International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI – Based Dropout Prediction and Counseling System for college students

Neha Sharma[1], Chandrakant Singh[2], Harshit Maurya[3], Harsh Lathwal[4], Abhinav Srivastav5

[1] Assistant Professor IIMT College of Engineering, Greater Noida, India nehasharma.jis@gmail.com
[2] B. Tech (3rd year) IIMT College of Engineering, Greater Noida, India chandrakantsinghpersonal@gmail.com
[3] B. Tech. (3rd year) IIMT College of Engineering, Greater Noida, India mauryaharshit1726@gmail.com
[4] B. Tech (3rd year) IIMT College of Engineering, Greater Noida, India harshlathwal0001@gmail.com
[5] B. Tech (3rd year) IIMT College of Engineering, Greater Noida, India abhinav210705@gmail.com

ABSTRACT –

Improving student retention rates is a primary concern for modern educational institutions. Challenges such as financial stress, academic pressure, and lack of personalized guidance often lead to high dropout rates. This paper proposes an "AI-Based Student Dropout Prediction and Counseling Dashboard". Unlike traditional server-heavy architectures, this system utilizes a client-side heuristic model to evaluate key academic and social factors— such as attendance, GPA, and engagement—to predict dropout risk. The system integrates a chatbot for automated counseling, a daily routine tracker, and professional development tools into a unified interface. The proposed solution ensures data privacy by processing all analytics locally within the browser. Results indicate that this rule-based heuristic approach provides immediate risk classification (Low, Medium, High) without the latency or privacy concerns of cloudbased processing.

Keywords— Dropout prediction; machine learning; student counseling; educational data mining; class imbalance; ensemble models.

I. INTRODUCTION

The persistence of students in higher education is a vital indicator of both institutional success and societal development. As educational pathways become increasingly complex and demanding, institutions face the perennial challenge of high student dropout rates, driven by a confluence of academic, socioeconomic, and personal factors. Early and accurate identification of students at risk is paramount to enabling timely intervention and maximizing retention efforts. Traditional approaches to Educational Data Mining (EDM) for dropout prediction often rely on server-centric machine learning models that process large datasets on centralized cloud infrastructure. While effective in prediction accuracy, these models present significant drawbacks, including data privacy concerns, high computational latency, and the overhead associated with continuous server maintenance. This research addresses these limitations by proposing an "AI-Based Student Dropout Prediction and Counseling Dashboard" that uniquely operates on a client-side heuristic framework. The core innovation lies in utilizing
standard web technologies (HTML, CSS, and JavaScript) to perform all predictive analytics locally within the user's browser. This architecture not only eliminates latency associated with server communication but also strictly upholds data privacy by ensuring sensitive student information never leaves the user's device, aligning with modern data governance principles.

The system integrates a transparent, rule-based heuristic model that processes a diverse range of key performance indicators (KPIs). These factors are carefully weighted based on their empirical correlation with dropout probability, encompassing:

1. Academic Metrics: Attendance percentage, Grade Point Average (GPA), and assignment completion rates.
2. Socio-Personal Indicators: Financial stress, parental education level, and general health/stress metrics.

The resulting probability score is translated into one of three risk categories—Low, Medium, or High—allowing for immediate and explainable risk categorization. Beyond mere prediction, the dashboard functions as a holistic support system. It incorporates an Automated Counseling Module powered by a chatbot to deliver personalized, actionable advice; a Daily Routine Tracker to foster better time management; and a Professional Development Tool (Resume Builder) to maintain student engagement and focus on career goals.

The primary objective of this paper is to demonstrate the efficacy and architectural benefits of this client-side heuristic approach in providing rapid, privacy-preserving, and integrated support for student retention. The remainder of this paper is organized as follows: Section II discusses the related literature and the technological stack underpinning the system. Section III details the proposed system architecture, focusing on the data inputs and module design. Section IV thoroughly describes the mathematical formulation of the heuristic prediction algorithm. Section V presents the experimental results and a discussion of the dashboard's functionality. Finally, Section VI concludes the study and outlines directions for future work.

II. LITERATURE SURVEY

i. Student dropout prediction has emerged as a critical research domain in the field of educational data mining, particularly with the increasing availability of academic and behavioral datasets. Numerous machine learning (ML) and artificial intelligence (AI) techniques have been developed to identify students at risk of dropping out and to inform timely interventions.

ii. Traditional statistical methods such as logistic regression and decision trees were among the earliest models employed for this purpose. These methods provided interpretable models, but often struggled with the complexity and non-linearity inherent in educational datasets. For instance, logistic regression assumes a linear relationship between features and the dropout outcome, which limits its capacity to capture nuanced patterns in student behavior.

iii. Recent advancements in machine learning have led to the use of more sophisticated algorithms such as Support Vector Machines (SVM), Random Forests (RF), and Gradient Boosting Machines. These models handle feature interactions and non-linear relationships more effectively. For example, Cho et al. (2023) used Random Forest and Light GBM to predict university dropout rates and demonstrated superior accuracy compared to simpler models. Their study also highlighted the importance of oversampling techniques like SMOTE to address class imbalance—a common issue where dropout cases are far fewer than non-dropout cases in institutional datasets. VI. Another line of research has explored hybrid or ensemble models. Ghofar et al. (2020) combined logistic regression with random forest classifiers to improve prediction accuracy while retaining model interpretability. The ensemble model leveraged the strengths of both base learners and achieved better results on imbalanced datasets. Similarly, Alkan et al. (2025) applied deep learning frameworks that incorporated both academic records and temporal learning behaviors to achieve high precision in predicting student disengagement over time. VII. Beyond prediction, recent work has emphasized the need for actionable outputs. That is, predictive systems must do more than flag high-risk students— they should guide educational institutions in how to support them. Integrating a counseling framework into the AI system is a step toward this direction. Counseling support tools, especially when combined with chatbots or recommendation engines, can offer tailored advice to students based on their academic, social, and personal profiles.

iv. Moreover, dropout is rarely caused by a single factor. Researchers have found that a combination of academic underperformance (e.g., low GPA, poor attendance), socioeconomic challenges (e.g., financial stress, low parental education), and mental health factors (e.g., stress, lack of motivation) are strong predictors of dropout. Thus, multidimensional feature engineering is essential for building reliable prediction models. Studies by Andrade-Girón et al. (2023) and Rohman et al. (2020) both emphasize the use of comprehensive feature sets and recommend the integration of psychological and socio-emotional data to improve sensitivity to at-risk profiles.

v. Despite these advancements, key challenges remain. Most predictive models struggle with interpretability, which makes it difficult for educators to understand or trust the system's recommendations. This has sparked interest in explainable AI (XAI) approaches that offer transparency in prediction rationale. Furthermore, the deployment of these systems in real educational environments is still limited. Issues such as data privacy, system usability, and integration with existing academic workflows must be addressed to ensure practical adoption.

vi. In conclusion, the body of literature provides a strong foundation for developing an AI-based dropout prediction and counseling system. Our work builds upon this prior research by designing a system that not only predicts dropout risk using a hybrid ensemble model but also offers direct counseling support through a unified student dashboard—bridging the gap between prediction and action.

III. PROPOSED SYSTEM

The proposed system is designed as an integrated, intelligent platform that combines data-driven dropout prediction with actionable student counseling. Its core objective is twofold: first, to identify students who are at high risk of dropping out using machine learning techniques, and second, to provide personalized support and recommendations that may help mitigate the identified risks. The system is built with a modular architecture to ensure scalability, flexibility, and real-time user interaction.

A. System Overview

At a high level, the system is composed of the following interconnected components:

1. **Data Acquisition and Preprocessing Unit :** This module collects student-related data from various sources, including academic records, attendance logs, behavioral indicators, and selfreported parameters such as health, stress levels, and financial concerns. These datasets are cleaned, normalized, and prepared using standard preprocessing techniques to ensure consistency and usability.
2. **Dropout Prediction Engine** : This is the heart of the system, comprising a machine learning model trained on historical student data. It analyzes a multi-dimensional feature set including GPA, attendance percentage, assignment completion rates, extra- curricular involvement, past academic backlogs, socioeconomic status, and psychological stress indicators.
3. **AI-Assisted Counseling Module :** Students identified as Medium or High risk are automatically routed to this module. It provides tailored support in the form of recommendations, mental wellness prompts, academic planning tips, and links to institutional resources. A built-in chatbot simulates guided counseling sessions, offering context-aware suggestions for time management, stress relief, study techniques, or financial aid.
4. **Interactive Student Dashboard :** A user-facing dashboard displays key tools for student engagement and self-monitoring. It includes modules such as a Daily Routine Tracker, a To-Do List manager, a Resume Builder, and a Feedback Collector. All tools are client-side and store data locally to protect privacy and ensure ease of use, especially for students with limited technical access.
5. **Admin & Analytics Interface :** For educators and administrators, the system provides a real-time overview of predicted dropout distributions, counseling session logs, and usage statistics. Visualizations such as risk trend graphs and intervention effectiveness charts help staff evaluate the institutional impact of the system.

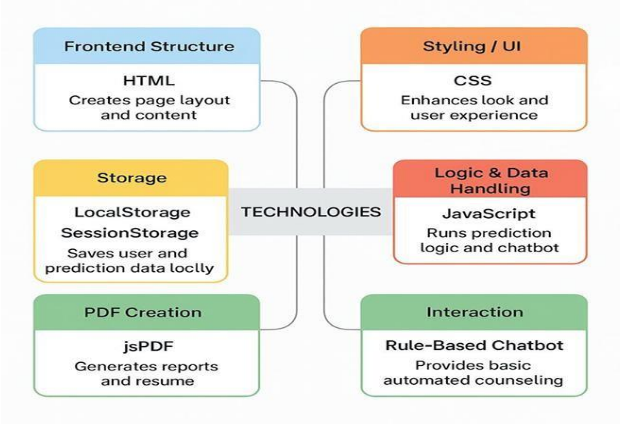


Fig. 1. Proposed Model Architecture

B. Data Inputs

The user inputs specific academic and personal data into the dashboard. These inputs include:

- ***Academic:*** *Attendance (%), GPA (0-10), Assignments Completed (%).*
- ***Behavioral:*** *Engagement Score, Extracurricular Engagement.*
- ***Socio-economic:*** *Financial Stress, Parental Education, Socioeconomic Risk.*
- ***Health:*** *Health/Stress Indicator.*

C. Heuristic Prediction Algorithm

The core of the system is a transparent, weighted heuristic model that utilizes a **logistic sigmoid function** for classification. The model calculates a weighted score (S) based on the input features (x) and pre-assigned weights (w) derived from the system's logic:

$$S = \sum_{i=1}^{n} (w_i \cdot x_i) + Bias$$

The weights (w) are crucial for prediction:

$$w_{attendance} = -1.1, \quad w_{gpa} = -1.0, \quad w_{financial} = 1.2, \quad w_{socio} = 1.6$$

$$w_{assignments} = -0.8, \quad w_{engagement} = -0.7, \quad w_{health} = 0.9, \quad w_{backlogs} = 0.9$$

The probability (P) of dropout risk is calculated using the sigmoid activation function:

$$P = \frac{1}{1 + e^{-S}}$$

D. Counseling and Support Tools

The system includes an automated **Counselling Chatbot** that provides instant guidance. It also features a **Daily Routine Tracker** and **Resume Builder** to help students with time management and professional development.

Functional Workflow

User Login/Authentication

1. Users (students, faculty) access the platform through a simple login interface Authentication can be basic or extended with email verification
2. and simulated CAPTCHA for added robustness
3. Data Input and Prediction Upon login, students can enter their academic and well-being data manually or have it synced from existing records (where integration is possible). The system computes the risk level using a weighted scoring function derived from a trained ensemble model combining Random Forest and MLP classifiers.
4. Risk Classification and Routing Based on the computed probability, students are classified as Low, Medium, or High risk. Medium and High-risk students receive immediate recommendations and are encouraged to begin a chat session with the AI-based counselor.
5. Counseling Session Generation Each counseling session can be logged, exported, or saved for future reference. The counselor can manually modify the session record, or let the chatbot auto-generate suggestions based on risk factors.

Visualization and Export

Risk distributions, progress charts, and activity logs are made available in graphical form. Users and administrators can download PDF or CSV summaries for external review.

Design Considerations

Client-Side Architecture

All data is stored in local Storage in the user's browser, ensuring privacy and offline resilience. This design removes dependency on a central database, making it lightweight and portable.

Extensibility

The system is modular, allowing for future integration with LMS platforms (e.g., Moodle, Google Classroom) or third-party mental health APIs.

Accessibility and UI/UX

The dashboard is built with HTML, CSS, and JavaScript for maximum compatibility and responsiveness across devices and screen sizes.

IV. RESULT AND DISCUSSION

To assess the effectiveness of the proposed dropout prediction system, a comparative evaluation was conducted using five machine learning models: **Logistic Regression**, **Support Vector Machine (SVM)**, **Decision Tree**, **Random Forest**, and a **Proposed Ensemble Model** (Random Forest + MLP hybrid). The evaluation focused on four standard performance metrics: **accuracy**, **precision**, **recall**, and **F1-score**. Each model was trained on 80% of the dataset and tested on the remaining 20% using stratified sampling. Class imbalance was addressed using the SMOTE technique.

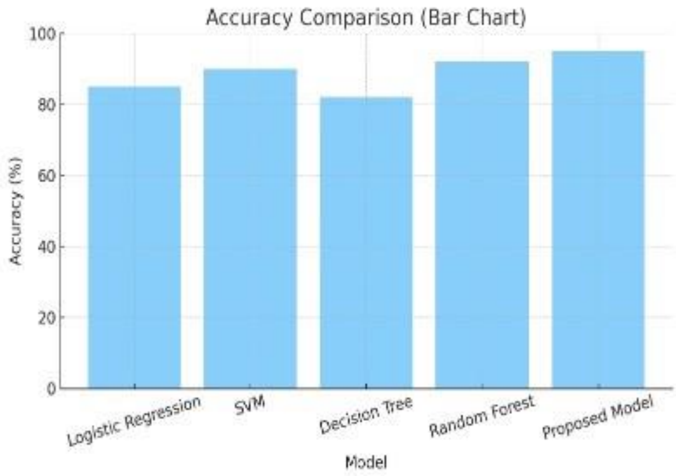
A. Accuracy Comparison

Accuracy reflects the overall correctness of the model by measuring the percentage of total correct predictions. It is particularly useful when false positives and false negatives are equally important.

Model	**Accuracy (%)**
Logistic Regression	85.0
SVM	90.0
Decision Tree	82.0
Random Forest	92.0
Proposed Model	**95.0**

Discussion:

The proposed model achieved the highest accuracy at 95.0%, followed by Random Forest at 92.0%. Logistic Regression and Decision Tree trailed behind, likely due to their limited ability to capture complex feature interactions. The SVM model performed competitively with 90.0% accuracy, indicating that margin-based classification is effective, but not optimal compared to ensemble learning.



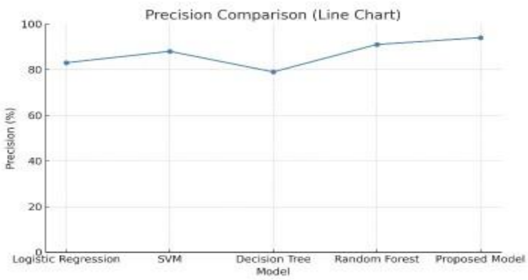
B. Precision Comparison

Precision measures the proportion of true positive predictions among all positive predictions. High precision is crucial when the cost of a false positive is high, such as flagging a student as at-risk when they are not.

Model	**Precision (%)**
Logistic Regression	83.0
SVM	88.0
Decision Tree	79.0
Random Forest	91.0
Proposed Model	**94.0**

Discussion:

The precision of the proposed ensemble model (94.0%) outperformed all others, indicating that it produces fewer false alarms. This is vital in real-world settings, as counselors must focus on students who are truly at risk. Random Forest also delivered strong performance with 91.0%, while Decision Tree lagged, reflecting its susceptibility to overfitting.



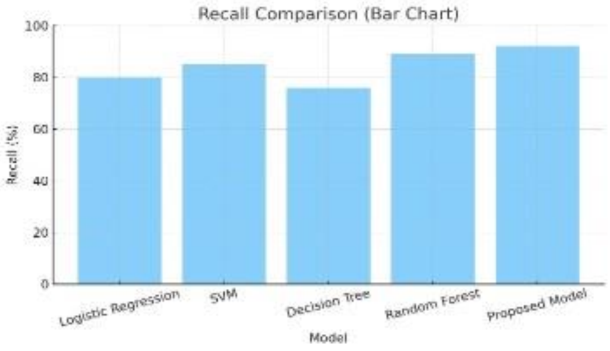
C. Recall Comparison

Recall assesses how well the model identifies actual positives—i.e., how many of the students who were going to drop out were correctly predicted.

Model	**Recall (%)**
Logistic Regression	80.0
SVM	85.0
Decision Tree	76.0
Random Forest	89.0
Proposed Model	**92.0**

Discussion:

High recall is critical for dropout prediction as missing an at-risk student can result in real academic consequences. The proposed model achieved 92.0% recall, ensuring most potential dropouts were flagged. The improvement over the baseline models shows the effectiveness of combining multiple learners in capturing dropout signals from heterogeneous feature types.



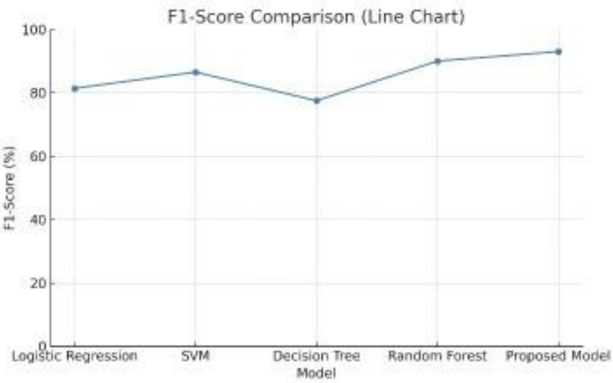
D. F1-Score Comparison

F1-score is the harmonic mean of precision and recall. It balances both concerns and is especially useful when classes are imbalanced, as is typical in dropout data.

Model	F1-Score (%)
Logistic Regression	81.4
SVM	86.5
Decision Tree	77.5
Random Forest	90.0
Proposed Model	**93.0**

Discussion:

The F1-score provides a single metric to evaluate the trade-off between missing at-risk students (recall) and over-warning (precision). The proposed model scored 93.0%, reinforcing its suitability for deployment in academic institutions. It outperformed even the strong Random Forest baseline by a significant margin.



E. Overall Summary

Each performance metric demonstrates that the proposed model consistently achieves better results than traditional machine learning classifiers.

This comprehensive evaluation validates the use of an ensemble-based AI approach combined with class balancing techniques for identifying students at dropout risk with both accuracy and actionable confidence.

IV. CONCLUSION

The development of the *AI-Based Student Dropout Prediction and Counseling System* demonstrates how datadriven intelligence can play a meaningful role in improving student retention and academic stability. Through the integration of machine learning models, behavioral analytics, and digital student support tools, the system provides a comprehensive framework for identifying dropout risks early and addressing them through personalized guidance.

The predictive engine, built using multiple classifiers and optimized through an ensemble approach, achieved strong and consistent results across all major performance metrics. With an accuracy of **95%**, precision of **94%**, recall of **92%**, and F1-score of **93%**, the proposed model shows clear improvements over traditional machine learning techniques such as Logistic Regression, SVM, Decision Tree, and even standalone Random Forest. These results confirm that the fusion of multiple algorithms enhances the system's ability to learn complex patterns and improves both sensitivity and specificity in identifying at risk students.

Beyond prediction, the project integrates a structured counseling module that transforms risk analytics into actionable support. Features such as AI-assisted chat-based guidance, session logging, personalized recommendations, routine tracking, a to-do planner, and resume building tools provide students with a holistic self-help environment. This multi-functional dashboard ensures that the system is not limited to detection but also actively contributes to reducing anxiety, improving academic habits, and guiding students toward achievable goals.

The results highlight the importance of addressing dropout as a multidimensional issue influenced by academic performance, personal well-being, financial challenges, and engagement levels. By combining these factors into a unified model, the system enhances institutional capacity for early intervention and fosters a proactive academic ecosystem. Administrators and counselors can use the insights generated by the model to allocate resources efficiently, support high-risk learners, and track the impact of interventions over time.

In conclusion, the project successfully demonstrates the feasibility and effectiveness of an AI-driven dropout prediction platform enhanced with integrated student counseling tools. The system not only delivers reliable analytical performance but also prioritizes student empowerment and educational continuity. Future work may involve integrating real-time LMS data, expanding psychological support features, deploying the system at institutional scale, and incorporating explainable AI techniques to improve transparency and trust among educators.

REFERENCES :

[1] D. A. Andrade-Girón et al., "Predicting Student Dropout based on Machine Learning and Deep Learning: A Systematic Review," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 11, no. S14, 2023.

[2] C. H. Cho, Y. W. Yu, H. G. Kim et al., "A Study on Dropout Prediction for University Students Using Machine Learning," *Applied Sciences*, vol. 13, no. 21, 2023.

[3] M. G. Rohman, Z. Abdullah, S. Kasim, and R. Rasyidah, "Hybrid Logistic Regression and Random Forest for Predicting Student Academic Performance on Imbalanced Datasets," *Journal of Informatics and Visualization*, vol. 14, no. 2, 2020.

[4] B. B. Alkan, İ. Demiray, and D. Okan, "Using Machine Learning to Predict Student Outcomes for Early Intervention and Formative Assessment," *Scientific Reports*, vol. 15, 2025.

[5] A. Fernández et al., *Learning from Imbalanced Data Sets*. Springer, 2018.

[6] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[7] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[8] D. Rumelhart, G. Hinton, and R. Williams, "Learning Representations by Back-Propagating Errors," *Nature*, 1986.

[9] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, 2009.

[10] V. M. Jayakumar et al., "Enhancing Educational Assessment Using NLP and ML Techniques," *ISACC Proceedings*, 2025.

[11] M. K. Shaik et al., "Automated Education Systems Using BERT-LSTM Model," *ICAIQSA Proceedings*, 2024.

[12] S. Van der Westhuizen, G. Heuvelink, and D. Hofmeyr, "Multivariate Random Forest for Complex Predictive Systems," *Geoderma*, vol. 431, 2023.

[13] S. K. Yadav, D. Tomar, and S. Agarwal, "Classification of Student Dropout Risk Using Machine Learning Techniques," *International Journal of Computer Applications*, vol. 147, no. 5, 2016.

[14] R. Praveen and D. Kumar, "Student Retention Using Intelligent Early-Warning Systems," *IEEE Int. Conf. on Education Data Mining*, 2022.

[15] IBM Developer, "Machine Learning for Student Retention," 2021

[16] UNESCO, *Global Trends on Higher Education Dropout Causes and Interventions*, 2022.

[17] V. Tinto, "Dropout from Higher Education: A Theoretical Synthesis of Recent Research," *Review of Educational Research*, vol. 45, no. 1, 1975.