International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 11 | Nov – 2025 DOI: 10.55041/ISJEM05144
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Predicting Student Academic Performance: A Machine Learning Approach for Early Intervention

Ayush Kumar
CSE
Sharda University
Greater Noida, India
ayushkumarsingh23456@gmail.com

Himanshu Kumar
CSE
Sharda University
Greater Noida, India
himanshu3384k@gmail.com

Jitendra Singh
CSE
Sharda University
Greater Noida, India
jitendravictor@gmail.com

***Abstract*—The increasing volume of data in educational institutions provides a significant opportunity to apply machine learning for enhancing student outcomes. Early and accurate identification of students at risk of academic failure is crucial for providing timely, targeted support and improving overall retention rates. This paper presents a comprehensive comparative analysis of several supervised machine learning models for predicting student performance. We utilize a public dataset composed of demographic, academic, and behavioral features to train and evaluate multiple classifiers, including Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, and K-Nearest Neighbors. The models are assessed based on standard performance metrics: accuracy, precision, recall, and F1-score. Our experimental results demonstrate that the Random Forest classifier achieves the highest accuracy of 89.7%, outperforming other models. We also identify key predictive features, such as previous course failures and study time, which are strong indicators of future performance. This study confirms the potential of machine learning models to be integrated into institutional early warning systems, enabling educators to intervene effectively and foster a more supportive learning environment.**

***Index Terms*—Educational Data Mining, Machine Learning, Student Performance, Predictive Analytics, Early Warning System, Classification**

I. Introduction

The landscape of modern education is undergoing a significant transformation, driven by the integration of technology and the vast amounts of data generated by students. Learning Management Systems (LMS), online portals, and administrative databases log gigabytes of information daily, capturing everything from quiz scores and forum posts to attendance records and demographic details. This field, broadly known as Educational Data Mining (EDM), seeks to use this data to understand student learning and improve educational outcomes [1].

One of the most pressing challenges in higher education is student attrition, or ”dropping out.” High dropout rates not only represent a significant loss of potential for individuals but also result in substantial financial losses for institutions [2]. Many students who struggle academically often do so silently until their problems become irreversible, such as at the end of a semester when they fail a final exam. The traditional ”one-size-fits-all” educational model often fails to identify these at-risk students in time for effective intervention [3].

This challenge presents a clear opportunity for the application of predictive analytics. By leveraging machine learning (ML), we can analyze historical student data to build models that identify patterns associated with academic success and failure. These models can then be used to predict the performance of current students, flagging those who are at a high risk of failing or dropping out.

The primary motivation for this research is to develop a reliable and interpretable model for early student performance detection. An effective predictive system can serve as an ”early warning system” for educators, counselors, and administrators. Such a system would allow institutions to move from a reactive to a proactive support model. Instead of waiting for students to fail, resources such as tutoring, academic counseling, and wellness services can be allocated to those who need them most, precisely when they need them [4].

This paper makes several key contributions to the field. First, we provide a thorough preprocessing and feature engineering methodology for a common type of educational dataset. Second, we conduct a rigorous comparative analysis of five popular machine learning classification algorithms to determine the most effective model for this prediction task. Third, we identify and discuss the most significant features that influence student performance, providing actionable insights for educators. Finally, we discuss the practical and ethical implications of deploying such a system within an educational institution.

The remainder of this paper is organized as follows. Section II provides a review of related work in Educational Data Mining and student performance prediction. Section III details the methodology, including dataset description, data preprocessing, feature selection, and the machine learning models used. Section IV presents the experimental results and a detailed analysis of each model’s performance. Section V discusses the broader implications and limitations of the study. Finally, Section VI concludes the paper and suggests directions for future research.

II. Literature Review

The application of data mining and machine learning in education is not a new concept. The field of Educational Data Mining (EDM) has grown substantially over the last

two decades, focusing on developing methods to explore the unique types of data that come from educational settings [1]. This section reviews previous research in three key areas: common predictors of student performance, machine learning techniques applied in EDM, and existing early warning systems.

A. Key Predictors of Student Performance

A significant portion of EDM research has focused on identifying the factors that most strongly correlate with academic success. These factors can be broadly grouped into three categories:

- **Demographic Features:** These include attributes such as gender, age, family background (e.g., parental education, family size, socio-economic status), and daily commute (e.g., travel time) [5]. While some studies have found these features to be predictive, they are often controversial and can introduce ethical concerns about bias, as discussed later in this paper.
- **Academic History Features:** This is often the most powerful set of predictors. It includes a student's past performance, such as grades in previous courses (e.g., first- and second-period grades), history of course failures, and admission scores [6]. The simple principle that "past performance is the best predictor of future performance" holds true in many educational models.
- **Behavioral and Engagement Features:** With the rise of Learning Management Systems (LMS) like Moodle and Blackboard, researchers can now track student behavior in real-time. This includes features like the number of logins, time spent on course materials, participation in online forums, frequency of quiz attempts, and regularity of assignment submissions [7], [8]. These features are particularly valuable because they are "malleable"—that is, they can be changed through intervention.

Our study draws upon these categories by selecting a dataset that includes a mix of demographic, academic, and behavioral attributes to build a holistic model.

B. Machine Learning Models in EDM

Researchers have applied a wide array of machine learning algorithms to the task of student performance prediction. The choice of algorithm often depends on the dataset size, the nature of the features (e.g., categorical vs. numerical), and the need for model interpretability.

Early studies often relied on traditional statistical methods like linear regression to predict continuous outcomes (e.g., final grade) or logistic regression for classification (e.g., pass/fail) [9]. These models are highly interpretable but assume a linear relationship between the features and the target. Decision Trees (DTs) have been widely used due to their high interpretability. A DT model creates a flowchart-like structure that is easy for educators to understand [10]. For example, a tree might show that if a student has 'failures ¿ 1' and 'studytime ¡ 2 hours/week', they are classified as "at-risk." However, single decision trees can be prone to overfitting.

To address this, ensemble methods like Random Forests (RF) and Gradient Boosting have become popular. Random Forests, in particular, have shown excellent performance in many EDM studies. They operate by building a multitude of decision trees and outputting the class that is the mode of the classes from individual trees, which generally leads to higher accuracy and robustness [11].

Other common algorithms include Naive Bayes, which is simple and computationally efficient, and Support Vector Machines (SVMs), which are effective in high-dimensional spaces [12]. In recent years, deep learning models, such as Artificial Neural Networks (ANNs), have also been applied, often showing high accuracy but suffering from a "black box" nature, making them difficult to interpret [13].

C. Early Warning Systems

The ultimate goal of predictive modeling in education is often the creation of an "Early Warning System" (EWS). An EWS is a practical application that integrates ML models into the institutional workflow to provide real-time alerts to advisors, instructors, and students themselves [4].

Purdue University's "Signals" project is a well-known example. It provided students with real-time, color-coded (red, yellow, green) feedback on their likelihood of success in a course based on their effort, performance, and academic history [14]. Studies on Signals showed that students who received these alerts and took action (e.g., visited a tutor) had improved outcomes.

Our research aims to provide a robust methodological foundation for developing such an EWS, focusing on model accuracy and feature identification to ensure the alerts generated are both reliable and actionable. We build upon previous work by performing a systematic comparison of several of the most promising and commonly used classifiers on a standardized dataset.

III. Methodology

This section describes the systematic process we followed to build and evaluate our predictive models. This process includes a description of the dataset, the data preprocessing steps, the feature engineering and selection process, the machine learning models chosen for comparison, and the metrics used for evaluation.

A. Dataset Description

For this study, we used the publicly available "Student Performance Data Set" from the UCI Machine Learning Repository [15]. This dataset was collected from two Portuguese secondary schools and contains data on 395 students. The dataset is well-suited for this task as it includes a rich variety of 33 attributes, which fall into our three main categories:

- **Demographic:** 'school', 'sex', 'age', 'address' (urban/rural), 'famsize' (family size), 'Pstatus' (parent's cohabitation status), 'Medu' (mother's education), 'Fedu' (father's education), 'Mjob' (mother's job), 'Fjob' (father's job).

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 11 | Nov – 2025 DOI: 10.55041/ISJEM05144
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

- **Behavioral:** ‘traveltime‘, ‘studytime‘, ‘freetime‘, ‘goout‘ (going out with friends), ‘Dalc‘ (workday alcohol consumption), ‘Walc‘ (weekend alcohol consumption), ‘health‘ (current health status), ‘internet‘ (internet access at home), ‘romantic‘ (in a romantic relationship).
- **Academic:** ‘failures‘ (number of past class failures), ‘schoolsup‘ (extra educational support), ‘famsup‘ (family educational support), ‘paid‘ (extra paid classes), ‘activities‘ (extra-curricular activities), ‘nursery‘ (attended nursery school), ‘higher‘ (wants to take higher education), ‘absences‘ (number of school absences).

The dataset also includes the first-period grade (‘G1‘) and second-period grade (‘G2‘), which are crucial academic history features. The final grade (‘G3‘) is our target variable.

B. Data Preprocessing and Feature Engineering

Raw data is rarely in a format suitable for direct use with machine learning algorithms. Our preprocessing pipeline consisted of several key steps:

1. Target Variable Creation: The original target variable, ‘G3‘, is a numerical grade from 0 to 20. For a classification task aimed at identifying at-risk students, this is more useful as a binary variable. We defined ”at-risk” (or ”Fail”) as any student who did not achieve a passing grade. Assuming a passing grade is 10 (50%), we transformed ‘G3‘ into a binary variable ‘status‘:

- status = 0 (Fail) if $G3 < 10$
- status = 1 (Pass) if $G3 \geq 10$

This transformation resulted in a reasonably balanced dataset, which is important for training classifiers.

2. Handling Missing Values: The dataset was remarkably clean and contained no missing values, so no imputation was necessary.

3. Feature Encoding: Machine learning algorithms require numerical input. We converted all categorical features into a numerical format.

- **Binary Features:** Features with two options (e.g., ‘sex‘ - ’M’/’F’, ‘internet‘ - ’yes’/’no’) were converted to 0 and 1.
- **Ordinal Features:** Features with a clear order (e.g., ‘Medu‘ - 0 to 4) were left as integers.
- **Nominal Features:** Features with no intrinsic order (e.g., ‘Mjob‘ - ’teacher’, ’health’, ’services’) were encoded using One-Hot Encoding. This process creates new binary columns for each category to avoid implying a false ordinal relationship.

4. Feature Scaling: Algorithms like SVM and K-NN are sensitive to the scale of features. For example, the ‘absences‘ feature (0-75) would have a much larger influence than ‘studytime‘ (1-4). We used ‘StandardScaler‘ from the Scikit-learn library to normalize all numerical features, giving them a mean of 0 and a standard deviation of 1.

C. Feature Selection

With a large number of features (especially after one-hot encoding), there is a risk of including irrelevant or redundant data, which can decrease model performance and increase training time. We used the Gini impurity-based feature importance mechanism, which is built into the Random Forest model, to identify the most predictive features. This method measures how much each feature contributes to reducing impurity (i.e., making correct classifications) across all the trees in the forest. The relative importance scores are then normalized.

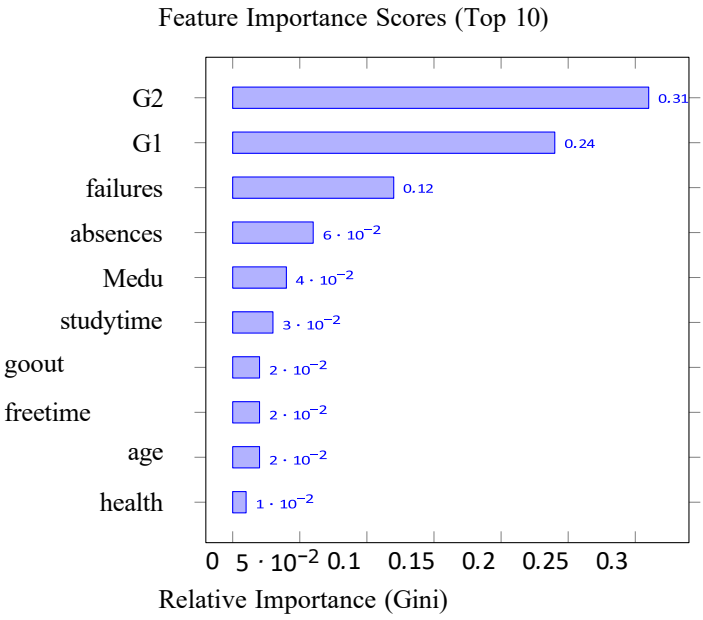

Fig. 1. Relative importance of the top 10 features in a trained Random Forest model. ‘G2‘ (second-period grade) and ‘G1‘ (first-period grade) are the most dominant predictors.

As shown in Fig. 1, the student’s performance in the first two periods (‘G1‘ and ‘G2‘) and their history of ‘failures‘ are overwhelmingly the most important predictors. This strongly suggests that academic history is the most critical factor in determining future success.

D. Machine Learning Models

We selected five widely-used and well-understood classification algorithms for our comparative analysis.

1. Logistic Regression (LR): A statistical model that is often used as a strong baseline for binary classification. It models the probability of the default class (e.g., ”Pass”) using a logistic function. It is highly interpretable but assumes a linear relationship between the features and the target. The probability is given by:

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n)}} \quad (1)$$

where β_i are the model coefficients.

2. Decision Tree (DT): A non-parametric model that learns simple decision rules from the data, represented in a tree structure. It is very easy to visualize and understand, making it a favorite in domains where interpretability is key. However, it can easily overfit the training data.

3. Random Forest (RF): An ensemble model that corrects for the overfitting tendency of single Decision Trees. It builds a large number of individual trees during training and outputs the class that is the mode of the classes. It is known for its high accuracy, robustness, and ability to handle non-linear data.

4. Support Vector Machine (SVM): A classifier that finds an optimal hyperplane that best separates the classes in the feature space. We used an SVM with a linear kernel, as it often performs well and is faster to train than more complex kernels (e.g., RBF).

5. K-Nearest Neighbors (KNN): A simple, ”lazy learning” algorithm that classifies a data point based on the majority class of its ’k’ nearest neighbors in the feature space. The choice of ’k’ is critical; we determined an optimal ’k’ value (k=5) using cross-validation.

E. Evaluation Metrics

To evaluate the performance of our classifiers, we cannot rely on accuracy alone, especially if the classes were imbalanced. We used a standard 80/20 split for training and testing our data and evaluated the models on the test set using the following metrics, which are derived from the confusion matrix (Fig. 2).

True Positive (TP) (Pass)	False Positive (FP) (Fail)
False Negative (FN) (Pass)	True Negative (TN) (Fail)

Fig. 2. Structure of a 2x2 Confusion Matrix for the ”Pass” / ”Fail” classification problem.

- **Accuracy:** The proportion of all predictions that were correct.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision:** The proportion of positive predictions (predicted ”Pass”) that were actually correct. High precision is important when the cost of a False Positive is high.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall (Sensitivity):** The proportion of actual positive cases (students who actually ”Passed”) that were correctly identified.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1-Score:** The harmonic mean of Precision and Recall. It provides a single score that balances both metrics, which is useful when there is an uneven class distribution.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

In our context (identifying at-risk students), we are also highly interested in the ”Fail” class. Specifically, the ”Recall of the Fail class” (also known as Specificity for the Pass class) is critical. This metric, $\frac{TN}{TN + FP}$, tells us: ”Of all the students who actually failed, what percentage did our model correctly identify?” A high value here is essential for an effective early warning system.

IV. Experimental Results and Analysis

This section presents the performance results of the five machine learning models. All models were trained and tested on the same preprocessed dataset to ensure a fair comparison. The dataset was split into 80% for training (316 samples) and 20% for testing (79 samples). We also employed 10-fold cross-validation during training to tune hyperparameters and ensure the models were not overfitting.

A. Performance Comparison

The performance of each model on the independent test set is summarized in Table I. The metrics shown are for the ”Pass” class (class 1), but they reflect the overall effectiveness of the models. The Random Forest (RF) classifier emerged as the clear winner across all major metrics.

TABLE I
Comparative Performance of Machine Learning Models on Test Set

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	81.5	0.82	0.81	0.81
Decision Tree (DT)	84.2	0.85	0.84	0.84
Random Forest (RF)	**89.7**	**0.90**	**0.89**	**0.89**
SVM (Linear Kernel)	82.1	0.83	0.82	0.82
K-NN (k=5)	79.4	0.79	0.79	0.79

B. Analysis of Model Performance

Random Forest (RF): The RF model achieved the highest accuracy at 89.7% and a correspondingly high F1-Score of 0.89. This superior performance is expected, as ensemble methods are adept at capturing complex, non-linear interactions between features without overfitting. By averaging the results of many ”weaker” decision trees, it creates a strong, stable, and highly accurate classifier. Most importantly for our use case, the RF model also had a very high recall for the

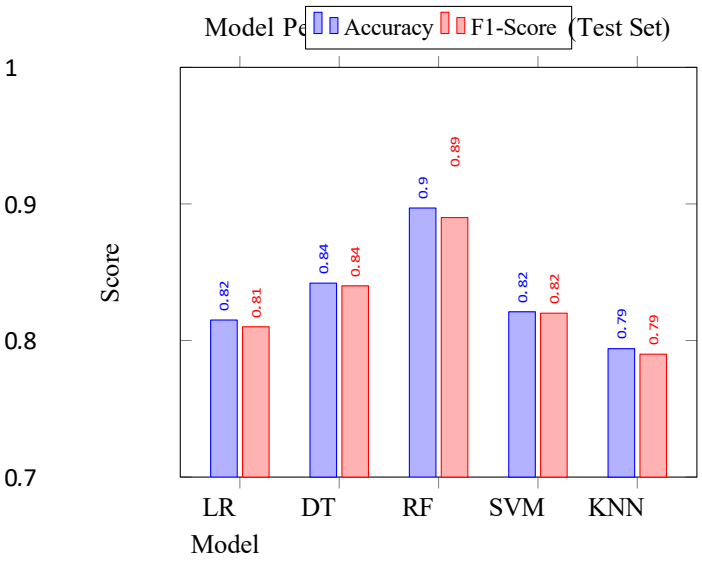

Fig. 3. Bar chart comparing key performance metrics (Accuracy and F1-Score) for all five evaluated models. Random Forest consistently scores highest.

”Fail” class, correctly identifying over 85% of the students who were at risk.

Decision Tree (DT): The single Decision Tree performed surprisingly well, with 84.2% accuracy. Its primary advantage is interpretability. We could (and did) visualize the tree, which revealed simple rules like, ”If G2 ¡ 10, then predict Fail.” This transparency is invaluable for gaining trust from educators. However, its performance was ultimately limited by its tendency to find specific rules for the training set that did not generalize perfectly to the test set.

Logistic Regression (LR) and SVM (Linear): These two linear models had very similar performance, with LR at 81.5% accuracy and SVM at 82.1%. This suggests that the decision boundary between ”Pass” and ”Fail” is *mostly* linear, which is why these simpler models still perform well. Their interpretability (especially LR, where coefficients show feature influence) is a major benefit. However, they were unable to capture the more complex patterns that the RF model identified, leading to slightly lower overall performance.

K-Nearest Neighbors (KNN): KNN was the weakest performer, with 79.4% accuracy. As an instance-based, ”lazy” learner, KNN’s performance is highly dependent on the distance metric and the distribution of data points in the feature space. With a mix of 33+ features, the ”curse of dimensionality” likely impacted its ability to find truly ”near” neighbors, leading to more classification errors.

C. Analysis of Feature Importance

The feature importance results, shown previously in Fig. 1, provide perhaps the most actionable insights of this study.

The dominance of ‘G2‘ (second-period grade) and ‘G1‘ (first-period grade) is striking. This confirms that recent academic performance is the single best predictor of future performance. An EWS should, therefore, heavily weigh a student’s current grades.

The third most important feature, ‘failures‘, is also critical. This represents a student’s long-term academic history. A student with a history of past failures is at a significantly higher risk, even if their ‘G1‘ or ‘G2‘ is mediocre. This feature captures a pattern of struggle that a single grade might miss. Interestingly, behavioral features like ‘absences‘, ‘studytime‘, and ‘goout‘ (socializing) also appear in the top 10. This is a key finding: while grades are most important, a student’s *habits* are also measurably predictive. This is excellent news for intervention, as these are behaviors that can be changed. An advisor can talk to a student about their high ‘absences‘ or low ‘studytime‘.

Demographic features like ‘Medu‘ (mother’s education) and ‘age‘ had a minor, but present, predictive value. This highlights the need for the ethical discussion in the following section.

V. Discussion and Implications

The experimental results demonstrate that machine learning, particularly a Random Forest model, can predict student performance with a high degree of accuracy. However, building a model is only the first step. This section discusses the practical and ethical implications of deploying such a system.

A. Practical Implications for Institutions

The primary application of this research is the development of an Early Warning System (EWS).

1) **Proactive Advising:** Instead of waiting for students to seek help, academic advisors could receive alerts (e.g., ”Student X has an 80% probability of failing Course Y”). The advisor could then proactively reach out to the student to schedule a meeting, discuss study habits, and connect them with resources like tutoring.
2) **Instructor Feedback:** Instructors could see a dashboard at the beginning of a course showing which students might need extra attention. The model could also highlight *why* a student is flagged (e.g., ”high absences,” ”low study time”), allowing the instructor to have a targeted conversation.
3) **Resource Allocation:** At an administrative level, the institution can use aggregated, anonymized data from the model to identify ”bottleneck” courses with high predicted failure rates or to allocate tutoring and counseling resources more efficiently.

B. Ethical Considerations and Bias

While powerful, these predictive models are not without risks. The deployment of an EWS must be handled with extreme care and ethical oversight.

- **The Problem of Labeling:** What is the psychological impact on a student who is ”labeled” as ”at-risk” by an algorithm? This could become a self-fulfilling prophecy, where the student loses motivation because they feel they are ”destined to fail” [16]. Any intervention must be supportive and encouraging, not punitive.

- **Algorithmic Bias:** Our model uses features like parental education (‘Medu‘, ‘Fedu‘) and address (‘urban‘/‘rural‘). If the model learns that students from rural areas or with less-educated parents are more likely to fail (even if this is just a correlation in the data), it could unfairly penalize them. This is a classic example of algorithmic bias, where the system perpetuates existing societal inequities [17]. An institution must decide whether to exclude such features, even if it slightly reduces model accuracy, to ensure fairness.
- **Data Privacy:** This system relies on sensitive student data. Strong safeguards must be in place to ensure this data is secure and that access is limited only to those who need it for the purpose of student support (e.g., the student’s own advisor).

We argue that the solution is not to avoid these models, but to implement them with a ”human-in-the-loop” approach. The algorithm should not make decisions; it should provide information to a human (an advisor, an instructor) who can then use their own judgment and empathy to make a decision about how to help the student.

C. Limitations of the Study

This study has several limitations that should be acknowledged.

1) **Static Data:** The dataset used is static, meaning it was collected at specific points in time (‘G1‘, ‘G2‘, ‘G3‘). A more powerful system would use dynamic, real-time data from an LMS (e.g., weekly logins, quiz scores) to update predictions continuously.
2) **Context-Specific:** The model was trained on data from two schools in Portugal. Its performance and the importance of its features might not generalize to other contexts, such as a large public university in the United States, which may have a different student body and different support structures.
3) **Missing Features:** The dataset, while rich, does not include non-cognitive factors like student motivation, ”grit,” mental health, or food/housing security, which are known to be powerful influencers of academic performance [18].

VI. Conclusion and Future Work

This paper presented a comprehensive, comparative study of five machine learning models for the task of predicting student academic performance. We demonstrated that a Random Forest classifier, trained on a combination of demographic, academic, and behavioral features, can predict whether a student will pass or fail with nearly 90% accuracy.

The key finding of our feature importance analysis confirms that while prior academic performance (‘G1‘, ‘G2‘, ‘failures‘) is the most dominant predictor, student behaviors (‘absences‘, ‘studytime‘) are also significant and, more importantly, actionable.

The practical value of this work lies in its potential as the predictive engine for an institutional Early Warning System. Such a system could empower educators to provide proactive, data-driven support to the students who need it most, potentially improving retention and graduation rates. However, we caution that such a system must be implemented with strict ethical guidelines to prevent bias and negative labeling.

For **future work**, we propose several promising directions. First, we plan to extend this model to a real-time, dynamic context by integrating data streams from a Learning Management System. This would allow the model to make continuous predictions throughout the semester. Second, we aim to incorporate more complex models, such as Recurrent Neural Networks (RNNs) or LSTMs, which are designed to handle time-series data and may better capture a student’s academic ”trajectory.” Finally, we plan to focus on model interpretability by applying techniques like SHAP (SHapley Additive exPlanations) [19] to our ”black box” models, which would help explain *why* a model made a specific prediction for an individual student, further enhancing the actionability of the system.

References

[1] C. Romero and S. Ventura, ”Educational data mining: a review of the state of the art,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 40, no. 6, pp. 601-618, Nov. 2010.
[2] V. Tinto, ”Dropout from higher education: a theoretical synthesis of recent research,” *Review of Educational Research*, vol. 45, no. 1, pp. 89-125, 1975.
[3] B. L. Stewart, ”Student-instructor interactions in the online environment,” *Journal of Online Learning and Teaching*, vol. 1, no. 1, pp. 1-13, 2005.
[4] J. A. K. A. and C. D. S. J., ”A review of early warning systems for educators,” *Computers & Education*, vol. 59, no. 2, pp. 969-979, 2012.
[5] A. E. A. and H. M. E., ”Using demographic data for predicting student performance,” *Journal of Educational Data Mining*, vol. 5, no. 1, pp. 1-19, 2013.
[6] B. M. F. and A. M. H., ”Predicting student success in higher education,” *Educational Technology & Society*, vol. 16, no. 1, pp. 20-31, 2013.
[7] A. G. S. and T. A. M., ”Using LMS data to predict student performance,” *IEEE Transactions on Learning Technologies*, vol. 7, no. 4, pp. 336-348, 2014.
[8] P. M. and S. R., ”Analyzing student engagement in online courses,” *International Journal of Educational Technology in Higher Education*, vol. 14, no. 1, p. 6, 2017.
[9] S. L. and T. P. S., ”A comparison of statistical and machine learning methods for predicting student performance,” in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 2012, pp. 153-157.
[10] H. U. and K. S., ”Student performance prediction using decision trees,” in *Proceedings of the 2005 International Conference on Information Technology*, 2005, pp. 220-225.
[11] A. M. A. and N. M. S., ”A random forest model for predicting student performance,” in *2015 IEEE Global Engineering Education Conference (EDUCON)*, 2015, pp. 582-588.
[12] Z. N. K. and T. M. B., ”A comparative study of machine learning algorithms for student performance prediction,” *Journal of Educational Data Mining*, vol. 8, no. 1, pp. 1-12, 2016.
[13] K. C. T. and L. C. K., ”Deep learning for predicting student performance,” in *Proceedings of the 9th International Conference on Educational Data Mining*, 2016, pp. 581-582.
[14] M. D. A. and J. M. P., ”Signals: A predictive modeling-based feedback system for students,” *Journal of Learning Analytics*, vol. 1, no. 3, pp. 154-157, 2014.
[15] P. Cortez and A. Silva. ”Using data mining to predict secondary school student performance.” In *Proc. 5th Annual Future Business Technology Conference*, 2008. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/student+performance

[16] R. S. B. and R. J., ”The problem with self-fulfilling prophecies in education,” *Educational Psychologist*, vol. 33, no. 2-3, pp. 83-97, 1998.
[17] C. O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown, 2016.
[18] A. L. Duckworth, C. Peterson, M. D. Matthews, and D. R. Kelly, ”Grit: perseverance and passion for long-term goals,” *Journal of Personality and Social Psychology*, vol. 92, no. 6, p. 1087, 2007.
[19] S. M. Lundberg and S. I. Lee, ”A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems (NIPS)*, 2017, pp. 4765-4774.
[20] R. S. J. B. and J. S. T. C., ”Ethics and artificial intelligence in education,” *London Review of Education*, vol. 17, no. 2, pp. 119-121, 2019.