iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: II Month of publication: February 2025

DOI: https://doi.org/10.22214/ijraset.2025.67144

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue II Feb 2025- Available at www.ijraset.com

Identification and Minimization of Students Academics Failure Using Machine Learning Techniques

Kirti Ambekar[1], Snehal Matade[2], Shruti Deshmukh[3], Arohi Birari[4], Prof. K. B. Dhomse[5]
[1, 2, 3, 4]Student of BE. Dept. of Information Technology MET-IOE, Nashik, India
[5]Internal Guide Dept. of Information Technology MET's Institute of Engineering, Nashik, India

Abstract: Early identification of students at risk of academic failure is critical for ensuring timely support and improving educational outcomes. Despite the availability of extra assignments and projects, many students struggle academically without early intervention. This project aims to address this issue by developing a predictive model that identifies high school students at risk of failing before the end of the academic period. By accurately predicting these students, teachers can provide additional assistance and resources to prevent failure and enhance overall student performance. The core objective of this project is to increase student success by offering targeted support to those identified as at risk. We will analyze various factors influencing student performance, such as academic history, behavior, and engagement, and use this data to develop a model that can predict potential failures early. By providing insights into students' risk levels, educators will be equipped to take timely actions, such as offering extra learning materials and personalized interventions, to improve academic outcomes. This project emphasizes the importance of early intervention in education and aims to reduce student failure rates.

I. INTRODUCTION

Early identification of students at risk of academic failure is critical for ensuring timely support and improving educational outcomes. Despite the availability of extra assignments and projects, many students struggle academically without early intervention. This project aims to address this issue by developing a predictive model that identifies high school students at risk of failing before the end of the academic period.

By accurately predicting these students, teachers can provide additional assistance and resources to prevent failure and enhance overall student performance. The core objective of this project is to increase student success by offering targeted support to those identified as at risk. We will analyze various factors influencing student performance, such as academic history, behaviour, and engagement, and use this data to develop a model that can predict potential failures early. By providing insights into students' risk levels, educators will be equipped to take timely actions, such as offering extra learning materials and personalized interventions, to improve academic outcomes. This project emphasizes the importance of early intervention in education and aims to reduce student failure rates.

II. LITERATURE SURVEY

The literature survey in this research highlights three studies that apply machine learning techniques to predict and reduce student academic failure. The first study by Ahmed Malik et al. (2021) in identifying at-risk students through machine learning models compares algorithms like K-Nearest Neighbours, Decision Trees, and Naive Bayes to identify at-risk students, noting scalability issues and feature independence assumptions. The second study by Sarah Brown et al. (2022) in reducing student dropout using machine learning techniques uses SVM and Logistic Regression to predict dropout risks, facing challenges with data quality and the incorporation of non-academic factors.

The third study by Rahul Patel et al. (2023) in adaptive learning platforms for minimizing student failures integrates machine learning into adaptive learning platforms to customize educational content, but faces high development costs and integration challenges with existing systems. These studies collectively emphasize both the potential of machine learning in educational contexts and the technical and practical limitations that need to be addressed.

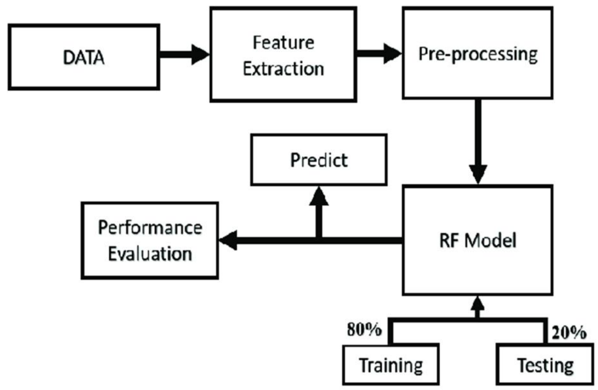
International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue II Feb 2025- Available at www.ijraset.com

III. METHODOLOGY

A. Random Forest Algorithm

Random Forest is a machine learning algorithm that creates "forest" of decision trees to make predictions. The idea behind it is that a group of trees, each slightly different from the others, will perform better as a group than any single tree alone.

Random Forest first builds several decision trees by selecting random samples of data and a random subset of features for each tree. This randomness reduces the risk of overfitting, where a model becomes too specific to its training data and fails on new data. Once the trees are built, each tree makes its own prediction. For classification tasks, Random Forest takes the majority vote among the trees to decide on the final class label. For regression tasks (predicting numbers), it averages the predictions from each tree.



IV. SYSTEM OVERVIEW

A. Hardware Requirements

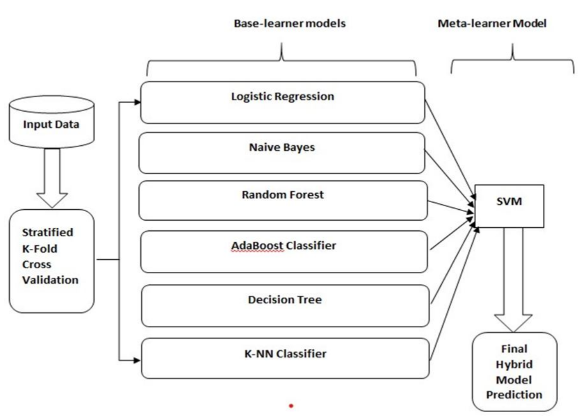
1) Modern PC or laptop (Intel i5/ Ryzen 5, 8GB RAM)
2) Adequate disk space
3) Reliable internet connection

.

B. Software Requirements

1) OS: Windows 10/11, macOS, or Linux
2) Programming Languages: Python, R
3) IDE: PyCharm, VS Code, or Jupyter Notebook
4) Libraries: scikit-learn, TensorFlow, Keras, PyTorch
5) Tools: Pandas, NumPy, Matplotlib
6) Version Control: Git

V. MODULE DESCRIPTION



A. Hybrid Ensemble Model Architecture

1) Base Learner

a) Logistic Regression: Performs binary classification by modeling the probability of a student passing based on features.
b) Naive Bayes: Uses probability estimates assuming feature independence.
c) Random Forest: Aggregates results from multiple decision trees to make predictions.
d) Decision Tree: Splits data based on feature values for classification decisions.
e) AdaBoost: Combines weak classifiers and adjusts weights of misclassified instances.
f) K-Nearest Neighbors (K-NN): Classifies based on majority vote from the nearest neighbors.

2) Meta-Learner

Support Vector Machine (SVM): Takes predictions from base learners and combines them to produce the final prediction by finding the optimaldecision boundary.

3) Stacking Process

a) Training Phase

- Base Learners Training: Each base learner is trained on the original training dataset.
- Creating Meta-Learner Dataset: Use predictions from base learners on a validation set to form a new dataset. This new dataset includes the predictions from base learners as features (D').
- Meta-Learner Training: Train the SVM meta-learner using this new dataset (D').

b) Prediction Phase

- Base Learners Prediction: Each base learner makes predictions on the test dataset.
- Meta-Learner Prediction: Combine the base learners' predictions using the trained SVM to make the final prediction.

c) Validation

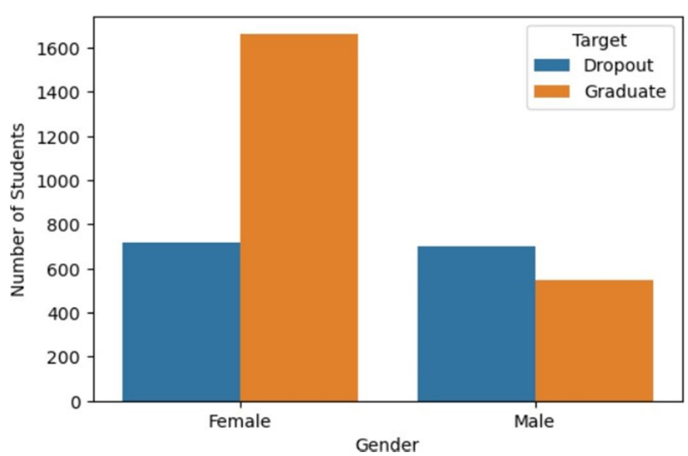
10-Fold Cross-Validation: Used to split the dataset into 10 folds. Each base learner is trained and evaluated on these folds to avoid overfitting and assess model performance.

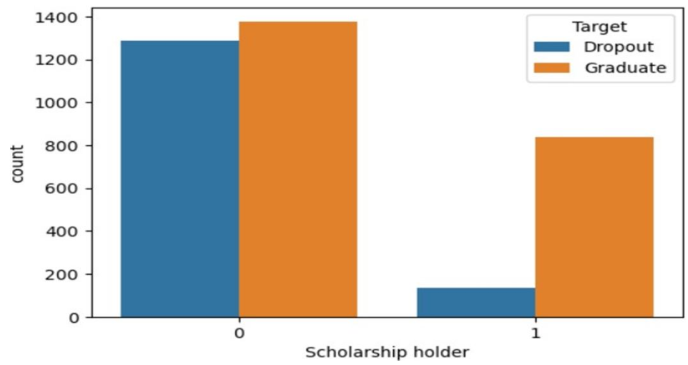
d) Output:

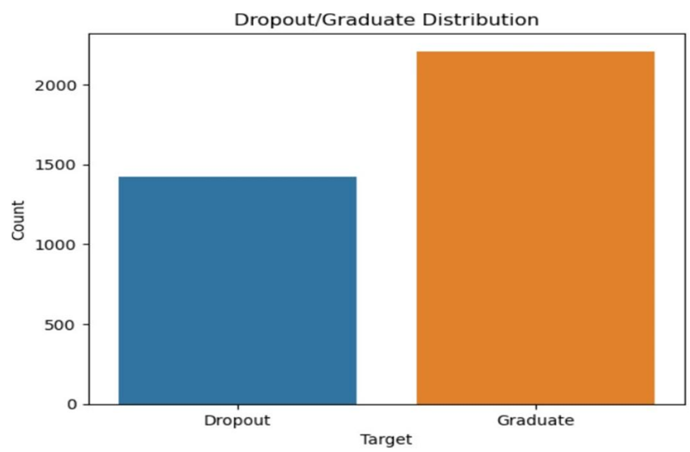
Final Predictions: The ensemble model provides final predictions based on the combined outputs of base learners through the meta-learner.

Design an algorithm that processes resumes one by one and extracts information in a logical order (personal information, education, work history, etc.).

VI. RESULT ANALYSIS







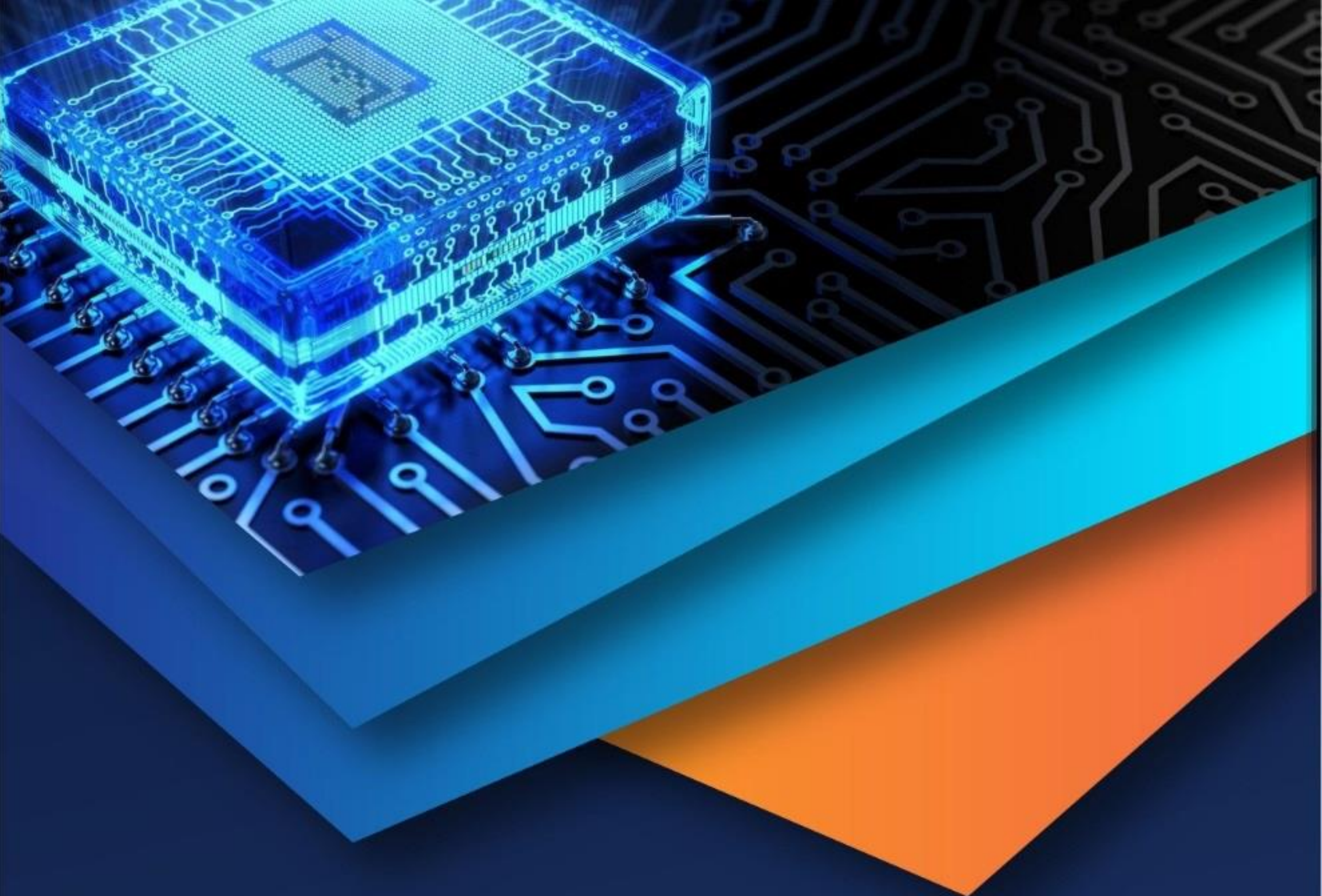

```
Accuracy: 0.92
              precision    recall  f1-score   support

           0       0.92      0.95      0.93       663
           1       0.92      0.87      0.89       426

    accuracy                           0.92      1089
   macro avg       0.92      0.91      0.91      1089
weighted avg       0.92      0.92      0.92      1089
```

VII. CONCLUSIONS

This project aims to develop a machine learning-based system to identify and minimize students' academic failure. By leveraging educational data and machine learning techniques, we can predict academic failure and provide early interventions to support students. The expected outcomes include improved educational outcomes, reduced academic failure rates, and enhanced student success. By leveraging machine learning, this project aims to create a system that predicts academic failure and minimizes it through early identification and targeted interventions. This approach not only improves student performance but also enhances the overall educational experience.

REFERENCES

[1] S. Lee and J. Y. Chung, ''The machine learning-based dropout early warning system for improving the performance of dropout prediction,'' , Jul. 2019.
[2] K. T. Chui, R. W. Liu, M. Zhao, and P. O. D. Pablos, ''Predicting students' performance with school and family tutoring using generative adversarial network-based deep support vector machine , Feb 2020.
[3] M. Adnan, A. Habib, J. Ashraf, S. Mussadiq, A. A. Raza, M. Abid, M. Bashir, and S. U. Khan, "predicting at risk students at different percentages of course length for early intervention using machine learning models , March 2021
[4] J. Berens, K. Schneider, S. Görtz, S. Oster, and J. Burghoff, ''Early detection of students at risk—Predicting student dropouts using administrative student data and machine learning methods,'' CESifo Group Munich, Dec. 30, 2022.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY