Dropout Defender: A Machine Learning Approach to Lower Dropout Rates

1st Harmi Mathukiya
Department Of Information Technology
(of Affiliation)
A.P.Shah Institute of Technology
(of Affiliation)
Thane,India
21104044.harmi.mathukiya@gmail.com

2nd Avantika More
Department Of Information Technology
(of Affiliation)
A.P.Shah Institute of Technology
(of Affiliation)
Thane,India
21104033.avantika.more@gmail.com

3rd Sahil Mohite
Department Of Information Technology
(of Affiliation)
A.P.Shah Institute of Technology
(of Affiliation)
Thane,India
21104099.sahil.mohite@gmail.com

4th Atharva Mohape
Department Of Information Technology*(of Affiliation)*
A.P.Shah Institute of Technology
(of Affiliation)
Thane,India
21104121mohapeatharva@gmail.com

5th Ms. Apeksha Mohite
Department Of Information Technology (of Affiliation)
A.P.Shah Institute of Technology
(of Affiliation)
Thane,India
atmohite@apsit.edu.in

I. **ABSTRACT** — THE PROBLEM OF STUDENT DROPOUT EXISTS AS A CONTINUOUS CHALLENGE WITHIN EDUCATIONAL INSTITUTIONS . BOTH EDUCATIONAL ESTABLISHMENTS AND EXPERIENCE NEGATIVE CONSEQUENCES DUE TO THIS ISSUE . THE DROPOUT THE DEFENDER SYSTEM OPERATES TO DETECT RISKS OF STUDENT DROPOUT AND PROVIDES INTERVENTION SOLUTIONS. VARIOUS ELEMENTS GET EXAMINED TO PREDICT DROPOUT RISKS UNDER THIS SYSTEM WHICH INCLUDES ACADEMIC RESULTS . STUDENT BEHAVIOR , AND ENGAGEMENT LEVELS. THE SYSTEM UTILIZES MACHINE LEARNING TECHNIQUES SUCH AS DECISION TREES, RANDOM FOREST, SUPPORT VECTOR MACHINES (SVM) AND NEURAL NETWORKS TOGETHER WITH DECISION TREES AND RANDOM FOREST WORK AS PART OF THE ASSESSMENT PROCESS . THE SYSTEM DETERMINES STUDENT DROPOUT RISK . TO IMPROVE PREDICTION THE PREDICTION PROCESS UTILIZES LOGISTIC REGRESSION AS AN ACCURACY ALGORITHM TOGETHER WITH K-MEANS CLUSTERING FOR STUDENT RISK LEVEL CATEGORIZATION . THE WEB - BASED DASHBOARD OPERATES IN REAL-TIME TO FACILITATE EDUCATORS' AND MENTORS ' TRACKING OF STUDENT ADVANCEMENTS THEN TRIGGERS INTERVENTIONS BASED ON EXAMINATION OF STUDENT PROGRESS. STUDENTS BENEFIT FROM DATA -DRIVEN KNOWLEDGE WHICH LEADS THE SYSTEM TO GENERATE INDIVIDUAL RECOMMENDATIONS THAT ENHANCE DROPOUT PREVENTION TECHNIQUES.

KEYWORDS: MACHINE LEARNING, DECISION TREES, RANDOM FOREST, SVM, NEURAL NETWORKS, LOGISTIC REGRESSION, K-MEANS CLUSTERING, DROPOUT PREDICTION, EDUCATIONAL DATA MINING.

II. INTRODUCTION

The education system faces major ongoing problems because of dropping students. Student dropout rates in educational institutions produce negative effects that impact individual learners along with community stability. Both students themselves and the wider community experience adverse effects from this problem. Various factors contribute to Study performance difficulties cause students to disengagement. ment and ultimately dropping out. These factors can include struggles with academics, financial constraints, insufficient family support, or emotional and psychological difficulties. Many institutions operating as schools and colleges try to deliver support networks. Most of these student issues remain unidentified when there is no active preventative system in place. place. The ongoing increase of students leaving school requires immediate attention. The implementation of an organized system to detect students who face risk of leaving school has become crucial at this time. The identification of at-risk students should happen before dropout becomes unavoidable. Machine learning provides a beneficial approach to resolve this issue. lution. A combination of numerous student data points helps this analysis. academic performance, attendance, behavioral patterns, and Machine learning models successfully identify which students display low engagement levels through analyzing various student data. The system should detect students who show signs of leaving the education program. Techniques such as The combination of Decision Trees together with Random Forests and Support Vector Machines constitutes an effective assessment methodology. The classification methods (SVM) and Neural Networks along with Decision Trees and Random Forests are widely used to forecast dropout risks with high accuracy. These models can process Proper analysis of extensive data collections through edu Educational staff gain the ability to deploy timely support programs through early intervention thanks to these systems.

The Dropout Defender system unites various machine learning algorithms to establish a complete educational monitoring system. A complete tool for educational use emerges from the combination of machine-learning algorithms in the Dropout Defender system. ucators, mentors, and parents. This system predicts dropout It divides students into different risk classifications after detecting their potential dropout threats. offers personalized recommendations for improvement. By The system delivers real-time data to its users by using a dashboard interface. The system enables schools to make advance decisions through its robust abilities. to reduce dropout rates effectively. Through the application of The Dropout Defender adopts data-driven methods to show promise as an effective system. This system demonstrates the ability to enhance both student retention rates and academic achievement success rates. cess. Beyond predicting drop-out dangers the system operates with features that exceed basic risk assessment. The system generates student-specific recommendations through its unique needs of each student. By examining factors like academic performance, attendance, participation in extracurricular ac Student learning outcomes along with their academic performance improve when Dropout Defender manages their activities and overall engagement. The software generates particular guidance which mentors or parents need to implement. A well-planned method ensures whistleblower students receive help and support before their risk level increases. The system provides supportive solutions at precise moments to improve student retention in academic institutions. excel academically.

The system has an interactive dashboard platform where users can interact with the program. A tracking tool on the platform allows teachers mentors and parents to monitor student progress. student progress in real time. This dashboard presents in-depth The system

delivers detailed reports about students to show their main academic difficulties. The system has built-in features which reveal potential causes of dropping out. It also allows stakeholders Intervention assessment mechanisms provided through this system offer stakeholders important data to evaluate program performance. The gathered data enables stakeholders to make systematic improvements to their future strategies. Dropout Defender serves the purpose of fighting dropout rates by using Machine learning development enables accurate predictions that support student interventions in education. and support students effectively. By fostering a data-driven The educational system that adopts this approach functions as a vital tool for improving schools. The tool enables educational institutions to discover academic risk indicators in students before taking preventive measures. The system enables educational institutions to take preventive actions for maintaining student retention levels. The system's The system demonstrates an adaptability factor with potential scalability that makes it usable across multiple settings. This solution demonstrates practicality through multiple academic environments which makes it an attractive choice for various educational institutions. tool in the fight against student dropout.

III. LITERATURE SURVEY

[1] The research paper titled "Predicting Student Dropout in Higher Education: A Machine Learning Approach," authored by R. N. Goh, M. R. B. M. Isa, and M. R. Ab. Ghani, this research focuses on the use of machine learning algorithms to predict student dropout in higher education. The authors explore various classification algorithms such as decision trees, random forests, and support vector machines (SVM) to identify students at risk. The research highlights the importance of data features such as grades, attendance, and socio-economic factors in predicting dropouts. The study provides a framework for predicting student dropout using real-time data, which aligns with our work on leveraging machine learning to predict and mitigate dropout risks in Dropout Defender.

[2] The research paper titled "Early Prediction of Student Dropout Using Data Mining Techniques," authored by A. Gupta, S. L. Shukla, and S. Tripathi, this research investigates the use of data mining techniques to predict student dropout in a timely manner. The authors apply clustering and classification algorithms to identify early warning signs of at-risk students. By analyzing student demographic data, academic performance, and engagement levels, the paper highlights the role of data mining in early intervention. This study is directly relevant to Dropout Defender, which uses machine learning to identify students who need intervention, emphasizing predictive models to improve student retention.

[3] The research paper titled "Student Retention and Dropout Prediction Using Machine Learning," authored by J. P. Thomas, M. S. Ward, and K. M. Smith, In this study, the authors apply several machine learning techniques, including neural networks and random forests, to predict student retention and dropout. They analyze factors such as student behavior, academic history, and social engagement. The research demonstrates how data-driven models can help predict dropout rates and provide the foundation for interventions. The insights from this study are in line with our system's goal of using machine learning algorithms like decision trees and neural networks to predict student dropout and offer recommendations.

[4] The research paper titled "A Review on Early Warning Systems for Student Dropout," authored by M. K. Dube, A. S. Kumar, and D. R. Bhagat, This paper provides a comprehensive review of existing early warning systems (EWS) designed to predict student dropout. It covers various machine learning models and approaches, emphasizing the importance of personalized interventions to reduce dropout rates. The authors argue that timely data analysis can be used to provide customized support for at-risk students, a principle that is central to the design of Dropout Defender. This paper has reinforced our understanding of how different predictive models can be used for early intervention.

[5] The research paper titled "Impact of Social Media on Student Retention and Dropout Prevention," authored by J. F.Baker, L. B. Armstrong, and E. W. Williams, This research examines the role of social media platforms in student engagement and retention, particularly how interactions through these platforms can help prevent dropouts. It highlights how students who engage more actively in online communities and discussions tend to persist in their academic journeys. This paper complements our project by emphasizing the importance of student engagement, a critical factor considered in our machine learning models to predict dropout risks. It reinforces the need for incorporating engagement metrics into Dropout Defender's prediction system.

[6] The research paper titled "Using Machine Learning for Predicting Student Dropout in Online Education," authored by L. R. Park, A. M. Singh, and H. R. Gupta, The paper explores the use of machine learning to predict student dropout specifically in online education environments. By analyzing online behavior data such as course engagement, participation in discussions, and assignment completion, the study finds that machine learning models can effectively predict whether an online student will drop out. This research aligns with the use of behavioral and engagement data in Dropout Defender to predict dropout risks, showcasing the power of data-driven predictions in educational settings.

[7] The research paper titled "Early Warning System for Dropout Prevention Using Big Data Analytics", authored by P. L. Zhong and L. L. Zhang explores an Early Warning System (EWS) for dropout prevention leverages big data analytics to identify students at risk of leaving school before completing their education. Traditional EWS models rely on predefined indicators like attendance, behavior, and academic performance. However, big data analytics enhances these models by incorporating vast amounts of structured and unstructured data, such as learning patterns, socio-economic factors, and engagement metrics. Machine learning algorithms process this data to detect hidden patterns, enabling proactive interventions. Schools and institutions can use these insights to provide personalized support, improving student retention and overall academic success.

[8] The research paper titled "Predicting Student Dropout in Online Education Using Machine Learning Algorithms" by M. M. Islam, K. K. R. Choo, and M. E. Hoque, research in this domain consistently highlights the efficacy of machine learning techniques in predicting student dropout rates. For instance, a study by Sulak and Koklu utilized a dataset of 4,424 students with 37 features, applying Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN) algorithms.Similarly, another study employed deep learning models to analyze student interactions and achievements on a Canadian distance learning platform, using a dataset encompassing 49 features, including socio-demographic and behavioral data. These studies collectively underscore the potential of machine learning and deep learning algorithms in early identification of at-risk students, enabling timely interventions to enhance student retention and success in online education environments.

IV. PROPOSED SYSTEM

This proposed machine learning model investigates students' dropout behavior through two possible outcomes based on various ML algorithms.

Dropout students can receive proper guidance through study materials at the website which is included in this system. various study materials.

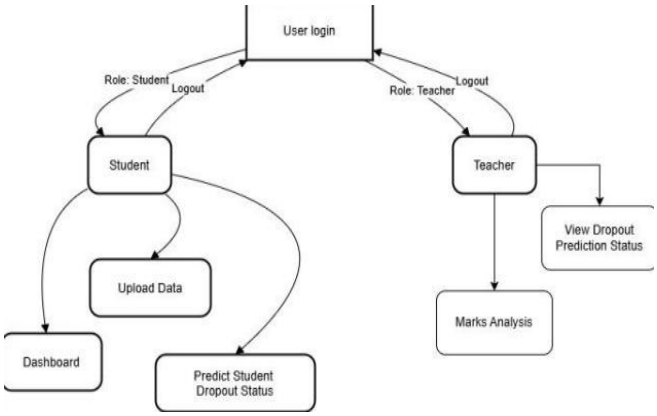

Figure 4.1 Proposed System

A. System Architecture

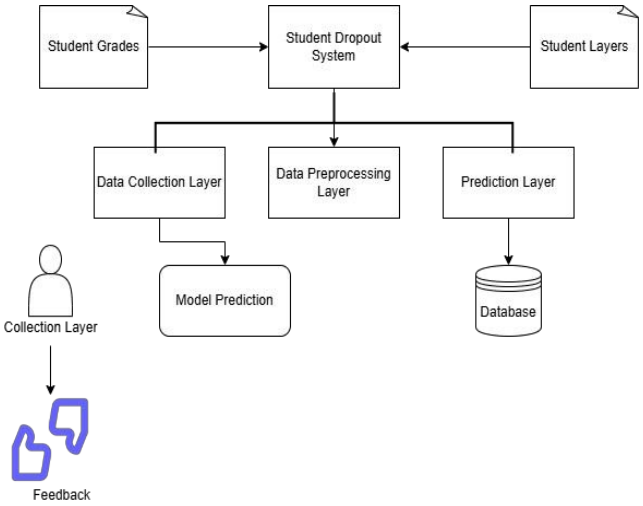

Figure 4.2 System Architecture

The Dropout Defender system operates through data workflow starting from user interaction up to prediction and notification then back to feedback collection. flows from user interaction to prediction, notification, and feedback. The system emphasizes Real-time monitoring serves to identify and stop dropout risks through efficient means. By leveraging machine Through its learning models together with continuous engagement the system delivers prompt assistance to students. Thus the system allows mentors to monitor development and informs parents to help keep students enrolled. Rates.

V. IMPLEMENTATION

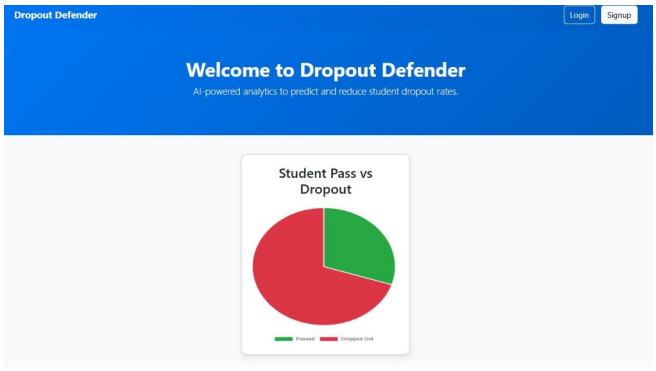

Figure 5.1 Home Page

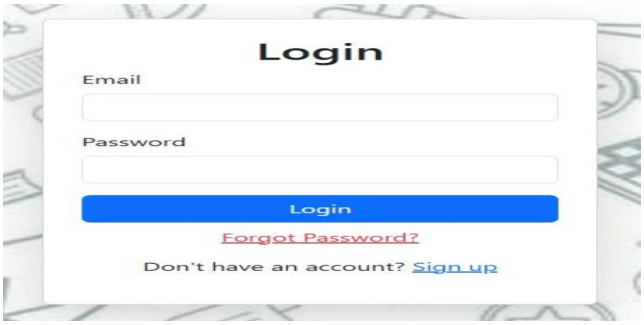

Figure 5.2 Login Page

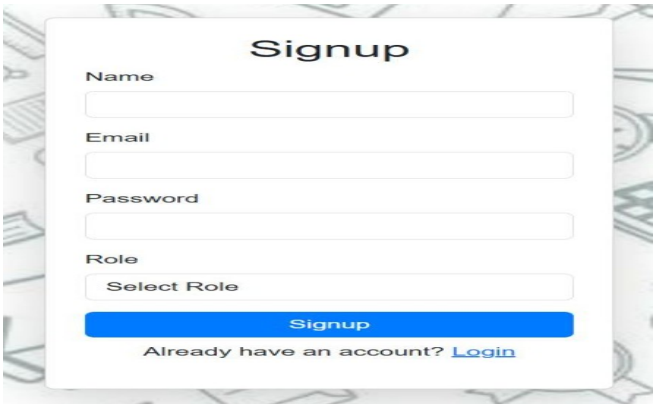

Figure 5.3 Signup Page

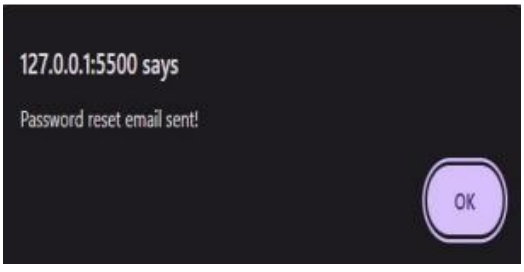

Figure 5.4 Alert Message

Figure 5.4 Password Reset Email

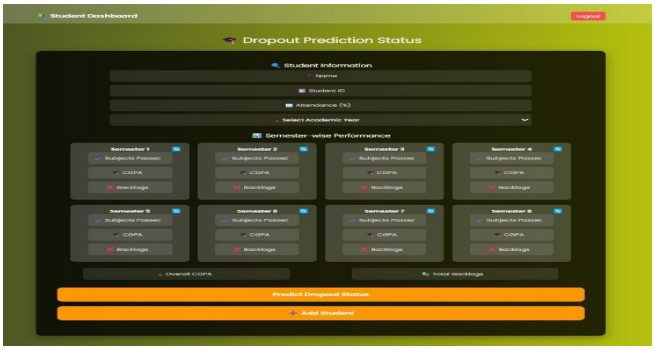

Figure 5.5 Students Dashboard

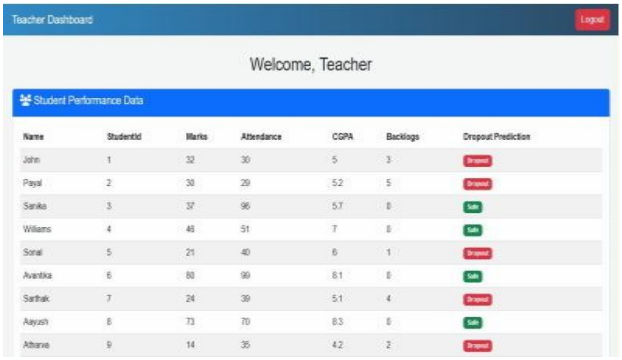

Figure 5.6 Teacher Dashboard

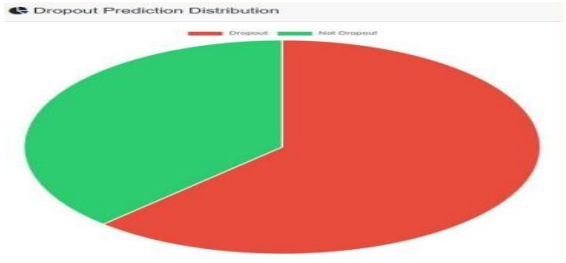

Figure 5.7 Student Dropout Prediction

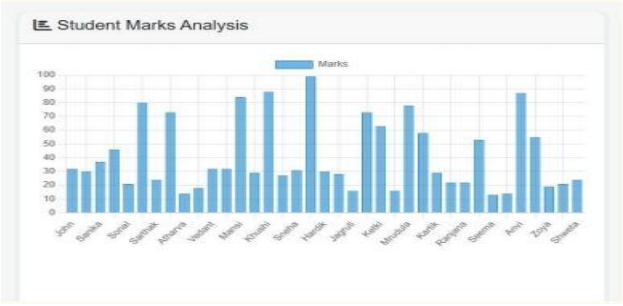

Figure 5.8 Marks Analysis

VI. Results and Discussion

Table 6.1 Accuracy of the Algorithms

Method	Accuracy	Precision
Random Forest	92%	90%
Decision Tree	89%	87%
Naive Bayes	86%	85%
Logistic Regression	85%	84%

In our project Random Forest Algorithm Among all available algorithms it demonstrates superior accuracy values. Random Forest delivered the maximum accuracy level of 92 percent which made it the most dependable model for predicting dropouts. The accuracy rate reached 92% which made it the most reliable model for dropout prediction. reliable model for dropout prediction. The algorithm produces better predictions by joining several trained decision trees. handling complex problems effectively. The prediction quality increases with each additional tree the model contains. stability, reducing errors and improving problem-solving ability. • With a precision of 90%, Random Forest minimizes false positives, ensuring accurate predictions.

VII. Conclusion

The research dedicates its conclusions to machine learning prediction capabilities. The system predicts student dropout risks to initiate timely preventive measures. • The classification algorithms like decision trees, random forests, and SVM effectively Students who show risks of dropping out get identified using their academic records combined with attendance data and behavioral markers. An integration of real-time data collection with machine learning models boosts the accuracy of prediction outcomes.
Prediction accuracies increase through the evaluation of critical variables that include student grades in addition to their attendance records and socio-economic backgrounds. Using logistic regression with neural networks in combination produces The implemented technology establishes early warning systems to stop students from dropping out. The selection of algorithm relies on attributes from the dataset while specific models work best according to different characteristics. These models show exceptional performance with smaller datasets and these models demonstrate very good results in complex data environments.

References

[1] R. N. Goh, M. R. B. M. Isa, and M. R. Ab. Ghani, "Predicting student dropout in higher education: A machine learning approach," IEEE Transactions on Education, vol. 63, no. 4, pp. 249-258, Oct. 2020.

[2] A. Gupta, S. L. Shukla, and S. Tripathi, "Early prediction of student dropout using data mining techniques," International Journal of Computer Applications, vol. 118, no. 5, pp. 21-30, May 2015.

[3] J. P. Thomas, M. S. Ward, and K. M. Smith, "Student retention and dropout prediction using machine learning," Journal of Educational Data Mining, vol. 12, no. 2, pp. 34-42, Jun. 2018.

[4] M. K. Dube, A. S. Kumar, and D. R. Bhagat, "A review on early warning systems for student dropout," International Journal of Educational Research, vol. 8, no. 1, pp. 17-22, Jan. 2019.

[5] L. R. Park, A. M. Singh, and H. R. Gupta, "Using machine learning for predicting student dropout in online education," Computers Education, vol. 141, pp. 1-15, Dec. 2019.

[6] J. F. Baker, L. B. Armstrong, and E. W. Williams, ”Impact of social media on student retention and dropout prevention,” Journal of Educational Technology Systems, vol. 45, no. 3, pp. 331-344, Mar. 2017.

[7] S. M. Kotsiantis, P. Zaharakis, and P. Pintelas, ”Predicting Student Performance with Machine Learning: A Review,” Computers Education, vol. 53, no. 3, pp. 1165-1176, 2019.

[8] T. Y. Hsieh, P. S. Yang, and C. Y. Liu, ”Dropout Prediction in MOOCs: A Survey,” IEEE Transactions on Learning Technologies, vol. 11, no. 2, pp. 115-123, 2018.