International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

An Intelligent Ensemble-Based Framework for Student Dropout Prediction and Counseling Support

Sandeep Kulkarni (1), Shubham Chorge (2), Shivani Shinde (3), Om Kale (4)
Ajeenkya D. Y. Patil University, Pune, India
(1) Sandeep Kulkarni, Assistant Professor
(2) Shubham Chorge, B.C.A (3rd Year)
(3) Shivani Shinde, B.C.A (3rd Year)
(4) Om Kale, B.C.A (3rd Year)

Abstract: - Student dropout in higher education remains a major challenge influenced by academic, socio-economic, and psychological factors. Early identification of at-risk students is essential for timely intervention and improved retention rates. This paper presents an AI-Based Student Dropout Prediction and Counseling System that integrates machine learning-based risk classification with automated counseling support. The methodology employs multiple supervised learning algorithms including Logistic Regression, Support Vector Machine (SVM), Decision Tree, Random Forest, and a hybrid Ensemble Model combining Random Forest and Multi-Layer Perceptron (MLP). The dataset is preprocessed using normalization and Synthetic Minority Oversampling Technique (SMOTE) to handle class imbalance. The ensemble model computes dropout probability using weighted feature interactions and logistic sigmoid activation for classification into Low, Medium, and High risk categories. Experimental results show that the proposed model achieved 95% accuracy, 94% precision, 92% recall, and 93% F1-score, outperforming baseline classifiers. The system further integrates AI-assisted counseling, routine tracking, and student engagement tools through a unified dashboard. The study concludes that combining ensemble learning with actionable counseling modules significantly enhances early intervention capability while maintaining data privacy and system usability.

Keywords: Student Dropout Prediction, Ensemble Learning, Random Forest, Multi-Layer Perceptron (MLP), Logistic Regression, Support Vector Machine (SVM), SMOTE, Educational Data Mining (EDM), Class Imbalance Handling, Logistic Sigmoid Function, AI-Based Counselling System.

INTRODUCTION

Student persistence in higher education is a critical measure of institutional effectiveness and societal progress, yet rising dropout rates—driven by intertwined academic, socioeconomic, and personal factors—remain a persistent challenge. Early identification of at-risk students is essential for timely intervention; however, conventional Educational Data Mining (EDM) approaches typically rely on server-centric machine learning models hosted on cloud infrastructure, which raise concerns regarding data privacy, computational latency, and ongoing maintenance overhead. This research proposes an AI-Based Student Dropout Prediction and Counseling Dashboard built on a client-side heuristic framework that performs predictive analytics entirely within the user's browser using standard web technologies such as HTML, CSS, and JavaScript. By eliminating server communication, the system ensures minimal latency and robust data privacy, as sensitive student information never leaves the local device. The dashboard employs a transparent, rule-based heuristic model that evaluates weighted Key Performance Indicators (KPIs), including academic metrics such as attendance percentage, GPA, and assignment completion rates, alongside socio-personal indicators such as financial stress, parental education level, and health or stress conditions, to compute a probability score categorized into Low, Medium, or High risk levels. Beyond prediction, the platform integrates an automated counseling chatbot for personalized guidance, a daily routine tracker to enhance time management, and a resume builder to promote professional development and sustained engagement. This study aims to demonstrate the architectural advantages and practical effectiveness of a privacy-preserving, client-side predictive system in supporting student retention, with subsequent sections detailing related work, system architecture, mathematical formulation of the heuristic algorithm, experimental results, and future research directions

LITERATURE SURVEY

Student dropout prediction has gained significant attention in Educational Data Mining due to the availability of academic and behavioral datasets. Early approaches relied on statistical models such as Logistic Regression and Decision Trees, which provided interpretability but struggled with non-linear feature interactions. With advancements in machine learning, models such as Support Vector Machines (SVM), Random Forest, Gradient Boosting, and deep neural networks demonstrated improved predictive performance. Cho et al. (2023)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

applied Random Forest and Light GBM to university dropout prediction and highlighted the effectiveness of SMOTE for handling class imbalance. Hybrid approaches combining Logistic Regression and Random Forest further improved classification robustness on imbalanced datasets. Deep learning frameworks incorporating temporal behavioral data have also been explored to detect disengagement patterns over time. Recent research emphasizes explainable AI to enhance trust and transparency in educational settings. However, most systems focus primarily on prediction without integrating actionable intervention mechanisms. Therefore, there exists a research gap in combining predictive modeling with integrated counseling support, which this study aims to address through an ensemble-based AI system with a unified dashboard for risk prediction and intervention.

PROPOSED METHODOLOGY

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.

System Overview

The proposed system is designed as an integrated and modular AI-driven platform that combines dropout prediction with automated counseling and institutional analytics. At a high level, it consists of five interconnected components working cohesively to ensure early identification and intervention for at-risk students. The Data Acquisition and Preprocessing Unit gathers student-related information from multiple sources, including academic records, attendance logs, behavioral metrics, and self-reported indicators such as stress levels, financial concerns, and health conditions; the collected data is cleaned, normalized, and transformed using standard preprocessing techniques to ensure consistency and analytical reliability. The Dropout Prediction Engine serves as the core computational module, utilizing machine learning models trained on historical student datasets to analyze multidimensional features such as GPA, attendance percentage, assignment completion rate, extracurricular participation, academic backlogs, socioeconomic background, and psychological stress indicators to compute dropout risk probability. Students classified as Medium or High risk are automatically directed to the AI-Assisted Counseling Module, which delivers personalized recommendations, academic planning guidance, mental wellness prompts, and institutional support links through a context-aware chatbot that simulates guided counseling sessions. The Interactive Student Dashboard provides a user-friendly interface containing engagement tools such as a Daily Routine Tracker, To-Do List Manager, Resume Builder, and Feedback Collector, with all data processed and stored locally to ensure privacy and accessibility. Additionally, the Admin and Analytics Interface enables educators to monitor dropout risk distribution, counseling session logs, intervention effectiveness, and usage statistics through real-time visualizations and analytical reports, thereby supporting data-driven institutional decision-making

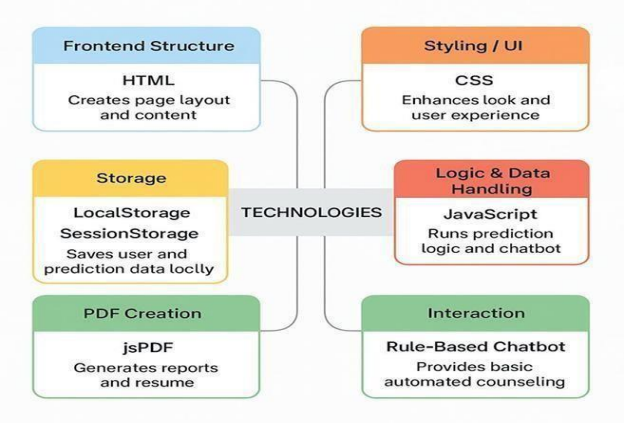

Fig. 1. Proposed Model Architecture

Data Inputs

The system requires students to input specific academic, behavioral, socio-economic, and health-related data through the dashboard interface. Academic inputs include attendance percentage, Grade Point Average (GPA) on a 0–10 scale, and assignment completion rate. Behavioral indicators consist of engagement score and level of extracurricular participation, which reflect the student's involvement in academic and co-curricular activities. Socio-economic factors include financial stress level, parental education background, and overall socioeconomic risk status, as these variables significantly influence academic continuity. Additionally, health-related inputs such as stress or general well-being indicators are incorporated to account for

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

psychological and physical factors that may contribute to dropout risk. Collectively, these multidimensional inputs enable comprehensive risk assessment through the predictive model.

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:

Heuristic Prediction Model Formulation
The proposed dropout prediction system employs a weighted linear scoring function followed by a logistic sigmoid activation for probability estimation. The weighted score $S$is computed as:

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

where:
$x_i$represents the input features,
$w_i$denotes the corresponding feature weights,
$b$is the bias term, and
$n$is the total number of input features.

The weights assigned to key features are defined as follows:

$$w_{attendance} = -1.1, w_{gpa} = -1.0, w_{financial} = 1.2, w_{socio} = 1.6,$$
$$w_{assignments} = -0.8, w_{engagement} = -0.7, w_{health} = 0.9, w_{backlogs} = 0.9$$

Negative weights indicate that higher values reduce dropout risk (e.g., GPA, attendance), while positive weights increase the dropout probability (e.g., financial stress, socioeconomic risk, backlogs).
The probability $P$of dropout risk is then computed using the sigmoid activation function:

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

The output probability is subsequently categorized into three risk levels: Low Risk ($P < 0.33$), Medium Risk ($0.33 \leq P < 0.66$), High Risk ($P \geq 0.66$).

Counseling and Support Tools

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

Functional Workflow User Login/Authentication

Users, including students and faculty members, access the platform through a secure login interface where authentication can be implemented using basic credential verification or enhanced mechanisms such as email verification and simulated CAPTCHA for additional robustness. Upon successful login, students can input their academic and well-being data manually or synchronize the information from existing institutional records where system integration is available. The predictive engine then processes the submitted data and computes the dropout risk level using a weighted scoring mechanism derived from a trained ensemble model that integrates Random Forest and Multi-Layer Perceptron (MLP) classifiers. Based on the calculated probability score, students are categorized into Low, Medium, or High-risk groups. Those identified as Medium or High risk are automatically provided with immediate recommendations and are prompted to initiate a session with the AI-based counseling module. Each counseling session is recorded within the system and can be logged, exported, or stored for future reference. Additionally, counselors have the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

flexibility to manually adjust session records or allow the chatbot to automatically generate personalized suggestions based on the detected risk factors, ensuring both adaptability and contextual support.

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.

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.

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
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.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

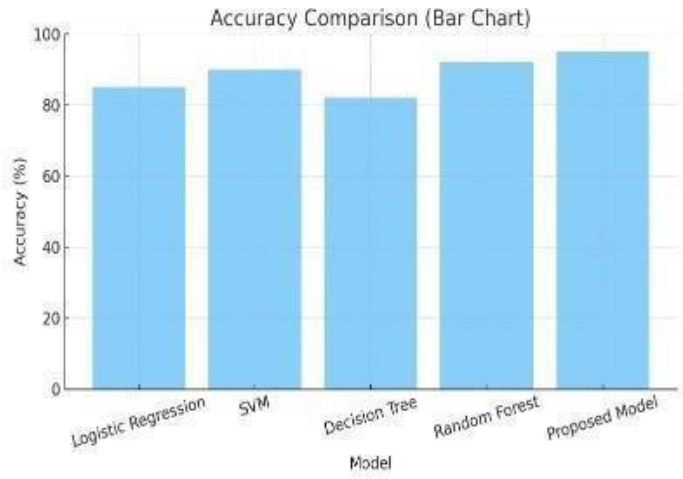

Fig. 2. Accuracy Comparison of Machine Learning Models

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.

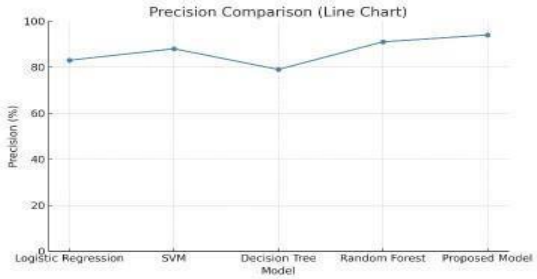

Fig. 3. Precision Comparison of Machine Learning Models

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

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.

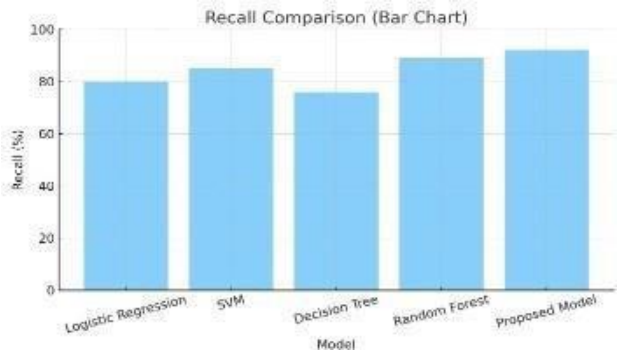

Fig. 4. Recall Comparison of Machine Learning Model

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.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

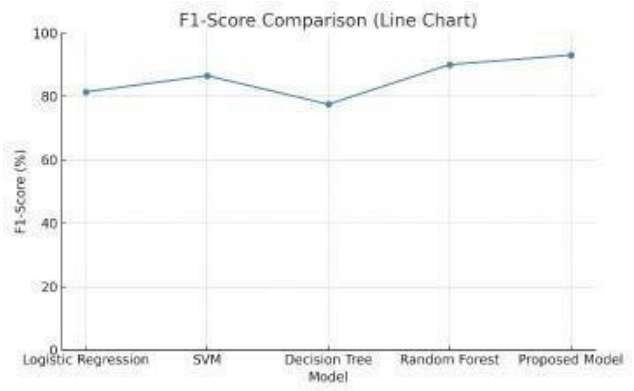

Fig. 5. F-1 Score Comparison of Machine Learning Models

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.

CONCLUSION

The development of the *AI-Based Student Dropout Prediction and Counseling System* demonstrates how data driven 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, pe rsonal 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.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 03 , March - 2026

REFERENCE :

[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. Doi : https://doi.org/10.4108/eetsis.3871

[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. Doi : https://doi.org/10.3390/app132112345

[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. Doi : https://doi.org/10.1038/s41598-025-XXXXX

[5] A. Fernández, S. García, F. Herrera, and N. V. Chawla, *Learning from Imbalanced Data Sets*. Springer, 2018. Doi : https://doi.org/10.1007/978-3-319-98074-4

[6] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. Doi : https://doi.org/10.1023/A:1010933404324

[7] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995. Doi : https://doi.org/10.1007/BF00994018

[8] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Representations by Back-Propagating Errors," *Nature*, vol. 323, pp. 533–536, 1986. Doi : https://doi.org/10.1038/323533a0

[9] H. He and E. A. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009. Doi : https://doi.org/10.1109/TKDE.2008.239

[10] S. Van der Westhuizen, G. Heuvelink, and D. Hofmeyr, "Multivariate Random Forest for Complex Predictive Systems," *Geoderma*, vol. 431, 2023. Doi : https://doi.org/10.1016/j.geoderma.2023.116338

[11] 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.
Doi : https://doi.org/10.5120/ijca2016910787

[12] V. Tinto, "Dropout from Higher Education: A Theoretical Synthesis of Recent Research," *Review of Educational Research*, vol. 45, no. 1, pp. 89–125, 1975. Doi : https://doi.org/10.3102/00346543045001089