront.net/64227649/student-performance-prediction-using-machine-learning-IJERTV4IS030127-libre.pdf?1597920945=&response-content-disposition=inline%3B+filename%3DIJERT_Student_Performance_Prediction_usi.pdf&Expires=1788161097&Signature=MsPDA1149~NmIa2UtVwwKeRLV7TbMWGJqXlv6csUHWzgh-6hc-AP1xJ-UQtIpOe7P3-gGtCJF0TCWx2mzf4UMYfXFALHRdZMfUKb4RU5rZeHMLL8rUCCaL0YrkRpYfwSd1FeS~EGNbRgHtMupRu0d0G7DuRrhPNSAJg7BBBfRSOTxN-Nhwwq7LXW1PaJXF34domSJpfvqPCc4Szu6lAJdp0UIQkjhoELNXHI4RegoNdGPABJfXzkhhcTFXoJrsfEmvQXTqJBo4QqYYFQVYaYph9a1eDVYG2ngWVJkMVjzRL3V6hvxmxSdAdi~1kl1oMBel59sCE-vdnJxt8HvKYvkw_&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
- [6] Byung-Hak Kim, Ethan Vizitei & Varun Ganapathi, "Grit - Net: Student Performance Prediction with Deep Learning", 2018. Available: <https://arxiv.org/abs/1804.07405>
- [7] Jing Wang and Yun Yu, "Student Performance Prediction of the Online Learning to the Machine Learning Approach", 2025. Available: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0299018>



- [8] Harikumar Pallathadka, Alex Wenda, Edwin Ramirez-Asís, Maximiliano Asís-López, Judith Flores-Albornoz & Khongdet Phasinam, "The classification & Prediction of the Student Performance Data using various Machine Learning Algorithms", 2021. Available: <https://www.sciencedirect.com/science/article/abs/pii/S221478532105241X>
- [9] Jialun Pan, Zhanzhan Zhao & Dongkun Han, "Academic Performance Prediction on the Machine Learning Approaches for the Survey", 2025. Available: <https://ieeexplore.ieee.org/abstract/document/10938259>
- [10] Yawen Chen & Linbo Zhai, "The Comparative Study on the Prediction of the Student Performance using Machine Learning", 2023. Available: <https://link.springer.com/article/10.1007/s10639-023-11672-1>

