



Your Dreams Our Goal
POORNIMA
UNIVERSITY

Member of Association of Indian Universities & Approved by UGC (Govt. of India) under 2(f) & 12(B)

FINAL DRC VIVA VOCE Presentation
ON
An Improved Ensemble Model for Predicting Student Performance
Based on Academic Achievement or Dropout Rates

MASTER OF TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
WITH SPECIALIZATION IN COMPUTER SCIENCE

UNDER THE SUPERVISION OF:

DR. VISHNU SHARMA
PROFESSOR & HEAD
POORNIMA UNIVERSITY,
JAIPUR

PRESENTED BY :

GARIMA BHATIA,
2021PUSCEMCEX10528
POORNIMA UNIVERSITY,
JAIPUR

CONTENTS

❖ Abstract

❖ Definition

❖ Motivation of research

❖ Introduction

❖ Literature Review

❖ Objectives

❖ Methodology

❖ Results & Discussion

❖ Conclusion

❖ Publication

❖ References

ABSTRACT

An important topic of research in the world of education is the prediction of student performance. Educational institutions may develop focused interventions and support systems by accurately forecasting student outcomes and identifying variables that affect academic achievement or dropout rates. The goal of this study is to create a machine learning-based model for predicting student performance based on academic achievement and dropout rates. The research makes use of a wide range of elements, such as demographic data, prior academic performance, socioeconomic circumstances, and other pertinent variables. The results of this study have important ramifications for educational institutions, decision-makers, and teachers, empowering them to invest funds wisely and implement interventions on time to enhance student outcomes. Overall, by utilising machine learning approaches, this research advances the subject of predicting student performance and lays the groundwork for future developments.

DEFINITION

- The term "Student Performance" refers to the academic achievements and overall educational outcomes of students in a given context. It can be measured through various indicators, including academic achievement, dropout rates, graduation rates, and other assessments of learning.
- **Academic Achievement:** This aspect of student performance evaluates how well students are performing academically in their studies. It can be measured through grades, class rankings, GPA (Grade Point Average), and performance on standardized tests.
- **Dropout Rates:** Dropout rates represent the percentage of students who leave school before completing their education. High dropout rates are generally considered a negative indicator of student performance and educational quality, as they may indicate issues with student engagement, support, and school environment.

MOTIVATION OF RESEARCH

As researchers, our first task is to define the issue facing the field of study.

Educational Quality and Accountability: Understanding student performance is essential for assessing the overall quality and effectiveness of the education system.

Academic Success and Student Well-being: Student performance directly correlates with academic success and overall well-being. High academic achievement is associated with better opportunities for further education, employment, and personal development.

Long-term Impact: Education plays a crucial role in shaping the future of individuals and society as a whole. Monitoring and improving student performance can have long-term positive effects on the personal growth and success of students and contribute to the overall progress of the nation.

Then I started work on this crucial field.



ORGANIZATION OF DISSERTATION REPORT

- **Chapter 1** deals with the introduction of Student Performance, working, applications, techniques, architecture. It also includes the problem statement and objective of dissertation work.
- **Chapter 2** deals with review of literatures related to Student Performance
- **Chapter 3** deals with Theoretical Aspect of Proposed Work.
- **Chapter 4** deals with the details of Methodology and overall system design flow along with experimentation details.
- **Chapter 5** deals with the experimental results and analysis.
- **Chapter 6** contains the conclusion of the work.

INTRODUCTION

❖ The use of machine learning techniques to forecast student performance and pinpoint variables that influence academic achievement or dropout rates has gained popularity in recent years. Academic success is a performance indicator that shows which pupils have reached particular objectives. Numerous stakeholders, including students, professors, and academic institutions, consider it crucial to predict students' success in particular courses or over a whole programme. The most crucial information for this study is student academic data since it provides the most comprehensive picture of the pupils. Machine learning has many different applications. Machine learning is used by search engines to more accurately build relationships between search terms and online pages.

LITERATURE REVIEW

I have read over 40 papers throughout my investigation. In my thesis, I analyzed 30 publications which are connected on machine learning algorithms for predicting student performance based on academic achievement and dropout rates.. Each study used a unique set of methods, was run on a unique set of data, and yielded a unique set of findings. Some researches employed MATLAB library modules, while the most majority relied on machine learning techniques like decision tree, random forest, support vector machine, machine learning, etc.

Name of Publisher	IEEE	International Conference and Journal
No. of Papers studied	10	27



7 papers were eliminated due to lack of data, Same algo. used and a closed research gap.



Finally, 30 papers reviewed (IEEE – 18 and International Conference - 12)

LITERATURE REVIEWS

- **Albreiki, B., Zaki, N., & Alashwal, H. 2021** The improvement of the learning environment requires the use of contemporary techniques, tactics, and applications from educational data mining. The most recent study provides practical methods for assessing the learning environment of students by reviewing and exploiting educational data using machine learning and data mining methodologies. Modern academic institutions operate in a very competitive and complex environment. Universities usually struggle with performance assessment, excellent education, performance assessment techniques, and future endeavors.
- **M. Chitti, P. Chitti and M. Jayabalan 2020** with the assistance of different Education Data Mining methodologies (EDM), the education sector is developing and expanding at an exponential rate, bringing new and enhanced options to the learning community. This research examines EDM with an emphasis on the variables impacting students' predictions, different algorithms applied, and gaps found. The paper also sheds light on how the prediction model's "black-box" judgments are produced, how different eXplainable AI (XAI) methodologies help to make the model's outcomes understandable, and how they help to provide results that are easy to explain.

LITERATURE REVIEW (CONT...)

Altabrawee, Hussein Osama & Qaisa & Ali, r, Samir. (2019) This study focuses particularly on the impact of utilizing the internet as a learning tool and the impact of students' use of social media on their academic achievement. The classification accuracy and the ROC index performance metric have been used to compare the models. Additionally, many metrics like the classification error, accuracy, recall, and the F measure have been calculated. The ANN (completely connected feed forward multilayer ANN) model produced the greatest results, measuring 0.807 in terms of performance, and 77.04% in terms of classification accuracy. In addition, the decision tree model revealed five elements as significant influences on students' performance

LITERATURE REVIEW (CONT...)

- **Hassan, H., Anuar, S., Ahmad, N.B. (2019)** Eight distinct group models were developed by combining five multi-classifiers—Random Forest, Bagging, AdaBoost, Stacking, and Majority Vote classifier—with three base-classifiers—Decision Tree, Artificial Neural Network, and Support Vector Machine. Additionally, academic, demographical, economic, and behavioral e-learning aspects were utilized in this study. The best accuracy of the classifier model was then optimized. A new model for forecasting student performance was subsequently developed. The outcome demonstrates that employing a meta-classifier model with optimized hyper parameters and combining demographics and behavior gave greater accuracy to predict students' success.
- **M. Nagy and R. Molontay (2018)** The models are based on data from 15,825 first-year students who registered at Budapest University of Technology and Economics between 2010 and 2017 and either graduated or left school. To deal with the issue of missing data, we employ imputation. After completing feature extraction and feature selection, a variety of classifiers, including Decision Tree-based methods, Naive Bayes, k-NN, Linear Models, and Deep Learning with varying input parameters, have been trained. The techniques were evaluated using tenfold cross-validation, and the top models were Gradient Boosted Trees and Deep Learning, with AUCs of 0.808 and 0.811, respectively.

01.	[Varma et al., 2021]	Designed to help the country ready its workforce to keep pace with the tremendous growth in the economy	'Along with a comprehensive re-vamping of the education set-up the policy also calls for a new look at human resource systems, such as performance management and training	
02.	[Mazumdar et al., 2022]	Voluntary (e.g., extra-role) employee actions toward companies are also the topic of this study since they are anticipated	when employees' expectations match the amount of fulfillment they experience in their psychological contract.	
03.	[Freire & Pieta, 2022]	There are two parts to this paper: The first part of the research analyzes the correlations between workers' psychological contract fulfillment, organizational citizenship behaviors, and their job happiness.	The second part of the paper focuses on creating a prediction model for assessing STP in the future, with the goals being (a) to understand the impact of STP on employee expectations, satisfaction, and behaviors, (b) to highlight the importance of machine learning techniques in understanding these relationships in an Asian context	
04.	Susomrith, 2020	That is to say, people often engage in social trade after carefully weighing the pros and cons of doing so. For an employment relationship to flourish, both the employee and the employer must believe that they are benefiting from the other's activities.	Employees are very perceptive to their employers' treatment of them, and they tend to respond positively when they perceive that their employers are making an effort to invest in them.	

Research objectives

The following are the primary goals of this study:

- To research different prediction methods and methodologies, and to evaluate their effectiveness using certain criteria.
- To evaluate the methods currently used to predict student performance based on academic success or dropout rates.
- To propose an improved ensemble model for student performance prediction using machine learning techniques on the basis of academic success or dropout rates.
- To find the experimental results and compare and analyze the results with the existing approaches.

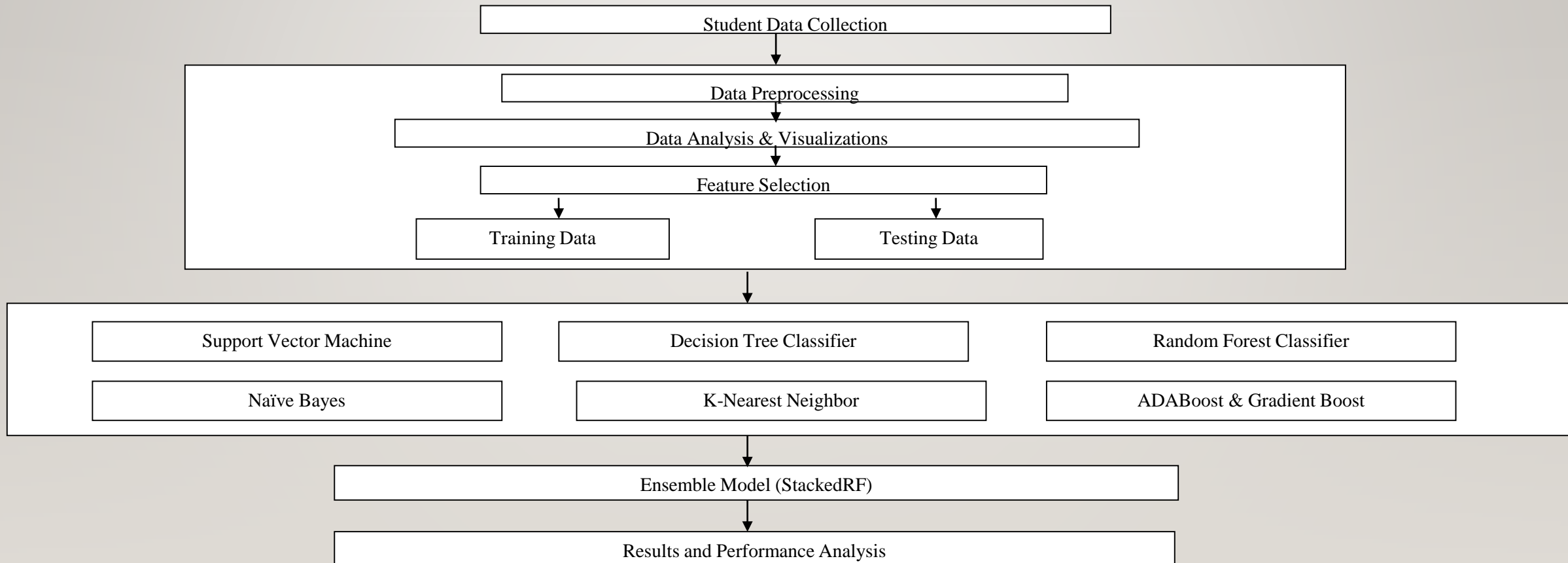
METHODOLOGY

- 1. DATA COLLECTION :** The paper's data came from a collection of student records kept by the institution. This step focuses on choosing the subset of all accessible data that you will be using. It's ideal for ML problems to start with a lot of data (examples or observations) for which you already know the ideal outcome.
- 2. DATA PREPROCESSING :** Data conversion and format transformation are the two processes that pre-processing often entails. The format translation operation is mostly carried out when any type of data is present, such as when.xls files are converted into.csv files. Second, data transformation is calculated with the assumption that the dataset contains data in many formats, which are afterwards changed into accurate forms.
- 3. FEATURE SELECTION :** The attributes that are considered for the feature selection are those which are required for effective prediction of student performance based on the academic success or dropout rates.

METHODOLOGY (CONT...)

- **COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS:** Machine learning algorithms are used by programs to find hidden patterns in data, anticipate outcomes, and improve performance based on prior performance. For different tasks in machine learning, a variety of algorithms may be used, such as simple linear regression for prediction problems like stock market forecasting and the KNN algorithm for classification problems.
- We choose that algorithm for further improvement of the performance metrics. The following is the detailed description of those algorithms which we have compared in this research.

PROPOSED MODEL



DATA COLLECTION

Marital status: The marital status of the student. (Categorical)
Application mode: The method of application used by the student. (Categorical)
Application order: The order in which the student applied. (Numerical)
Course: The course taken by the student. (Categorical)
Daytime/evening attendance: Whether the student attends classes during the day or in the evening. (Categorical)
Previous qualification: The qualification obtained by the student before enrolling in higher education. (Categorical)
Nationality: The nationality of the student. (Categorical)
Mother's qualification: The qualification of the student's mother. (Categorical)
Father's qualification: The qualification of the student's father. (Categorical)
Mother's occupation: The occupation of the student's mother. (Categorical)
Father's occupation: The occupation of the student's father. (Categorical)
Displaced: Whether the student is a displaced person. (Categorical)
Educational special needs: Whether the student has any special educational needs. (Categorical)
Debtor: Whether the student is a debtor. (Categorical)
Tuition fees up to date: Whether the student's tuition fees are up to date. (Categorical)
Gender: The gender of the student. (Categorical)
Scholarship holder: Whether the student is a scholarship holder. (Categorical)
Age at enrollment: The age of the student at the time of enrollment. (Numerical)
International: Whether the student is an international student. (Categorical)
Curricular units 1st sem (credited): The number of curricular units credited by the student in the first semester. (Numerical)
Curricular units 1st sem (enrolled): The number of curricular units enrolled by the student in the first semester. (Numerical)
Curricular units 1st sem (evaluations): The number of curricular units evaluated by the student in the first semester. (Numerical)
Curricular units 1st sem (approved): The number of curricular units approved by the student in the first semester. (Numerical)

REQUIREMENTS

❖ Software Requirements

- Google Colab using Python (Version 3.6.9).
- Coding Language – MATLAB 2021
- IDE - Visual Studio code

RESULTS AND DISCUSSION

- ❖ Graphical User Interface Presentation of project
- ❖ Showing Result
- ❖ Advantages

PERFORMANCE ANALYSIS OF ALGORITHMS

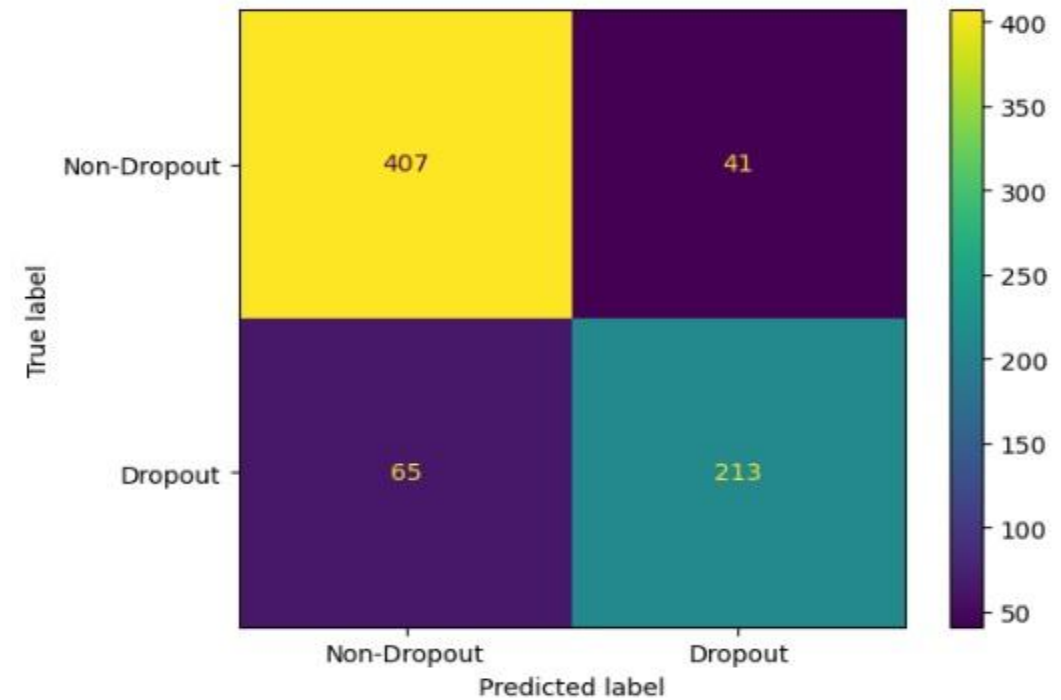
NAÏVE BAYES CLASSIFICATION REPORT

```
*****
Classification Report
*****
      precision    recall  f1-score   support

     0       0.86      0.91      0.88        448
     1       0.84      0.77      0.80        278

 accuracy          0.85          726
 macro avg       0.85      0.84      0.84          726
 weighted avg    0.85      0.85      0.85          726

*****
```



PERFORMANCE ANALYSIS OF ALGORITHMS CON.

LOGISTIC REGRESSION CLASSIFICATION REPORT

```
Precision : 0.9201101928374655
Recall : 0.9201101928374655
Accuracy : 0.9201101928374655
F1 Score : 0.9201101928374655
```

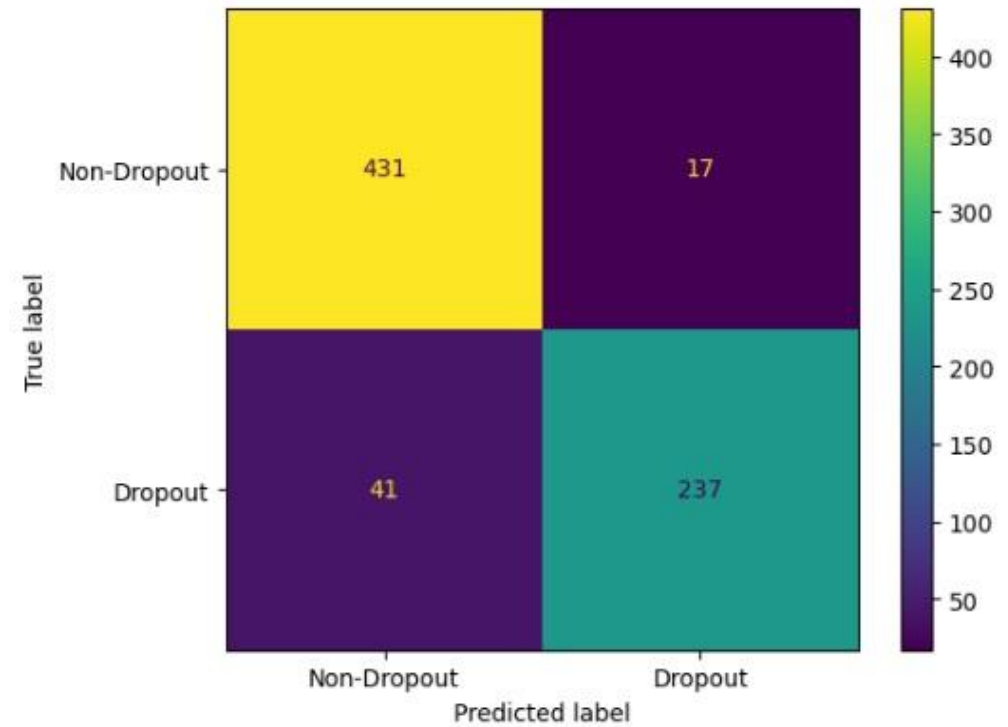
```
[[431 17]
 [ 41 237]]
```

```
*****
                        Classification Report
*****
      precision    recall  f1-score   support

     0         0.91      0.96      0.94         448
     1         0.93      0.85      0.89         278

 accuracy          0.92          0.92         726
 macro avg         0.92      0.91      0.91         726
 weighted avg      0.92      0.92      0.92         726

*****
```



PERFORMANCE ANALYSIS OF ALGORITHMS CON.

RANDOM FOREST CLASSIFICATION REPORT

```
Precision : 0.9201101928374655
Recall : 0.9201101928374655
Accuracy : 0.9201101928374655
F1 Score : 0.9201101928374655
```

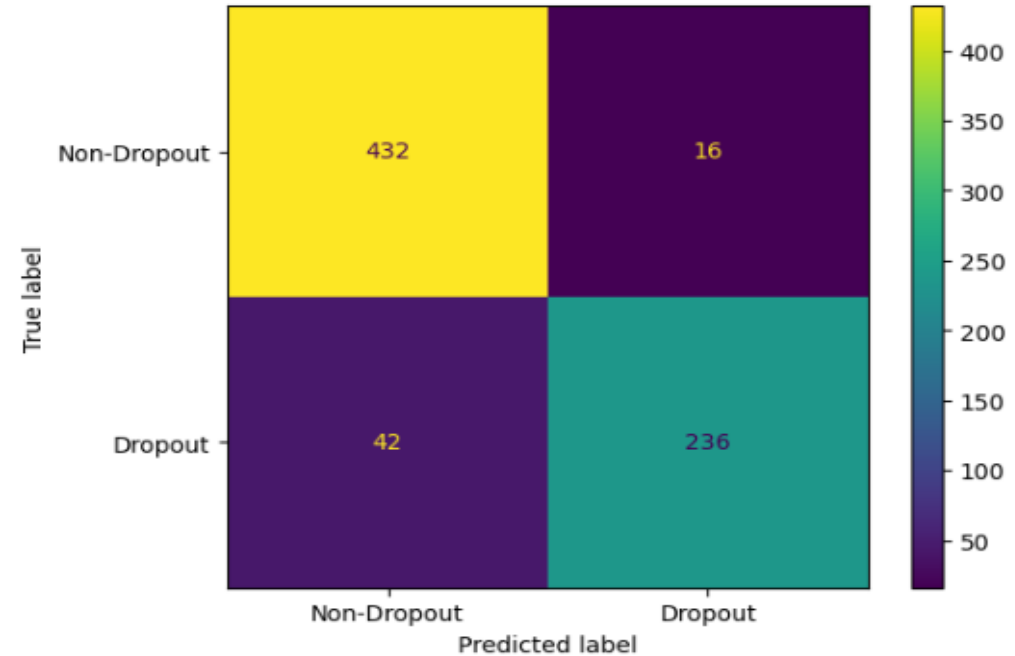
```
[[432 16]
 [ 42 236]]
```

```
*****
                        Classification Report
*****
      precision    recall  f1-score   support

     0         0.91      0.96      0.94        448
     1         0.94      0.85      0.89        278

 accuracy          0.92          726
 macro avg         0.92      0.91      0.91        726
 weighted avg      0.92      0.92      0.92        726

*****
```



PERFORMANCE ANALYSIS OF ALGORITHMS CON.

XGBOOST CLASSIFICATION REPORT

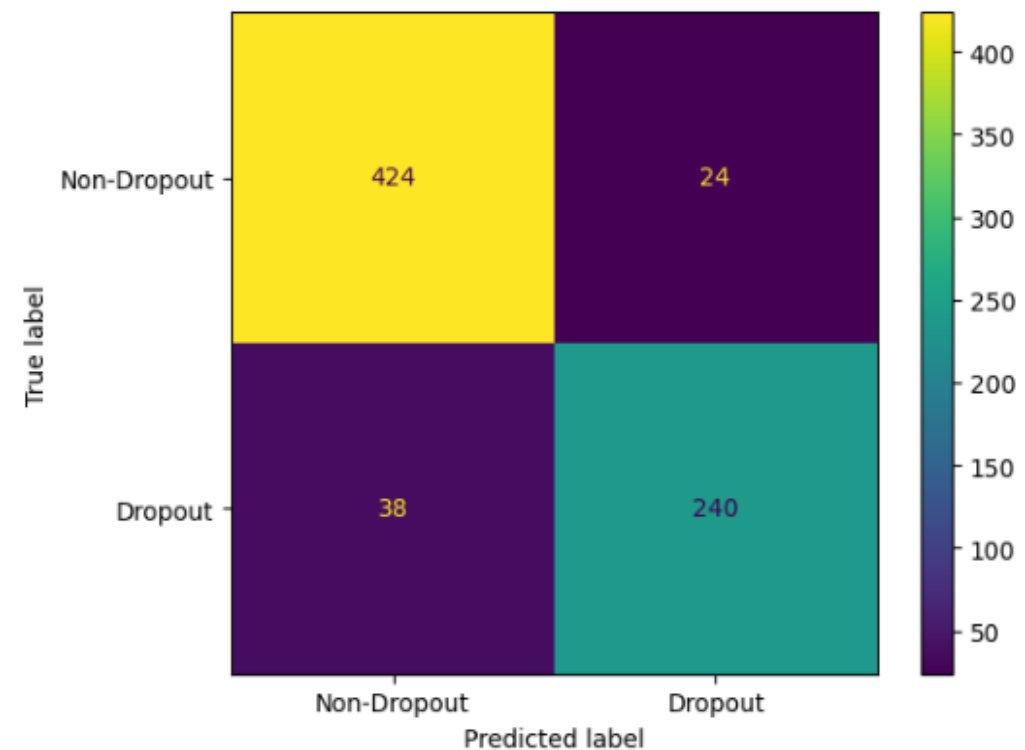
```
Precision : 0.9146005509641874
Recall : 0.9146005509641874
Accuracy : 0.9146005509641874
F1 Score : 0.9146005509641874
```

```
[[424 24]
 [ 38 240]]
```

```
*****
Classification Report
*****
```

	precision	recall	f1-score	support
0	0.92	0.95	0.93	448
1	0.91	0.86	0.89	278
accuracy			0.91	726
macro avg	0.91	0.90	0.91	726
weighted avg	0.91	0.91	0.91	726

```
*****
```



PERFORMANCE ANALYSIS OF ALGORITHMS CON.

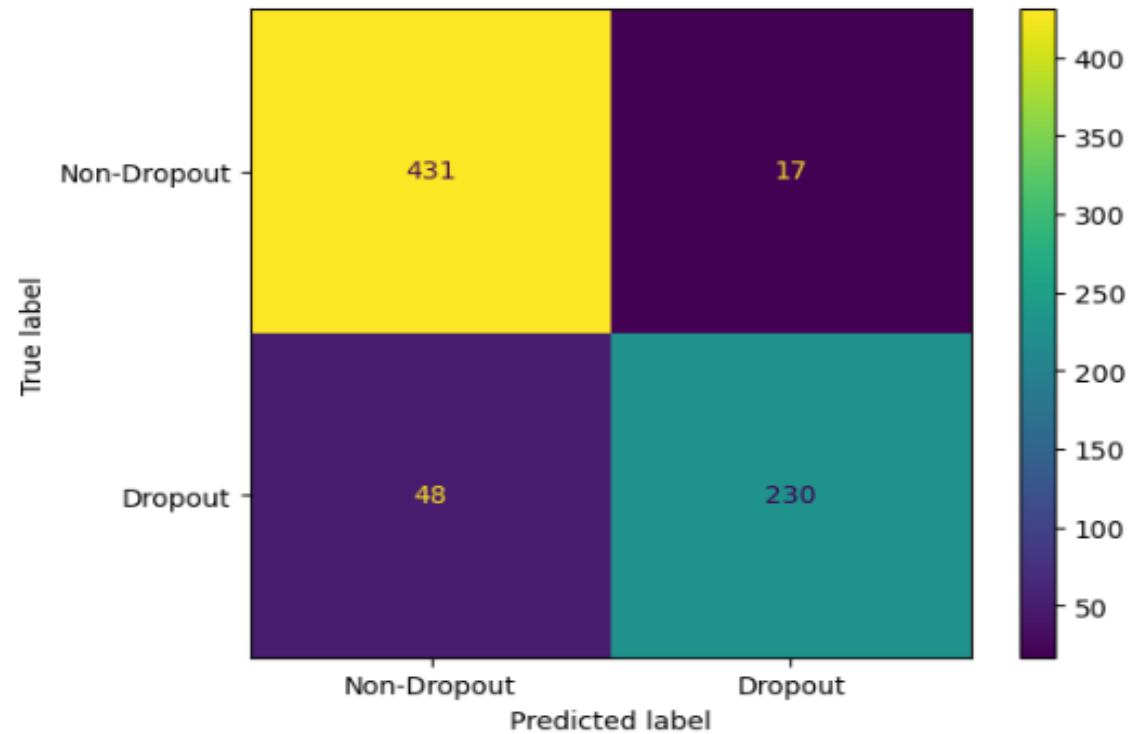
SUPPORT VECTOR MACHINE CLASSIFICATION REPORT

```
Precision : 0.9104683195592287
Recall : 0.9104683195592287
Accuracy : 0.9104683195592287
F1 Score : 0.9104683195592287
```

```
[[431 17]
 [ 48 230]]
```

Classification Report

	precision	recall	f1-score	support
0	0.90	0.96	0.93	448
1	0.93	0.83	0.88	278
accuracy			0.91	726
macro avg	0.92	0.89	0.90	726
weighted avg	0.91	0.91	0.91	726



PERFORMANCE ANALYSIS OF ALGORITHMS con.

Multi Layer Perceptron Classification report

Precision : 0.8994490358126722

Recall : 0.8994490358126722

Accuracy : 0.8994490358126722

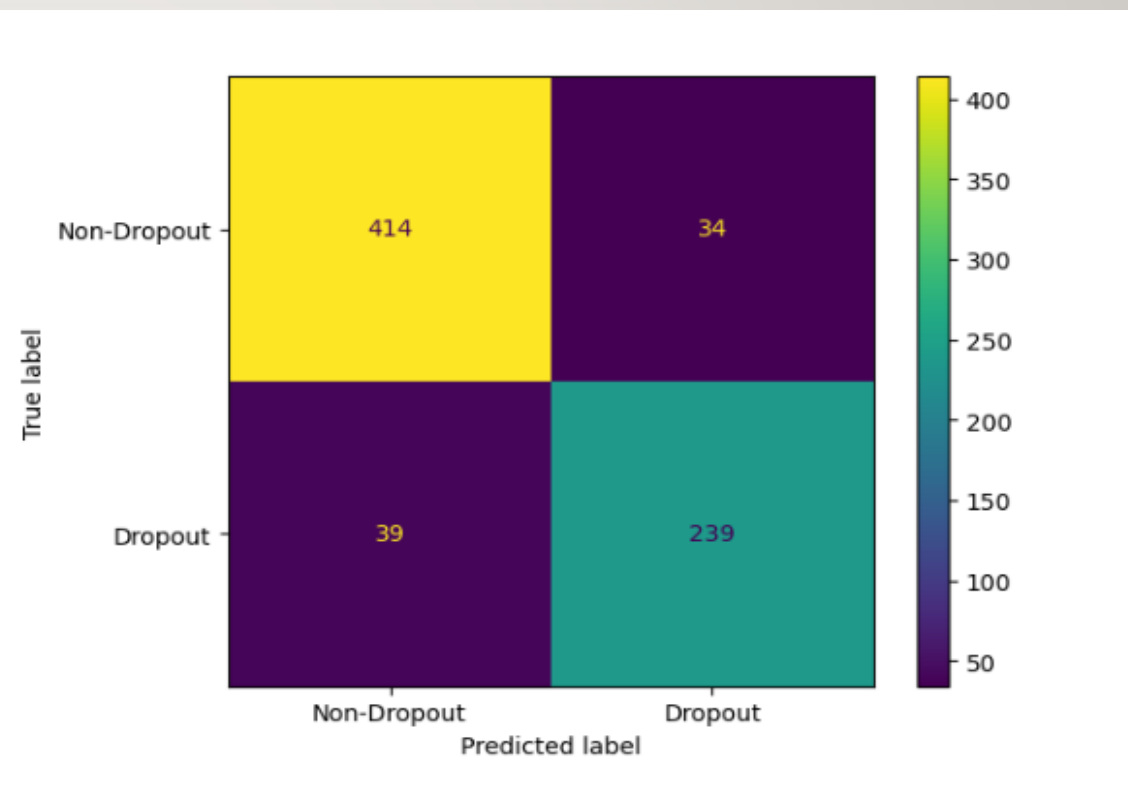
F1 Score : 0.8994490358126722

[[414 34]

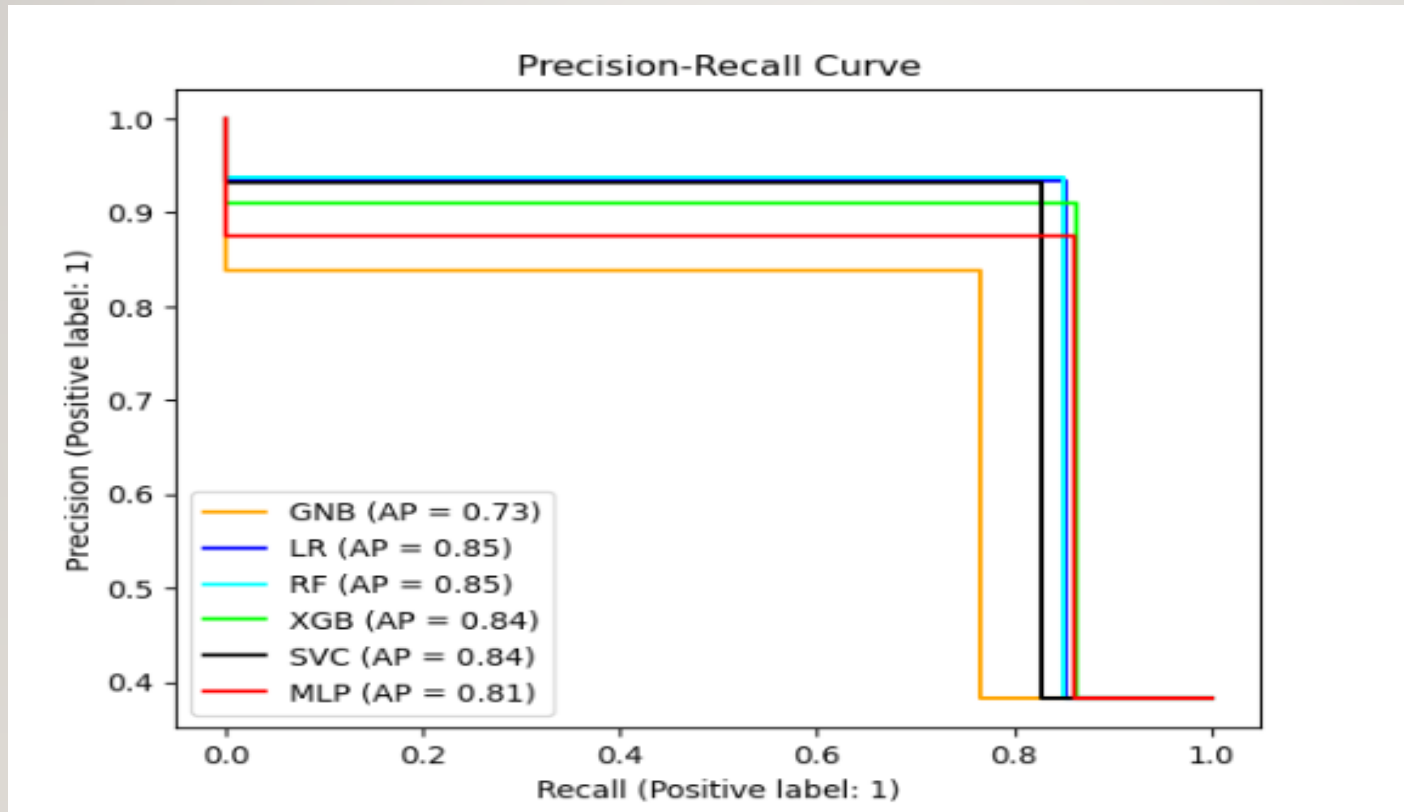
[39 239]]

Classification Report

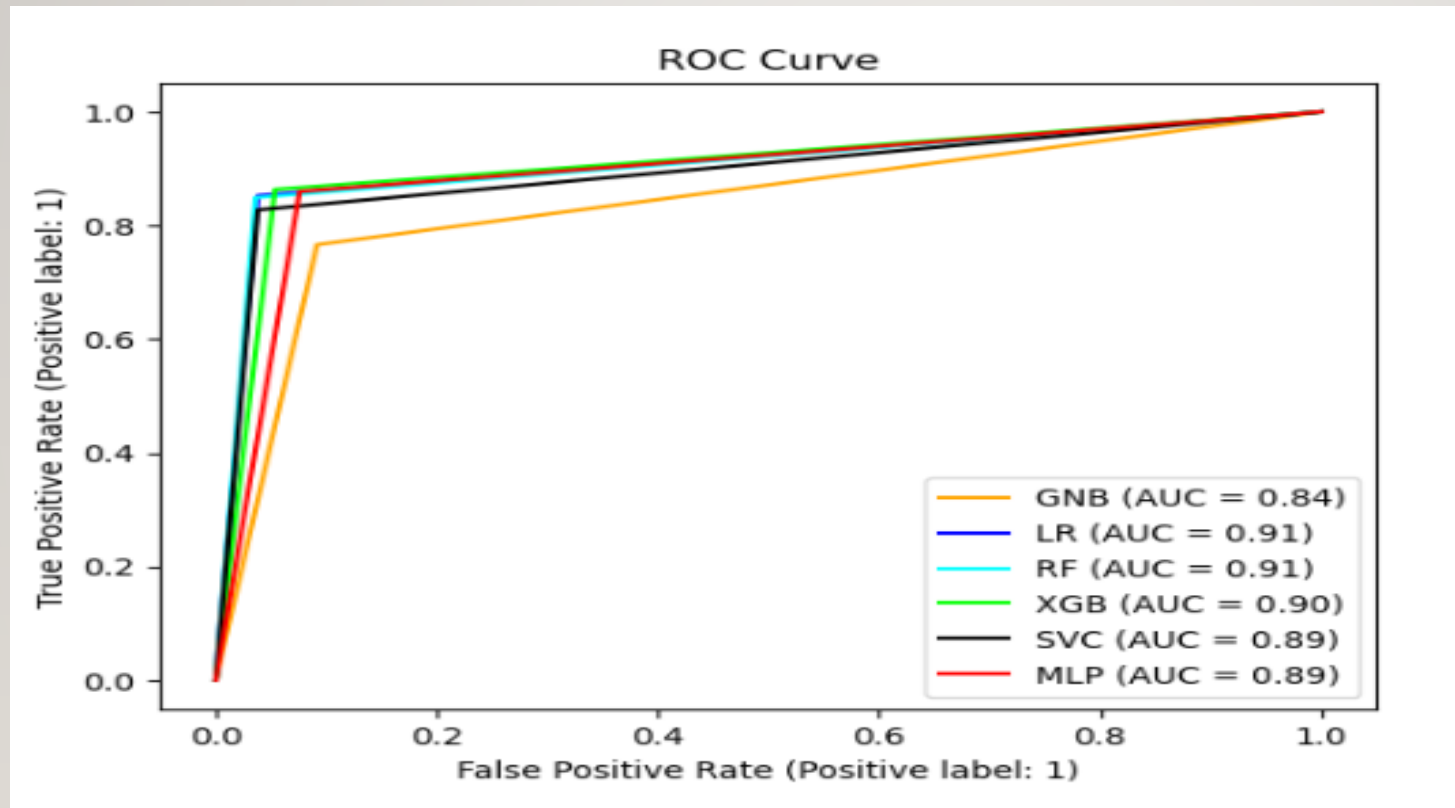
	precision	recall	f1-score	support
0	0.91	0.92	0.92	448
1	0.88	0.86	0.87	278
accuracy			0.90	726
macro avg	0.89	0.89	0.89	726
weighted avg	0.90	0.90	0.90	726



COMPARATIVE PERFORMANCE EVALUATION



COMPARATIVE PERFORMANCE EVALUATION



CONCLUSION

- The objective of this study is to utilizing machine learning to predict student performance based on academic achievement or dropout rates . It has enormous potential to enhance educational results. It may be used to find at-risk pupils, provide interventions that are specific to them, and personalize education. Addressing this approach's drawbacks and difficulties is necessary, though, since these include biases in training data, privacy issues, and the requirement for human experience. We can leverage the advantages of predictive analytics to build a more efficient and inclusive educational system by fusing the strength of machine learning with the knowledge and expertise of educators.

FUTURE SCOPE

The prediction model offers useful information that may guide the development of evidence-based policies and programs to lower dropout rates and improve educational equity. Overall, by utilizing machine learning approaches, this research advances the subject of predicting student performance and lays the groundwork for future developments. The model incorporates the forecasts from other trained models when using an average ensemble, such as Random Forest

PUBLISHED PAPER



**SARDAR PATEL INSTITUTE OF
ECONOMIC AND SOCIAL RESEARCH**

anveshak

A bi-annual journal

CERTIFICATE OF PUBLICATION

This is to certify that the paper entitled

**Comparative Evaluation of Machine Learning Algorithms for Predicting Student
Performance Based on Academic Achievement and Dropout Rates**

Authored by

Garima Bhatia

M.Tech Scholar in Poornima University, Jaipur

ज्ञान-विज्ञान विमुक्तये

SUBMITTED PAPER

ISSN: 2348 - 2117



International Journal of Engineering Technology and Computer Research (IJETCR)

Available Online at www.ijetcr.org

Volume 11; Issue 01; January-February; 2023; Page No. 36-43

Prediction of Student Academic Performance using ML

Garima Bhatia, Dr. Vishnu Sharma

M.Tech. Scholar in Poornima University, Jaipur

Professor in Poornima University, Jaipur

Conflicts of interest: Nil

Corresponding author: Garima Bhatia

PLAGIARISM REPORT

All Chap

ORIGINALITY REPORT

11 %
SIMILARITY INDEX

7 %
INTERNET SOURCES

5 %
PUBLICATIONS

7 %
STUDENT PAPERS

PRIMARY SOURCES

1	www.researchgate.net Internet Source	2 %
2	www.javatpoint.com Internet Source	1 %
3	Submitted to Sreenidhi International School Student Paper	1 %
4	Submitted to Staffordshire University Student Paper	1 %
5	Submitted to University of Bolton Student Paper	1 %
6	Submitted to Queen's University of Belfast Student Paper	1 %
7	Submitted to Babes-Bolyai University Student Paper	1 %

REFERENCES

- [1]Albreiki, B., Zaki, N., & Alashwal, H. (2021).A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques. *Education Sciences*, 11(9), 552.
<https://doi.org/10.3390/educsci11090552>
- [2]Harikumar Pallathadka,Alex Wenda, Edwin Ramirez-Asís, Maximiliano Asís-López, Judith Flores-Albornoz, Khongdet Phasinam (2023). Classification and prediction of student performance data using various machine learning algorithms. *Materials Today: Proceedings*, Volume 80, Part 3, 2023, Pages 3782-3785, ISSN 2214-7853,
<https://doi.org/10.1016/j.matpr.2021.07.382>.
- [3]M. Chitti, P. Chitti and M. Jayabalan, "Need for Interpretable Student Performance Prediction," 2020 13th International Conference on Developments in eSystems Engineering (DeSE), Liverpool, United Kingdom, 2020, pp. 269-272, doi: 10.1109/DeSE51703.2020.9450735.

- [4]Chen,Y., Zhai, L.A comparative study on student performance prediction using machine learning. Educ Inf Technol (2023). <https://doi.org/10.1007/s10639-023-11672-1>
- [5]Altabrawee, Hussein Osama & Qaisa& Ali, r, Samir. (2019). Predicting Students' Performance Using Machine Learning Techniques. JOURNAL OF UNIVERSITY OF BABYLON for pure and applied sciences. 27. 194-205. 10.29196/jubpas.v27i1.2108.
- [6]Hamsa, Hashmia & Indiradevi, Simi & Kizhakkethottam, Jubilant. (2016). Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm. Procedia Technology. 25. 326-332. 10.1016/j.protcy.2016.08.114.
- [7]Chen, Y., Zhai, L. A comparative study on student performance prediction using machine learning. Educ Inf Technol (2023). <https://doi.org/10.1007/s10639-023-11672-1>



- [8]. Bolden-Barrett, V. (2017). Turnover costs employers \$15,000 per worker. <https://www.hrdiver.com/news/study-turnover-costs-employers-15000-per-worker/449142/> (accessed on January 1, 2021).
- [9]. Burkov, A. (2019). The Hundred-Page Machine Learning Book, ISBN-13: 978-1999579500.
- [10]. Conlon, S. J. (2021). Why Do Data Scientists Want to Change Jobs: Using Machine Learning Techniques to Analyze Employees' Intentions in Switching Jobs. INTERNATIONAL JOURNAL OF MANAGEMENT & INFORMATION TECHNOLOGY, 16, 59–71. <https://doi.org/10.24297/ijmit.v16i.9058>.
- [11]. Fallucchi, F., Coladangelo, M., Giuliano, R., & William De Luca, E. (2020). “Predicting Employee Attrition Using Machine Learning Techniques.” Computers 9(4), 86. <https://doi.org/10.3390/computers9040086>.

- [12]Altabrawee, Hussein Osama & Qaisa& Ali, r, Samir. (2019). Predicting Students' Performance Using Machine Learning Techniques. JOURNAL OF UNIVERSITY OF BABYLON for pure and applied sciences. 27. 194-205. 10.29196/jubpas.v27i1.2108.
- [13]Hamsa, Hashmia & Indiradevi, Simi & Kizhakkethottam, Jubilant. (2016). Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm. Procedia Technology. 25. 326-332. 10.1016/j.protcy.2016.08.114.
- [14]Chen, Y., Zhai, L. A comparative study on student performance prediction using machine learning. Educ Inf Technol (2023). <https://doi.org/10.1007/s10639-023-11672-1>
- [15]Altabrawee, Hussein Osama & Qaisa& Ali, r, Samir. (2019). Predicting Students' Performance Using Machine Learning Techniques. JOURNAL OF UNIVERSITY OF BABYLON for pure and applied sciences. 27. 194-205. 10.29196/jubpas.v27i1.2108.



THANK YOU

- ANY QUERY