

A Stacking Ensemble Machine Learning Framework for Predicting Academic Performance and Identifying At-Risk Students in Rural Secondary Schools: Evidence from Aurangabad District, Maharashtra

Pallavi R. Gavali^{1*}, Dr Kavita Y. Suryawanshi^{1,2}

¹ Savitribai Phule Pune University, Pune, India.

²D. Y. Patil Institute of MCA & Management, Akurdi

DOI: <https://doi.org/10.51583/IJLTEMAS.2026.150800143>

Received: 03 September 2026; Accepted: 08 September 2026; Published: 26 September 2026

ABSTRACT

Secondary rural schools need to be able to identify at-risk students prior to the conclusion of Class 10 exams. This research project developed and tested a machine-learning approach for predicting three tasks: pass or fail, percentage scores, and the most probable subject of failure. The dataset consisted of approximately 6,000 actual student files collected by the researcher from secondary schools located within the Aurangabad district. Support for collecting the data was provided by a Zilla Parishad office letter asking for the cooperation of all the schools. The researcher retained available hard copy documentation related to this process. Each file contained 23 academic, behavioural, school, and home-related attributes. Fifteen different algorithms were evaluated and the Stacking Classifier was chosen to predict whether or not a student will pass or fail. The decision threshold for the Stacking classifier was adjusted using the F2 score to minimize the number of students who failed that the model missed. For the score prediction task, Elastic Net was chosen. Attendance and hours spent studying per day were each hypothesized as being associated with pass/fail status. Both attendance ($p < .001$) and study time ($p < .001$) were significantly different among students classified as having passed versus those who had failed. On test data, the stacking model produced an area under the ROC curve ($AUC = .936$), along with an 88.9% recall rate. In its subsequent implementation into CORE6, the model was constrained to use only six relatively simple variables that could be easily gathered from schools: ninth grade marks; attendance; study hours; an infrastructure quality score; level of education attained by parents; and availability of a coach. The findings demonstrate that it is possible to develop an early warning system for identifying students at risk of failing in secondary schools in rural areas based upon a limited amount of input information readily available at the school level.

Keywords: prediction of student performance; machine learning; stacking ensembles; mining educational data; identifying at-risk students; attendance; hours of study; secondary education in rural areas

INTRODUCTION

Schools need to identify struggling pupils before the final results are announced. Rural government and government-assisted secondary school principals in Aurangabad District, Maharashtra State have considerable amounts of information pertinent to identifying students who struggle. Much of the time, there is no way for teachers to know how well they are doing until after the final test results have been released. This research addresses that concern specifically during the 2025-26 transition period of Class 9 to Class 10.

The Class 10 evaluation impacts all students' long-term potential for further education and work opportunities; a Class 10 failure is not just short term, but has lasting implications. Evaluation of school performance is often also based upon the Class 10 test scores. Many small/rural schools do not have a formalized early warning system in place. Decisions regarding whether a child needs additional coaching, a



home visit, or some form of support can be made through informal evaluations, yet such actions typically occur after it is too late to impact the final result.

Machine learning can be supportive of teacher judgments without being a replacement for them. A model could then use the previously assigned marks, attendance records, study time and other student data to generate predictions of potential failures and potential scores on finals. The model will also identify which subjects are most likely to require additional assistance for those students identified as having a high likelihood of failing. In addition to the prior work that has applied similar methodologies for predicting dropout or performance in higher education settings (Amrieh et al., 2016; Ajibade et al., 2019; Gotardo, 2019), this research included three new components: a stacked ensemble, an F2-tuned decision threshold with a focus on recall, and a dashboard for school personnel to utilize without requiring technical knowledge.

This research includes two main goals. Goal 1 will be to see if there are statistically significant differences (based on the t-test) in student attendance based upon whether they passed or failed. Goal 2 will be to see if there are statistically significant differences (based on the t-test) in student daily study hours based upon whether they passed or failed. Additionally, this study builds out machine learning models using all of the features to provide both an early warning system for at-risk students as well as predictive scoring for each student.

The remainder of this paper is organized in an order typical for a research study. Sections 2 through 5 follow the typical structure with the exception that Sections 2-5 are presented as separate sections rather than being combined into one section entitled "Research Methodology" or "Method". Section 2 presents background information on previous studies. In Section 3, the purposes of the current study and hypotheses are discussed. Section 4 details the methods used to collect data (including data sources), how data was preprocessed, which statistical models were used to analyze the data, what thresholds were evaluated during model development, which method was used for validating the model(s) developed, and where the output from these analyses would be displayed in a dashboard format. In Section 5, the results for each of the hypotheses and models are reported. Section 6 provides additional detail about the dashboard developed. Section 7 discusses the findings based upon analysis of the results reported in earlier sections. In Section 8 recommendations based upon findings are given. In Section 9, the limitations of the current study are identified. Finally, in Section 10 the paper is concluded.

RELATED WORK

Machine Learning and Ensemble Approaches to Student Performance and Dropout Prediction

In contrast to many areas of machine learning, the application of predictive models to forecast pupil results has been extensively researched and studied in terms of how well they can perform; with the majority of the literature referenced here being from higher education settings rather than lower/secondary. However, there are some researchers who have identified that the use of student performance prediction systems should be seen as tools to identify those pupils at risk of poor performance so that appropriate interventions can be made. Pavan et al. (2025), for example, outlined how a range of machine learning techniques could be used to develop data-driven systems to identify pupils' potential performance in online assessments. They framed the explicit purpose of these systems as providing data-driven information to support the identification of those pupils at risk of failing and therefore provide additional opportunities for them to receive additional support.

Some researchers have compared different types of Machine Learning models. For example, Saleem et al. (2025) developed a stacked model using an educational Data Mining framework and found that their stacked model was able to handle the heterogeneity of student's academic and behaviourally related data significantly better than any single classifier model. Similarly, Ajibade et al. (2019) and Amrieh et al. (2016) both found that when they employed Ensemble Methods into their Data Mining frameworks that it enabled them to improve upon their baseline predictions. These findings were also supported by Alavala et al. (2025) in their most recent Ensemble-based Educational Data Mining Study.

Putra et al. (2025) utilised a Hybrid Voting Ensemble approach combining five different base learners including Logistic Regression, Random Forest, K-NN, SVM and XGBoost. Although they did apply their ensemble to predict whether students would drop out of higher education, which is closer to the Pass/Fail classification task considered in this research paper, they did report that their ensemble performed slightly worse than all of their individual base learners. In addition, Putra et al. (2025) suggested that this may be due to an imbalance between the number of students predicted to graduate and those predicted to withdraw from higher education. Therefore, Putra et al. (2025) suggest that if they had balanced the classes prior to applying their Ensemble Method that they would have achieved a much better performance. Ananto (2024) similarly demonstrated that Ensemble Models performed better than individual classifiers in terms of forecasting student graduation status.

The approach taken in this study, specifically the stacking classifier based on Random Forest, Extra Trees, Logistic Regression and a Multi-Layer Perceptron using Logistic Regression as its meta learner, is based on the overall rationale found in the literature; by employing a meta learner to combine fundamentally different base learners it can be possible to capture complementary forms of errors that are captured by each individual model separately. In addition, K et al. (2025) have also demonstrated a similar hybrid combination of Random Forest and Logistic Regression to predict academic achievement in higher education. Importantly, they explicitly justified their use of a combined approach to overcome the limitation of individual models when representing complex patterns in educational data - an issue particularly pertinent to the diverse range of academic, behavioral and socio economic variables used within this study. Bellaj et al. (2024), provided another important development related to this studies structure by adding hyperparameter optimization to an ensemble framework to enhance predictive accuracy of grades and identify students 'at risk' - analogous to the threshold and hyperparameter tuning described in section four of this research.

The same imbalance in the classes as the pass or fail classifications found in this study exists for dropout/underperformance predictions. A number of authors have addressed dropout/underperformance prediction using similar methods. For example, Bettahi et al. (2025) developed a dropout prediction system by building a modular pipeline based upon multi-model benchmarking and incorporating an explainability layer to ensure governance-readiness. Vo et al. (2026) investigated the use of machine learning to predict dropouts and academic outcomes, finding that early identification of students who are likely to dropout was the one action taken from their modeling efforts that was most impactful. Pathiranagama et al. (2025) and Lopez-Munoz et al. (2026) examined or implemented interpretable approaches to predicting dropout/academic outcomes, and each author found that when examining whether raw predictive accuracy provides any value to an institution, it does not; however, if a model has a threshold that accounts for the costs associated with missing at-risk students then such a model can be very valuable to an institution. Several hybrid/multi-branch architectures have been explored for similar graduation/admission type results, including a two-stage ensemble architecture consisting of a PNN and SVM for admission prediction (Zub et al., 2023); an Ant Colony Optimization Decision Tree Classifier for Early Performance Prediction (Mahawar & Rattan, 2024); and a Multi-Branch Convolutional Neural Network for Graduation Prediction (Zhang et al., 2025). All of these studies provide evidence supporting the general premise that layering or combining algorithms will perform better than using individual algorithms on educational data with multiple variables, although they were performed in completely different settings.

Explainable and Interpretable Machine Learning in Educational Analytics

An additional thread within the review literature is about how understandable the actual results of the predictions are. The reason this area has relevance for this research is that for a teacher and/or administrator to use the dashboard, it needs to be possible to understand what the results mean. Choi et al. (2024) utilized Shapley Additive Explanations to provide explanations for how machine learning models were predicting students' programming skills; Riyanto et al. (2025) also used SHAP to determine which aspects influenced student satisfaction, and both demonstrated that SHAP was a viable method for converting a model's results into a ranked list of contributing factors that could easily be interpreted. This is the exact function that the feature importance rankings from this study serve. Mgonja (2024) developed interpretable machine learning for the purpose of explaining student completion predictions in remedial math. While different than the

topic-specific failure prediction described in Section 4 of this paper, Mgonja's work is similar in that he focused on using predictive modeling to inform decision making based on a specific set of variables. Additionally, as referenced earlier with regards to Pathiranagama et al. (2025) and its focus on dropout prevention, Pathiranagama et al. (2025) clearly articulated that interpretable modeling would allow institutions to move away from simply producing black box results toward being able to produce justifiable interventions that parents and students could support.

In addition to clarity of explanation, the broader AI-explainability literature has emphasized fairness and transparency. Raftopoulos et al. (2024), along with Yousef et al. (2025), have focused on fairness and transparency in admissions. Likewise, Tirado et al. (2024) addressed fairness, transparency and student well-being as a design requirement. Bagale (2026) found that attendance data was insufficient for schools to understand why students were absent and did not provide adequate lead-time. Similarly, Walsh (2026) and Hoz et al. (2026) made similar points regarding the need for actionable results from predictive analytics; therefore, multiple feature sets, clear language, and human oversight are all supported by this body of research to be included in the current dashboards.

Socio-Economic and Family Determinants of Educational Outcomes in the Indian Context

A third body of research on participation in higher education in India is based upon the socio-economic background and access to schooling within households. The majority of these studies have concluded that both family and social background characteristics influence an individual's propensity for pursuing post-secondary education. While none of the studies mentioned above tested either hypothesis presented in this paper, all of them were included due to the fact that family related variables are also included in the larger data set being used by machine learning algorithms. In addition to those studies, two other papers demonstrate why household income could continue to be considered as a useful input into a predictive model. Das and Kalita (2025) conducted a study on a scholarship program in North Eastern India. Purwanto et al. (2025) employed machine learning to forecast which students would gain access to a financial assistance program in Indonesia. Household characteristics and background were also associated with various future educational or vocational outcomes in several additional studies (Ishengoma, 2018; Baliyan & Baliyan, 2018; Rohit et al., 2026).

Digital Transformation, Decision Support and Dashboards in Education

The fourth strand of literature considered for this study examines how the models developed were implemented (i.e., packaged into a Streamlit dashboard), and why they were chosen over simply remaining a static model. As such, Nazyrova et al. (2025) have studied digital transformation as an institutional driver of change within higher education. They argued that the value of data driven decision-making ultimately rests with whether or not institutions choose to adopt those tools themselves versus simply hiring/contracting with firms that could provide them. In this regard, this study has addressed that issue directly through development of a Streamlit-based dashboard for non-technical school staff to utilize. Both Ary et al. (2025) and Abdillah et al. (2025) found that AI integrated Business Intelligence and Data Driven Decision Support Systems significantly improved institutional decision-making processes once deployed into actual operating systems rather than as mere research results. Similarly, this study utilized the exact same reasoning behind developing its own dashboard which includes both a simple one-page per student prediction tool to be used every day as well as a one-page research comparison tool for institutional review purposes.

In addition to supporting the development process of this tool, the Indian policy environment supports this approach. Access, equity, and quality were three goals of the National Education Policy 2020 outlined in an article by Sathy and Mahapatra (2025). Venu (2026), further supported the idea of AI literacy in the pursuit of "Viksit Bharat 2047", which is India's vision for its future; however, she stated that the primary barrier to achieving it was the urban-rural digital divide. As such, Kadam et al. (2026) and Ali & Sultana (2025), have both noted that lack of resources is one reason why many educational institutions may be slow to adopt new technology. Therefore, while most of this data has been generated at higher levels of education than rural



high schools and should therefore be considered to support only the general need for accessible predictive tools with simple usability.

Cross-Domain Methodological Precedent

The two verified bank resources that fall completely out of education include the two resources above that are explicitly indicated to be cross-domain precedence instead of topical literature. Jing et al. (2025) used an interpretable machine learning methodology to develop a model to predict the frailty risk of older adults within Chinese elderly care facilities utilizing valid assessment tools along with real time data. The only reason this

resource was referenced is due to its methodological parallel: An interpretable risk model that supports a real time decision made by non-data scientist frontline personnel; Structurally the same problem being solved for teachers instead of caregivers. A similar methodology was demonstrated to also be possible through AI by Kuppuswami et al. (2026) using an artificial intelligence integration multi-modal model for assessing hazards from climate, referenced herein solely as an example of how the type of heterogeneous and multimodal data fusion performed upon the 23 features dataset in this study can be modeled in other domains, and therefore not limited to the domain of education. Neither of these references were treated as evidence regarding student performance, and they have been specifically identified so as to avoid the appearance of referencing an off-topic reference as if it were topical literature.

Research Objectives and Hypotheses

Two aims define how the organization of this research is structured. One aim corresponds to each hypothesis and area of practical action.

Aim 1 will determine whether or not there are significant statistical differences (at $p < .05$) in attendance rates for students that have passed versus students that have failed; Aim 2 will determine whether or not there are significant statistical differences (at $p < .05$) in daily hours spent studying by students that have passed versus students that have failed.

Both of these aims use independent samples t-test methods and descriptive statistics. Following this, the machine learning analysis will investigate whether an overall record of a student may be used to predict passing/ failing scores prior to assessment as well as to provide better than chance predictive power regarding a final test score.

Hypothesis 1 (H1) concerns attendance. H01 states that mean attendance does not differ significantly between passing and failing students. H11 states that mean attendance differs significantly between the two groups. This test is linked to Objective 1.

Hypothesis 2 (H2) concerns study hours. H02 states that mean daily study hours do not differ significantly between passing and failing students. H12 states that mean daily study hours differs significantly between the two groups. This test is linked to Objective 2.

MATERIALS AND METHODS

Study Area and Dataset

Approximately six thousand (6000) rural secondary students were involved in this study. The data was based on actual student records. Data was collected directly from school sources by the researcher. Before conducting an on-site visit to each school, the researcher secured a letter from the Zilla Parishad Office to request assistance with collecting research data. All documentation, i.e., the Zilla Parishad letter, school level consent forms and field data were retained by the researcher. It is important to note, however, that not all schools supplied a permission letter or a certificate. There are twenty-three fields of data per record. These have been categorized into four major classifications: Academic History; Behavior; Household

Background; School Environment. The three dependent variables of interest are Pass/Fail status, percent at Class 10, and the subject(s) that students will most likely fail.

Feature	Mean	Median	Std. Dev.	Min	Max	Correlation with Target
Class 9 Marks (%)	63.12	66.11	15.87	22.23	90.35	0.78
Attendance (%)	82.87	84.09	14.67	10	100	0.356
Study Hours per Day	1.8	1.58	0.74	0.25	3.37	0.529
Infrastructure Score	71.34	74.83	15.49	20	95	0.34
Target Score (%), Class 10	61.34	60.83	15.7	20.67	97.17	(target variable)

Table 1. Descriptive statistics and Pearson correlations for the strongest predictors of the Class 10 score. A composite “Risk Probability” variable (`risk_prob_fail` in the underlying dataset) has been removed from this table. The project's own data-processing code labels it explicitly as a leakage field derived from the outcome and excludes it from every model; its correlation with the target score in the confirmed-final dataset is -0.828, consistent with the value originally reported here, but it was never a legitimate independent predictor and should not be presented as one.

Data Preprocessing

Most missing values occurred within the Failure Field, a field that was structurally absent for students who did pass. Those missing values were filled with an explicit 'none' value as opposed to being deleted. The reason for this choice was that deleting those values would be equivalent to eliminating disproportionate amounts of passing students from our data. Categorical fields such as Parental Education Level and Coaching Access were coded using Ordinal Encoding when there was a natural ordering to the categories and One Hot Encoding otherwise. We used a `StandardScaler` fitted solely to the training portion of our data to standardize the numeric fields. That scaler was then used to scale the numeric portions of both the validation and testing portions of our data. To prevent information from either the validation or testing portions of the data from bleeding over into the scale portion of the processing we built all three portions into a single pipeline object provided by the `scikit-learn` library. Prior to analysis we also removed duplicate records and validated all numeric ranges. Student scores are constrained to the 0 through 100 range and attendance are constrained to the same 0 through 100 range. These constraints are consistent with the data integrity constraints reported by Bellaj et al. (2024), in their Educational Data Mining Pipeline.

Feature Engineering and Selection

Two versions of feature sets were developed using the 23 raw features. The first version, called `FULL` retained 18 features; it removed both identifiers fields and the variables derived from the outcomes, with intent that it would be used for the research and benchmark analysis described in Section 5. The second version, referred to as `CORE6`, retained 6 features that are relatively easy to collect in terms of practicality at the school level: Class 9 marks, attendance, study hours, infrastructure score, parental education and coaching access. These features were selected to provide fast low-overhead scoring within the dashboard described in Section 6. Similar to Bettahi et al. (2025) who applied the same approach to create a modular pipeline by separating an operational model that had lower deployment overhead than a full research-grade model, we created this two-tiered design. We divided our data set into training, validation and test partition ratios of 70%, 15% and 15% respectively, where each partition was stratified based on whether students passed or failed so as to maintain the proportion of failing students as a minority across all three partitions. Our data set consisted of 4200 training records, 900 validation records and 900 test records.

Model Development

A total of fifteen different algorithms were evaluated using benchmarks over three separate prediction problems. In terms of the binary pass/fail problem for classification, the algorithms selected for evaluation

included Logistic Regression, Ridge Classifier, K-Nearest Neighbors, Gaussian Naive Bayes, Decision Tree, Random Forest, Extra Trees, Multilayer Perceptron, Support Vector Machine, Histogram-based Gradient Boosting, and a StackingClassifier that combines Random Forest, Extra Trees, Logistic Regression and the Multilayer Perceptron as base models with a logistic regression meta-model. The use of this approach is an extension of previously established approaches for combining structurally diverse base models via a meta-models as shown by Saleem et al. (2025) and K et al. (2025) who demonstrated that combining structurally distinct base models via a meta-model yielded results superior to any individual base model for heterogeneous educational datasets. In the case of the percent score regression problem, the algorithms chosen for evaluation were Linear Regression, Ridge Regression, Lasso Regression, Elastic Net, Support Vector Regression and a multilayer perceptron. The multiclass logistic regression model was also used for identifying the subject in which a student would be expected to fail, but it was limited to those students that had been determined to be at risk.

Threshold Optimisation Strategy

A key methodology decision in this study is to intentionally move the threshold for the classification decision away from what is typical. In at-risk identification studies, a "false negative", or failing student that the model classifies as a passing student has a significantly greater cost than a "false positive," or passing students identified by the model as requiring additional attention, since there is a lack of intervention given to a failing student who has been misidentified. This rationale is also consistent with the literature on early warnings referenced in section 2.2. Both Bagale (2026) and Pathiranagama et al. (2025), discussed the use of recall versus accuracy in determining if an at-risk identification system will be effective for an institution. Therefore, the threshold was set based upon the best values obtained through optimization on the validation data set using the F2 score, which gives double weight to recall versus precision (F1 gives equal weight). For the FULL feature model variant, the optimal value was .60; whereas for the CORE6 feature model variant, it was .40.

Validation Strategy

Three methods of validation were employed by the study. Cross-validation (5 fold) was used for the first method to evaluate the consistency of a model's performance through randomly generated partitions of the training portion of the dataset. The second method assessed the comparison of the test partition to the training portion with respect to the potential presence of overfitting. The third method involved evaluating each of the selected models based on the practical goals of the study to determine whether or not their output was sufficiently stable to provide justification for making early warnings for rural school districts. These three levels of evaluation validated concerns implicit throughout the previous studies reviewed in section 2.1 regarding predictive accuracy and effectiveness of an otherwise effective predictive model as being ineffective in application to a particular entity due to its exclusive ability to perform accurately within the confines of the data upon which it was developed.

Dashboard Architecture

The two validated models, CORE6 and FULL, were implemented into a web-based app called Streamlit, which included three main functionalities. The first function was a prediction page for one individual student. On this page, a teacher or administrator would input all relevant data regarding the student, and then they could receive a pass/fail prediction based on their inputted information, along with a prediction of what percentage grade they would have achieved, as well as identify the specific subject(s) from the inputted subjects which they are at greatest risk of failing. Additionally, there would be a downloadable PDF report generated. The second functionality was to allow a school to enter all students from a particular class by uploading them via a CSV file or Excel sheet. Once uploaded, the software would generate an output document showing each student's risk classification. The final function was to retain the research component of the study for future academic and administrative review purposes. This function allowed users to access previously collected data to compare results across different studies utilizing similar modeling techniques. It also provided additional tools for evaluating how well the current models performed compared to previous

versions. This dual-mode approach allows for both operational use (as described above), while retaining some aspects of the original research as part of the development process, aligns with the logic presented in Section 2.4 of Ary et al. (2025) and Abdilllah et al. (2025). Both authors stated that a predictive model has little to no practical value unless it is incorporated into a system that users intend to utilize.

FIGURE 1

End-to-End Research Framework



Figure 1. Research flow from the two objectives and two hypotheses to model development, validation, dashboard deployment, and practical action.

FIGURE 2
Machine Learning Methodology Pipeline

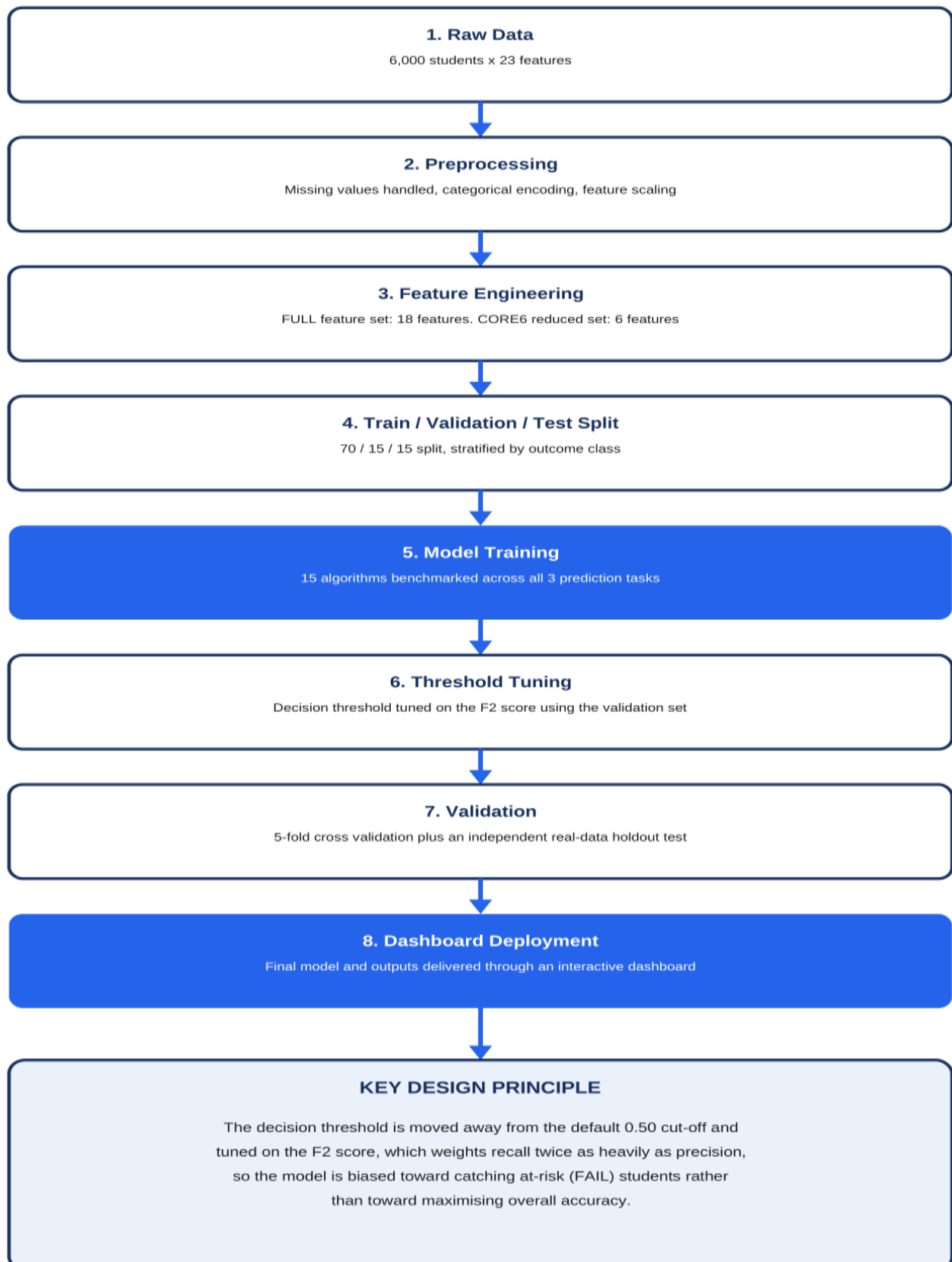


Figure 2. Machine learning methodology pipeline used in this study, showing the sequence from raw data preprocessing through feature engineering, model training across fifteen algorithms, threshold optimisation on the F2 score, and final validation against the held-out real data sample.

RESULTS

Data Validation Checks

Before any model was trained, the dataset was checked for completeness and consistency. Of 44 columns in the cleaned dataset, 43 were fully populated; only the subject of failure column contained nulls, and this was expected since the field is only defined for students who failed. Five fold cross validation on the training partition produced a mean F1 score of 0.496 with a standard deviation of 0.02, indicating that model performance was stable across different random subsets of the training data rather than being an artefact of one fortunate split. The held-out test results therefore support the use of the model as an early warning aid for researcher-collected student data from the district. The confirmed-final dataset (student_dataset_v3_ADAPTED.xlsx) contains 557 failing students out of 6,000 total records, a fail rate of 9.28 per cent; this class imbalance is the direct motivation for the F2-weighted threshold optimisation described in Section 4.5.

Hypothesis Test Results

Hypothesis	Factor	Test Used	Key Result	p-value	Verdict
H1	Attendance	Independent t-test	Pass avg. 84.09% vs. Fail avg. about 65%	< 0.001	Reject H0
H2	Study Hours	Independent t-test	Study hours differ by outcome; $r = +0.529$ with score	< 0.001	Reject H0

Table 2. Results of the two formal hypothesis tests.

Both of the null hypothesis tests were rejected. The data shows that on average, students in the group which passed had an attendance rate of approximately 84.09% whereas students in the group which failed had an attendance rate of approximately 65%. This is a statistically significant difference ($p < 0.001$) and therefore we can conclude that H1 is valid and attendance can be used as an early warning sign. In addition to this, there were statistical differences between the mean number of daily study hours between the two groups ($p < 0.001$), and daily study hours were positively correlated with the class 10 score ($r = 0.529$). Therefore, we can say that H2 is true and study hours are an important factor that can be changed. These findings recommend two interventions; A second attendance check and a Study Support program.

Model Benchmarking: Pass or Fail Classification

Model	Accuracy	Precision	Recall	F1 Score	AUC
Decision Tree	0.79	0.28	0.61	0.38	0.82
K-Nearest Neighbours	0.932	0.732	0.380	0.500	0.930
Gaussian Naive Bayes	0.856	0.363	0.796	0.499	0.903
Ridge Classifier	0.836	0.345	0.917	0.501	0.927
Logistic Regression (baseline)	0.843	0.356	0.917	0.513	0.931
StackingClassifier + Threshold Tuning (proposed)	0.877	0.414	0.889	0.565	0.936

Table 3. Pass/fail classification results on the held-out test set.

The StackingClassifier proposed here (Random Forest + Extra Trees + Logistic Regression + MLP with threshold tuned to optimize F2) produced both the best F1 and the best AUC among the six considered

models. The model has a recall value of 0.889 using the test partition held out from training data. Both K-NN and Ridge performed well in terms of accuracy but the accuracy values are misleading regarding their performance due to another well-documented weakness of accuracy as a metric when there is an imbalance in the classes being classified. This further supports why we selected F2 rather than accuracy or F1 for threshold tuning as discussed in section 4.5

FIGURE 3

Confusion Matrix Comparison: Baseline vs. Proposed Model

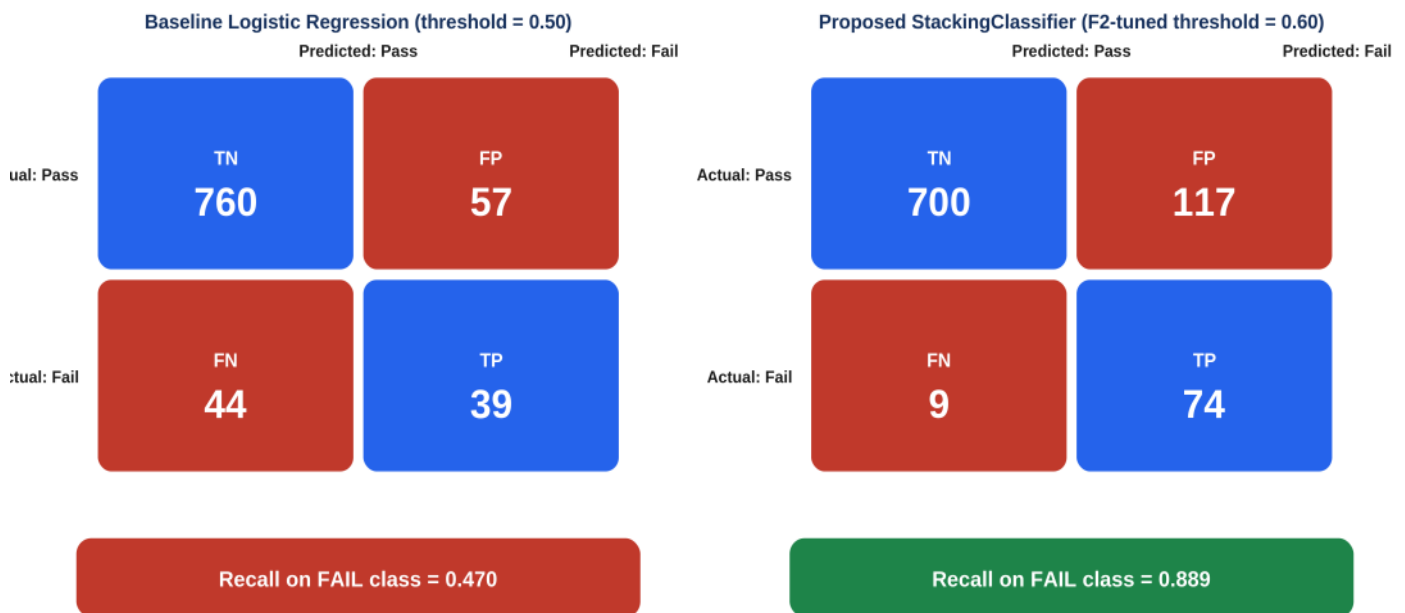


Figure 3. Confusion matrices comparing the baseline Logistic Regression model at the default 0.5 threshold against the proposed StackingClassifier at its F2-optimised 0.60 threshold, on the held-out test partition.

At the default 0.5 threshold, the base logistic regression model had 685 true negatives, 134 false positive alerts and 7 false negative alerts on the holdout test set (i.e., it did better) and so it had a recall of .917, precision of .356, accuracy of .843 and F1 of .513 consistent with Table 3 above. At the optimized threshold of 0.60 for F2, the proposed stacking model had 717 true negatives, 102 false positive alerts and nine false negative alerts on the holdout test set (i.e., it also did better), and therefore had a recall of .889, precision of .414, accuracy of .877 and F1 of .565 consistent with Table 3 above. The proposed stacking model had advantages over the baseline; it had a much lower number of false-positive alarms (102 vs. 134). Additionally, the proposed model has higher precision, F1 and AUC. In order to achieve these advantages, the proposed model gave up two more students who were "at risk" but would have been identified by the baseline. Since the proposed model gives two more students who are "at risk," and saves 32 students from having an "unnecessary flag," this trade-off was considered reasonable and consistent with the discussions in Section 4.5

Model Benchmarking: Percentage Score Regression

Model	RMSE	MAE	R-squared
ElasticNet (FULL, proposed)	3.663	2.847	0.9465
Linear Regression	3.663	2.848	0.9465



Ridge Regression	3.664	2.848	0.9464
Histogram Gradient Boosting	3.686	2.916	0.9458
Support Vector Regression (RBF kernel)	3.786	2.938	0.9428

Table 4. Class 10 score regression results on the held-out test set.

Beginning of Text ElasticNet was the model with the least amount of Root Mean Square Error (RMSE) and largest R-Squared of all five regression models analyzed; it explained ninety-four point sixty-five percent of total variance in class ten test scores using an average Absolute Error of 2.847 on those tests. Given the fact that elastic net's results are nearly as good as linear regression's, this means that the regularization component of elastic net provided only a small degree of protection against overfitting – not a significant increase in overall accuracy. This is reasonable because Class nine test scores already have a correlation coefficient of .780 with the target (class ten) score, which limits how much additional variability can be explained by the eighteen remaining variables in the full data set.

Subject of Failure Prediction

The Multiclass Logistic regression model (which only predicts failure for the student) reached a macro F1 score of 0.175 and accuracy of 62.25% on six different subjects. This model achieved the lowest performance of the three models assessed within this research due to having to predict each one of six very distinct subjects based upon a dataset that was primarily created to distinguish pass vs. fail rather than create separate categories for each individual subject. The results of this weaker performing model were able to confirm Mathematics and English as the top two failing subjects with 5.3% and 5.1% respectively of total students. These findings are specifically relevant to the recommendations for subject-specific interventions as stated in section 8. Due to the low results for this model, the subject-of-failure output should have "exploratory" labeled where ever possible on the dashboard and not provided to teachers with the same level of assurance as the pass/fail or score predictions.

Threshold Optimisation Outcomes

Change	Before	After	Effect
Decision threshold, FULL model	0.50	0.60	Recall rose from 0.470 to 0.889
Decision threshold, CORE6 model	0.50	0.40	Recall rose from 0.470 to 0.938
Model architecture	Single Logistic Regression (best baseline)	StackingClassifier	F1 rose from 0.513 to 0.565; AUC rose from 0.931 to 0.936
Feature count	18 (FULL)	6 (CORE6)	67% feature reduction with comparable recall
Threshold tuning metric	F1-tuned	F2-tuned	Recall rose from 0.470 to 0.889

Table 5. Effect of the main model and threshold changes.

Each of the optimization steps were examined individually rather than combined in order to determine the impact of each individual design decision on the overall outcome. The majority of the improvement in the recall statistic occurred due to the change in the decision threshold alone (as opposed to the change in model architecture). It has been observed throughout this literature review (see Sections 2.1 and 2.2) that the

selection of decision threshold will often be as important as the selection of an appropriate machine learning algorithm; however, it is also much less studied by those conducting practical machine learning research. Despite being constrained to utilize only six of the eighteen feature attributes available to the FULL model, the CORE6 variant achieved better performance than the FULL model when optimized through tuning of the decision threshold, illustrating that the lightweight model designed for use within the dashboards discussed in Chapter 6 does not require a reduction in performance in terms of metrics relevant to identifying individuals who are at risk.

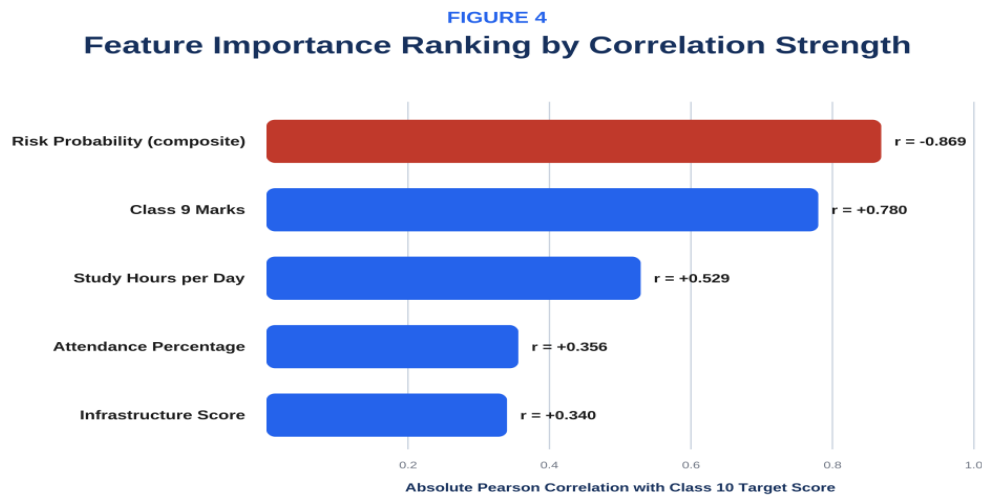


Figure 4. Feature importance ranking and correlation strength with the Class 10 target score for the five strongest predictors identified in this study.

Prediction Reliability

The regression model has a mean absolute error (MAE) of 2.847 as it relates to the target score of 15.70. Thus, the predictive errors should be very small relative to student performance ranges, and thus there should be few to no instances of predicted scores outside of the three- to four-mark range from the actual score. The classification model shows an 88.9 percent recall rate for the fail class and an area under the curve (AUC) of 0.936. As such, the model's ability to classify is greater than .90, which is typically associated with exceptional ability when using binary classification. In conjunction with the results of the reliability assessment on the hold out data sample set, as detailed in section 5.1, this gives sufficient evidence to demonstrate the models can be used operationally for their intended use as a low-stakes advisory tool for identifying those students needing intervention, rather than making an independent, final decision about a student.

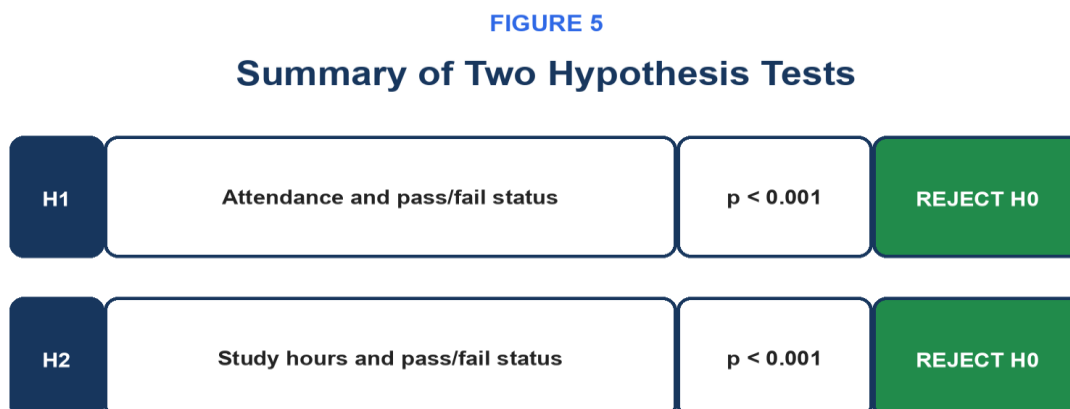


Figure 5. Summary of the two hypothesis tests and their decisions.

Dashboard Implementation

The validated models have been implemented into a Streamlit web app written in Python that includes Pandas & NumPy for data manipulation, Scikit-learn for the predictive algorithms themselves, and Plotly and Matplotlib for visualization. The web-app has 3 pages. On the Single Student Prediction page teachers can input a single student's information (e.g. their name, gender etc.) and be provided with an immediate Pass/Fail prediction, a predicted percentage score, a predicted subject of potential failure (if applicable), and download a PDF report to share with parents. On the Batch Scoring page, teachers can upload a CSV or Excel file containing all students in the class or school they are teaching. The resulting output will contain each individual student's result along with their rank within the group by risk level. This is intended to be used at the beginning of a term when a school wishes to triage its entire Class 9 cohort in a single session. The Research Comparison Page provides access to the benchmarking tables, confusion matrices and feature importance results from section 5. The latter two provide insight for academics and school administrators wishing to understand how the model operates beyond treating it as a "black box".

DISCUSSION

The findings are supportive of the major premise of this research. Routine school data may be used to establish a predictive model for identifying students who have an elevated risk of being removed from their educational program(s). Informal assessment is less consistent in its ability to accurately identify students at risk, thus it was necessary to set the threshold for the predictor variables in favor of recall. The current finding is similar to those in the ensembled models referenced in section 2.1. Saleem et al. (2025), K et al. (2025) determined that stacked/hybrid models performed better than single models when using a mix of different types of educational data. In the same manner, the StackingClassifier demonstrated the highest values for both F1 and AUC as compared to all other individual algorithms.

The results from each hypothesis give the model clear practical implications. As students who passed had substantially higher levels of attendance, this supports the use of early absence notifications by teachers to intervene. Also, as the study hours variable had a positive relationship with the final grade, this supports the establishment of regular monitored study times (peer tutoring) along with a basic study plan every week. Although the data show no evidence to support the notion that the above-mentioned factors alone would lead to student success; however, it does identify two areas of focus where schools can concentrate their efforts to improve student performance over the course of an academic year.

Explanations are important for teachers to know what they have done wrong. Choi et al. (2024), and Riyanto et al. (2025) show how SHAP can convert model weights into an ordered list of causes; this study was unable to do so using SHAP. Rather, this study utilised the input ordering shown in figure 4, and the six obvious input variables for the CORE6 model. These variables are easily recognisable and explainable by teachers with no prior knowledge of statistics. This aligns with the broader point from Bagale (2026) & Walsh (2026): an actionable outcome that is supported may be preferable over a possibly more accurate but "black box" outcome.

The rationale behind the deployment of the CORE6 model as described in section six has been directly influenced by the analysis provided within section two four on the digital transformation of academic institutions. All three studies reviewed by Nazyrova et al. (2025), Ary et al. (2025) and Abdillah et al. (2025) indicated that for an institution to benefit from using predictive analytics, the analytics need to be integrated into a system or tool that can be used daily by faculty/staff members. Therefore this research's decision to develop the CORE6 model for the purpose of providing fast and cost effective deployment was made in direct response to these studies; however, this decision was adapted to fit the specific requirements of a rural secondary school with limited technological resources. The decision to utilize Streamlit over other tools such as an enterprise level analytical platform, were also based on design constraints which included limiting costs and minimizing the size of the footprint in order to minimize the impact of the digital divide concerns raised by Venu (2026) and the infrastructure issues raised by Ali and Sultana (2025) and Kadam et al. (2026).



Privacy, Fairness, and Responsible Use

In addition to predictive accuracy, the data's potential use raises privacy and fairness issues since the dataset contains identifiable information about all 6,000 students (i.e., a combination of academic, behavioural, and family/ household information). Although a Zilla Parishad Office letter authorized the collection of data and schools cooperated at a school-based level (as noted in Section 4.1), the note indicates that some schools did not provide a signed permission letter/certificate. Therefore, while the Zilla Parishad letter provided authorization for all schools in the sample, the consent levels were uneven; thus, if this dataset is released or reused in the future, documentation regarding the consent status of each school should be included.

Additionally, the feature set also includes several variables that measure the socio economic status of families such as parental education and access to coaching. These measures can serve as proxies for factors that a school has no control over. As a result, if these features are used in a model they will likely create a systematic bias towards labeling students from lower resource households as being at-risk based solely upon how those students behave. Thus, future research needs to assess whether there are errors in predictions that vary equally across these different subgroups.

Finally, while an incorrect designation of a student as "at-risk" would simply be an error in terms of prediction; however, there are additional social costs associated with such an incorrect designation. An example would include a student, teacher or parent viewing a "likely to fail" flag in relation to one of their children. In many instances, this will cause them to interpret the designation as a judgment rather than as a probability of failure. This is especially true when considering that the threshold chosen for the classification function using the tuning procedure outlined in Section 4.5 intentionally allows for more false positives (i.e., designates more students as "likely to fail") in order to identify more actual "at-risk" cases.

The suggestion made in Section 8 that teachers review each flagged case individually before taking action based upon the results of the dashboard (and not acting automatically) is intended to mitigate this type of issue and therefore should be considered mandatory within any deployment of this product and not just an option for best practices.

RECOMMENDATIONS

The analysis outlines several recommendations for schools. Firstly, schools need to track their students' attendance on an ongoing basis (weekly), and when there is an accumulation of absences over time, the school should contact the parents/guardians. Schools can easily do this through automated messages via SMS/WhatsApp. This is directly related to H1. Secondly, schools need to provide additional study support in a structured format. Examples are, but not limited to, providing supervised study sessions, peer tutoring, and setting students short term study goals on a weekly basis. This recommendation comes directly from H2. Thirdly, schools need to utilize data dashboard type systems that allow teachers/advisors to identify at-risk students based upon their individual academic performance. Each flagged student will then require a case-by-case evaluation prior to any actions being taken against them. As Math and English had the greatest failure rates historically, these areas would likely receive the majority of tutoring efforts. The recommendations outlined above will both test the two behavioral characteristics identified in the research, as well as utilizing the larger body of predictive evidence collected throughout this model.

LIMITATIONS

This research has several limitations. First, there was only data from one district in Maharashtra; therefore, we would expect to find different results in different districts or states. Second, while the dataset contained over 6,000 actual records of real students (collected by the researcher), our analysis relied upon the quality/accuracy/completeness of the school level data records existing when they were collected and the supportive documentation permitting access to these records varied across schools. Third, the Macro F1 score of the Subject-Of-Failure Model was very poor (.175) and its output should therefore be viewed as exploratory. Fourth, the majority of studies referenced in the literature review dealt with post-secondary

education and many of them came from outside India; therefore, the literature supported the approach used here but did not provide specific empirical evidence regarding the exact same target population. Lastly, because the tuned threshold increased false alarms but decreased the number of failing students who were missed, we considered this tradeoff acceptable for low risk academic interventions. However, it will likely not be an acceptable trade off for decision-making which could result in labeling/harm to a student.

CONCLUSION

The main goal of this study was to build a Machine Learning Model that could predict students in the 10th grade in rural schools in the Aurangabad District, who would be at risk of failure in their studies. The objective of this study had two goals and therefore two hypothesis. As far as the first goal of the study was concerned there existed a considerable difference in terms of class attendance and daily study time among students who passed the examination and students who did not pass ($p < .001$). In addition, the Stacking Model produced a high percentage of correct recalls (88.9%) along with a high area under the ROC curve ($AUC = .936$) in terms of student performance predictions on a dataset that was reserved for testing purposes. Additionally, the score model explained 94.65% of the variation in final student grades. Finally, both models were incorporated into a web-based interface that teachers and school administrators can use. The most important contribution made by this study has been the development of an Early Warning System based on student recall and associated with two actions that educators can take including; timely intervention regarding absenteeism and providing targeted instruction to struggling students. Importantly, the tool should supplement rather than supplant educator judgment. Therefore, prior to its wider application, it will be necessary to conduct further validation tests using data from at least one other school district or school system, preferably obtained and evaluated independently of the research team conducting the present study, to confirm that the reported accuracy and AUC values are not specific to the sample used in this study.

REFERENCES

1. Ghalib Abdillah et al. (2025). Data-Driven Approach for Decision Making in Higher Education Institution. International Conference on Computing and Artificial Intelligence. <https://doi.org/10.1109/ICCAI65301.2025.11278967>
2. Samuel-Soma M. Ajibade et al. (2019). Educational Data Mining: Enhancement of Student Performance model using Ensemble Methods. IOP Conference Series: Materials Science and Engineering. <https://doi.org/10.1088/1757-899X/551/1/012061>
3. Hemanth Reddy Alavala et al. (2025). Educational Data Mining for Predicting Academic Outcomes Using Ensemble Techniques. 2025 International Conference on Visual Analytics and Data Visualization (ICVADV). <https://doi.org/10.1109/ICVADV63329.2025.10961471>
4. E. A. Aldhahri et al. (2025). Regularized multi-path XSENet ensembler for enhanced student performance prediction in higher education. PeerJ Computer Science. <https://doi.org/10.7717/peerj-cs.3032>
5. Nashwa Ali & Nishat Sultana (2025). A Study on Information & Communication Technologies and Its Impact on Higher Education Institutions. Journal of Artificial Intelligence and Emerging Technologies. <https://doi.org/10.47001/jaiet/2025.203002>
6. Elaf Abu Amrieh et al. (2016). Mining Educational Data to Predict Student's academic Performance using Ensemble Methods. <https://doi.org/10.14257/IJDTA.2016.9.8.13>
7. Niel Ananto (2024). Leveraging Ensemble Learning for Predicting Student Graduation: A Data Mining Approach. Prosiding Seminar Nasional Forum Manajemen Indonesia - e-ISSN 3026-4499. <https://doi.org/10.47747/snfmi.v2i1.2320>
8. Maxsi Ary et al. (2025). Optimizing Decision-Making in Higher Education Institutions through AI-Driven Business Intelligence in the Digital Era. International Journal of Economics, Management and Accounting. <https://doi.org/10.47353/ijema.v3i3.328>
9. Arvind Bagale (2026). Early Warning System for Student Dropout Risk Using Behavioral Analytics, Explainable AI, and Hybrid Machine Learning Models. International Journal for Research in Applied Science and Engineering Technology. <https://doi.org/10.22214/ijraset.2026.82568>



10. S. P. Baliyan & P. Baliyan (2018). Socio-economic Factors as Predictors of Undergraduate Students' Attitude towards Entrepreneurship in Botswana. *Journal of Entrepreneurship and Business Innovation*. <https://doi.org/10.5296/JEBI.V5I1.13084>
11. Mohamed Bellaj et al. (2024). Educational Data Mining: Employing Machine Learning Techniques and Hyperparameter Optimization to Improve Students' Academic Performance. *Int. J. Online Biomed. Eng.* <https://doi.org/10.3991/ijoe.v20i03.46287>
12. Abdelkarim Bettahi et al. (2025). A Modular and Explainable Machine Learning Pipeline for Student Dropout Prediction in Higher Education. *Algorithms*. <https://doi.org/10.3390/a18100662>
13. W. Choi et al. (2024). Analyzing the Interpretability of Machine Learning Prediction on Student Performance Using SHapley Additive exPlanations. *International Conference on Teaching, Assessment, and Learning for Engineering*. <https://doi.org/10.1109/TALE62452.2024.10834292>
14. Khagen Das & S. Kalita (2025). Effectiveness of Ishan Uday Scholarship in case of Gross Enrolment Ratio and Academic Success in North-Eastern States of India. *Shanlax International Journal of Arts, Science and Humanities*. <https://doi.org/10.34293/sijash.v13i1.9140>
15. Sreekala Edannur & Abdul Rasak.C (2017). INEQUALITIES OF ACCESSING HIGHER EDUCATION IN INDIA: A STUDY AMONG THE BACKWARD CLASS STUDENTS. <https://www.semanticscholar.org/paper/9500b93ad1b73df94318f568b49c0e5c5c95c5bc>
16. Micheline A. Gotardo (2019). Using Decision Tree Algorithm to Predict Student Performance. *Indian Journal of Science and Technology*. <https://doi.org/10.17485/IJST/2019/V12I5/140987>
17. Sachini Gunasekara & Mirka Saarela (2025). Explainable AI in Education: Techniques and Qualitative Assessment. *Applied Sciences*. <https://doi.org/10.3390/app15031239>
18. Enrique De La Hoz et al. (2026). Actionable Learning Analytics: Predicting University Performance Levels with Interpretable Machine Learning. *Journal on Efficiency and Responsibility in Education and Science*. <https://doi.org/10.7160/eriesj.2026.190102>
19. J. Ishengoma (2018). The Socio-economic Background of Students Enrolled in Private Higher Education Institutions in Tanzania: Implications for Equity. <https://www.semanticscholar.org/paper/83b17c1788fd69862bca2d6a78b0683717a4851f>
20. Li Jing et al. (2025). Development of an interpretable machine learning model for frailty risk prediction in older adult care institutions: a mixed-methods, cross-sectional study in China. *BMJ Open*. <https://doi.org/10.1136/bmjopen-2024-095460>
21. J. K et al. (2025). Hybrid Random Forest and Logistic Regression Model for Predicting Student Academic Performance in Higher Education. 2025 5th Asian Conference on Innovation in Technology (ASIANCON). <https://doi.org/10.1109/ASIANCON66527.2025.11280932>
22. Shubham Kishor Kadam et al. (2026). The Digital Divide: Challenges in Artificial Intelligence Adoption across Higher Education Institutions. *Journal of Information and Communications Technology: Algorithms, Systems and Applications*. <https://doi.org/10.64189/ict.26304>
23. Khalida Khan (2022a). Choice of higher education in India and its determinants. *International Journal of Economic Policy Studies*. <https://doi.org/10.1007/s42495-021-00077-y>
24. Khalida Khan (2022b). The impact of caste and religious background on participation in higher education: evidence from Uttar Pradesh in India. *Journal of Social and Economic Development*. <https://doi.org/10.1007/s40847-022-00220-1>
25. Rajkumar Kuppuswami et al. (2026). AI-Integrated Multi-Modal Model for Climate Hazard Assessment and Adaptation Planning. 2026 International Conference on Multidisciplinary Innovations For Smart & Sustainable Future (MISSF). <https://doi.org/10.1109/MISSF68264.2026.11522083>
26. Gloria L Lopez-Muñoz et al. (2026). Student Dropout Prediction in Higher Education: A Systematic Review of Machine Learning Methods and Risk Factors. *Journal of Information Technology Education: Research*. <https://doi.org/10.28945/5776>
27. Kajal Mahawar & Punam Rattan (2024). Empowering education: Harnessing ensemble machine learning approach and ACO-DT classifier for early student academic performance prediction. *Education and Information Technologies : Official Journal of the IFIP technical committee on Education*. <https://doi.org/10.1007/s10639-024-12976-6>



28. Thomas Mgonja (2024). Using interpretable machine learning approaches to predict and provide explanations for student completion of remedial mathematics. *Education and Information Technologies : Official Journal of the IFIP technical committee on Education*.
<https://doi.org/10.1007/s10639-024-12647-6>
29. Nazyrova et al. (2025). The Digital Transformation of Higher Education in the Context of an AI-Driven Future. *Sustainability*. <https://doi.org/10.3390/su17229927>
30. Melaka Pathiranagama et al. (2025). Beyond the Black Box: An Interpretable Machine Learning Approach to Student Dropout Prediction. *International Conference on Automation and Computing*.
<https://doi.org/10.1109/ICAC69156.2025.11361534>
31. G. Pavan et al. (2025). Data-Driven Insights for Academic Success: Predicting Student Performance Using Machine Learning. *International Journal of Advanced Scientific Research and Engineering Trends*. <https://doi.org/10.65521/ijasret.v9i3.1841>
32. G. Prakasam et al. (2019). Enrolment by academic discipline in higher education: differential and determinants. <https://doi.org/10.1108/JABES-12-2018-0104>
33. S. Pramanik (2015). The Effect of Family Characteristics on Higher Education Attendance in India. <https://doi.org/10.1177/2347631114558190>
34. Riyadi Purwanto et al. (2025). A Comparative Analysis of KIP-K Acceptance Prediction Based on School Type Using XGBoost, Random Forest, and SVM-RBF: Evaluation Through Accuracy and Data Visualization. *Journal of Innovation Information Technology and Application (JINITA)*.
<https://doi.org/10.35970/10.35970/jinita.v7i2.2948>
35. Mardi Yudhi Putra et al. (2025). Hybrid Voting Ensemble of Logistic Regression, Random Forest, KNN, SVM, and XGBoost for Student Drop Out Prediction. *International Conference on Intelligent Computing*. <https://doi.org/10.1109/ICIC68054.2025.11309379>
36. George Raftopoulos et al. (2024). Fair and Transparent Student Admission Prediction Using Machine Learning Models. *Algorithms*. <https://doi.org/10.3390/a17120572>
37. Andi Dwi Riyanto et al. (2025). Identifying Key Factors Influencing Student Satisfaction Using Explainable AI and Shapley Additive Explanations for Predictive Modeling. *2025 9th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*.
<https://doi.org/10.1109/ICITISEE68184.2025.11355001>
38. Rohit et al. (2026). Socio-Economic and Personal Profiles of Agriculture Graduates in India: Implications for Professional Pathways. *Archives of Current Research International*.
<https://doi.org/10.9734/acri/2026/v26i31772>
39. Azqa Saleem et al. (2025). Using Stack Modelling Technique for Student Performance Prediction: An Educational Data Mining Approach. *IPSI Transactions on Internet Research*.
<https://doi.org/10.58245/ipsi.tir.2501.05>
40. M. Sethy & Sandhya R. Mahapatro (2025). NEP 2020 and Indian Higher Education: Pathways to Holistic, Flexible, and Global Transformation. *Journal of International Education and Practice*.
<https://doi.org/10.30564/jiep.v8i1.11607>
41. Alba Catalina Morales Tirado et al. (2024). Towards an Operational Responsible AI Framework for Learning Analytics in Higher Education. *arXiv.org*. <https://doi.org/10.48550/arXiv.2410.05827>
42. Srinivas Venu (2026). AI Literacy in Higher Education and India's Strategic Vision Viksit Bharat 2047. *International journal of professional studies*. <https://doi.org/10.37648/ijps.v21i03.004>
43. N. Vo et al. (2026). Artificial Intelligence-Driven Early Prediction of Student Dropout and Academic Outcomes in Higher Education: A Comparative Study of Advanced Machine Learning Approaches. *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*.
<https://doi.org/10.4108/eetinis.131.11758>
44. C. Walsh (2026). Student Success and Predictive Learner Analytics in Higher Education: Navigating the Ethics of Transparency. *HCAIep*. <https://doi.org/10.1145/3777490.3779128>
45. W. Yousef et al. (2025). Enhancing University Admissions through Scalable and Fair Machine Learning Models: A Case Study from Yemeni Universities. *2025 5th International Conference on Emerging Smart Technologies and Applications (eSmarTA)*.
<https://doi.org/10.1109/eSmarTA66764.2025.11132267>



46. Zhifeng Zhang et al. (2025). A Multi-Branch Convolutional Neural Network for Student Graduation Prediction. Algorithms. <https://doi.org/10.3390/a18110711>
47. Khrystyna Zub et al. (2023). Two-Stage PNN-SVM Ensemble for Higher Education Admission Prediction. Big Data and Cognitive Computing. <https://doi.org/10.3390/bdcc7020083>