



A Framework for Explainable Early Prediction of Student Academic Performance Using Machine Learning Techniques

Nanaparapu Venkata SatyaNarayana

Department of Computer Science and Artificial Intelligence
Central University of Andhra Pradesh
Email: nanaparapuvencat@gmail.com

Dr. P. Sumalatha (Assistant Professor)

Department of Computer Science and Artificial Intelligence
Central University of Andhra Pradesh
Email: sumalatha.psl@gmail.com

Abstract—The earlier at-risk students can be detected, the better students' performance and lower the dropout rate can be achieved. The goal of this research is to develop an explainable machine learning framework for early predicting students' academic performance on Open University Learning Analytics Dataset (OULAD). It employs the information available from early-stage features such as demographics, students' engagement and initial assignment grades to make predictions. Several ML models were applied and a Random Forest classifier obtained the best prediction performances over other methods concerning accuracy, precision, recall and F1-score. The proposed system added the technique of SHAP (SHapley Additive Explanations) for both global and local interpretation of model predictions and it also provides explanations through a user-friendly dashboard. The outcomes indicate that this proposed framework provides an excellent and efficient solution to make decisions based on data that yields a timely and insightful system with accurate predictions and clear explanations.

Index Terms—Early Student Performance Prediction, Machine Learning, Random Forest, Explainable Artificial Intelligence (XAI), SHAP, Educational Data Mining, OULAD Dataset, Predictive Analytics, Student Engagement, Decision Support System

I. INTRODUCTION

In recent years, with the development of digital learning environment and online education, the volume of education data (such as student interactions, test results, behavior within VLE, etc.) grows rapidly. Education Data Mining (EDM) and Learning Analytics (LA) research are developed and make full use of the data to improve student learning and academic achievement and also teaching quality. Prediction of student performance is emerging as one of the most important research fields in all data-driven education applications.

A. Purpose of Student Performance Prediction

The purpose of predicting student performance is to identify students who are at risk of failing early in the course. Traditional assessment, like test and manual observation, is often implemented at the middle of the course and can identify failing students at later stage, so there are limited time for intervention and fail in the end due to not getting support in time. Early prediction system is therefore needed.

B. Role of Machine Learning in Student Performance Prediction

The machine learning (ML) technique have emerged as a very powerful method for education data analysis to uncover the hidden patterns and relationships. Random Forest, Decision Tree and Logistic Regression are frequently used in student performance prediction and show good results in prediction accuracy. But previous studies merely take prediction accuracy as criteria to measure their work and ignore other factors, such as early-stage prediction, interpretability and usability.

C. Limitations of Existing Prediction Approaches

One of the great weakness in many previous studies was the lacking of emphasis on the early stage prediction part. So far, majority of studies had been trying to make predictions based on the full data and the final score of a course. Such approach cannot perfectly meet on-line teaching needs. Based on that fact, we propose an early prediction approach where student performance prediction could be performed at the early stage of a course and help students who will fall behind in time.

D. Lack of Explainability in Machine Learning Models

Also, the ML models applied for student performance prediction have been referred as 'black box' which were hard to understand by users. These 'black boxes' have less trust when we get some critical outcomes, so there are limits of these models in the real teaching context. Fortunately, the explainability methods of AI could make the model interpretable. SHAP (Shapley Additive explanations) is considered in the work for showing the prediction results of student's behavior and engagement at the course level.

E. Need for User-Friendly Decision Support Systems

Except prediction and interpretability, we also identified a clear demand for user-friendly decision support systems which are easier to be employed by educators and administrators. Only the development aspect has been concentrated on for



majority of research without taking the ease of use in applications into account, which widens the gap between research and practice.

- **The best choice is to have a system** that can not only give the prediction results but also make the prediction in an explainable way to aid decision-making by the stakeholders.
- **Hence, in our study, we designed an Early Student Performance Prediction System** based on ML, feature engineering and explainable AI in an integrated fashion.
- **The models are trained with the early student data** (e.g. Student interactions and assessments in 30 days) for prediction and the features extracted are from the student academic history and behavior data.
- **Random Forest classifier was selected as the prediction model** due to its robustness and excellent performance on large datasets.
- **The SHAP values are used for explanation**, both globally and locally.

Finally, the system is realized in an interactive Streamlit based dashboard to allow educators to have an overview on prediction and its key factors, and further analyze the student's behaviors and make adequate decisions.

The primary purpose of the study is to construct a system with the state-of-art predictive power and ability to provide opportunity for early interventions and transparency, and to eliminate the gap between models and practice by proposing an complete system design.

II. LITERATURE REVIEW

Student performance prediction is becoming a hot topic for research in Educational Data Mining and Learning Analytics for the purpose of improving academic achievement and reduce drop-out rate [1, 2]. In the last decade, the authors have observed the extensive usage of data-driven methods in modeling of student behavior and prediction of student performance [1, 3].

Initially, the studies on student performance prediction use conventional statistical methods, like linear regression, correlation etc., in modeling relationships between student attributes and performance [4]. However, these methods failed to capture the complex non-linear relationships that exist in the data. With the progress in machine learning algorithms, complex models such as Decision Trees, Nave Bayes, Support Vector Machines, Logistic Regression were extensively used for prediction of student performance that provide better predictive accuracy and is capable of handling massive datasets [5, 6].

In the last few years, ensemble methods such as Random Forest and Gradient Boosting algorithms have gained tremendous attention for their excellent performance and robustness [7]. Especially the Random Forest models have gained extensive utilization in student performance prediction due to its high generalization accuracy, ability to overcome the overfitting problem, and handling of heterogeneous data, and

many works claim ensemble methods obtain a better prediction accuracy and stability than the traditional statistical models for student performance prediction [3, 7].

A. Research Gaps

Although there has been considerable work in predicting students' performance using machine learning techniques, there are several key weaknesses in the current literature. A major flaw that exists among most of the papers reviewed is that prediction accuracy using all data (with final exam data included) is the goal, instead of predicting early intervention. This makes it less useful.

- No early stage prediction on partial data, so timely intervention is impossible.
- No explainable models, so not trust by teachers.
- No combination of prediction, interpretation, and deployment in one system.
- No practical systems with user-friendly interface between research and application.

This work aims to close the above gaps by proposing an integrated framework that brings together early prediction, explainable AI, and an interactive decision support system.

III. METHODOLOGY

The proposed research outlines an explainable machine learning approach for early prediction of students' academic performance by developing a system comprising of data preprocessing, predictive modeling, and interpretation mechanisms [1, 5]. The presented method is implemented and applied on the Open University Learning Analytics Dataset (OULAD) [2]. This dataset includes student demographical information, student's VLE interaction and student's assessment.

In the first step, data preprocessing is performed where the dataset is cleansed to identify and deal with missing values, outliers and to transform categorical attributes to a numeric representation compatible with the machine learning algorithms [6, 8]. Then feature engineering is carried out and useful features such as the students' engagement level, their scores and frequency of interaction with the course are extracted which will aid the prediction at the early stages of their academic tenure [3].

The Random Forest algorithm has been used as the main predictive model given its robustness, ability to capture non-linear dependencies in data and its good generalization capabilities [7]. The model is trained at an early stage of students' participation to enable quick identification of students at risk. Standard metrics such as accuracy, precision, recall, and F1-score have been used to evaluate the model performance [4, 6].

The SHAP values are computed so as to provide the explanation of model predictions by highlighting the impact of individual features on each prediction at the local and global level, thus providing a higher transparency [5]. In the last step



the system has been implemented by creating a Streamlit based dashboard where educators can input student information and generate predictions with explanation to take better academic decisions [1, 3].

A. Evaluation Metrics

To monitor how good is the model or what is the effectiveness of it, some evaluation metrics can be employed. In this paper we take into account the following:

Accuracy:

Accuracy measures the percentage of correctly classified predictions of the model. This indicates how many of the prediction made by model is correct, which might not be a good parameter to choose for the imbalanced datasets.

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

Precision

This can be done to assess the correctness of the model in terms of class prediction. Precision focuses on positive predictions:

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

Recall:

Recall measures the number of positive cases correctly predicted. The F1-score is a measure we should care about since it takes into account the precision and recall of a model and provides a single number to quantify both metrics combined. This measure becomes particularly significant when the cost of false positive is very different from the cost of false negative.

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

Once we see that a model is performing as it should, we have to look and see what the prediction and the observed output correspond to. We do this in an organized and clear manner so that we are able to see what the model predicted when and what actually happened so we know how the model performs in different circumstances. If we closely examine the outputs of the prediction, we are able to know that the model is performing as needed.

F1 Score:

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

B. Model Development Workflow

These algorithms are then compared to identify which produces the best prediction model. First, the pre-processed OULAD dataset (a cleaner and processed student data) is imported. Then early stage attributes (demographic attributes,

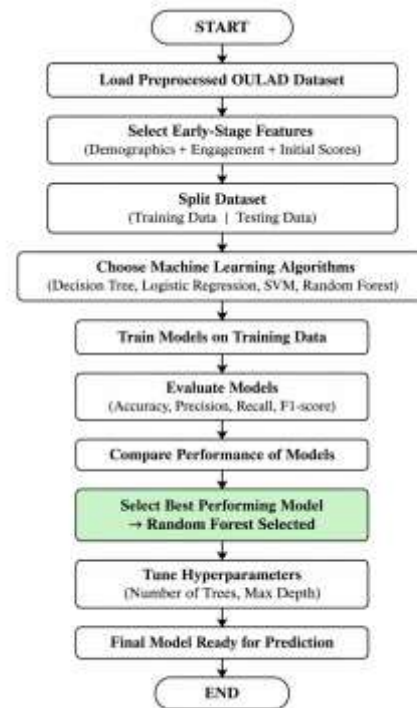


Fig. 1. Random Forest Model Training and Evaluation Pipeline for Early Student Performance Prediction

engagement attributes and assessment attributes in the beginning stages of course) are extracted for an early prediction.

Then the dataset is split into the training and testing data for building and evaluating the models. The machine learning algorithms of Decision Tree, Logistic Regression, SVM, Random Forest are used to select the best predictive model for the early stage prediction. Various algorithms are trained by using training data, and assessed on a number of evaluation metrics including accuracy, precision, recall, and F1-score.

A comparison is performed on the various algorithms to find the most effective predictive model. Random Forest is selected to be the best model in the evaluation after comparisons based on its accuracy. Further tuning of the Random Forest's hyperparameters (e.g., number of trees and maximum depth of trees) to improve its performance and effectiveness in prediction. The model developed is then built and used as the final prediction model.

IV. RESULTS AND DISCUSSION

According to the experimental outcomes, the early pre-

diction framework that we have designed is able to predict student academic achievements using limited early-stage data effectively. In terms of accuracy, precision, recall and F1-score, Random Forest outperformed the other classifiers that we have examined in this study which confirms its ability to process the complex non-linear data in education system.



TABLE I
 PERFORMANCE METRICS OF MODELS ON STUDENT ACADEMIC PERFORMANCE

Model	Acc.	Prec.	Recall	F1	Remarks
Logistic Regression	78.6	0.76	0.74	0.75	Performs well on linear relationships but limited in capturing complex patterns
Decision Tree	81.2	0.79	0.78	0.78	Easy to interpret but prone to overfitting
Naïve Bayes	76.8	0.75	0.72	0.73	Fast but assumes feature independence
SVM	83.5	0.82	0.80	0.81	Effective for classification but less interpretable
Random Forest	88.9	0.87	0.86	0.86	Best performance with strong balance and supports explainability (SHAP)

The features related to engagement and score on the early assessment contribute most significantly to prediction results. The study shows that behavioral features play critical roles in early prediction. In addition, SHAP could improve model interpretability as we know the contributions made by each feature.

A. Model Performance Evaluation

Using the typical metrics of accuracy, precision, recall and F1 score, various ML models such as Decision Tree, Logistic Regression, SVM, and Random Forest were tested. The result of testing shows that the ensemble learning approach yields better performance than a single model in handling the non-linear and complicate relationship present in the dataset. The highest performing model among all was the Random Forest classifier.

B. Early Prediction Effectiveness

The developed system is aimed to predict students’ performance at the beginning of the course based on a few features but with a high significance-students’ engagement and first assignment grades. The prediction was achieved with a reliable accuracy even with fewer amount of data. This would enable students who are at risk of failing the course to be identified as early as possible to receive assistance and compared to a prediction based on final semester’s performance, this system represents an advancement.

The designed system provides with an interactive dashboard on the web to help us predict the performance of students on real time and aids decision making for students. The users can feed the initial student activity using an easy to use interface which includes parameters such as mean score, score standard deviation, minimum/maximum score, number of assignments, total clicks and active days. All these are the major performance features for a student at the early stages of the course. The prediction is displayed as output with a

Advanced Student Performance Prediction System

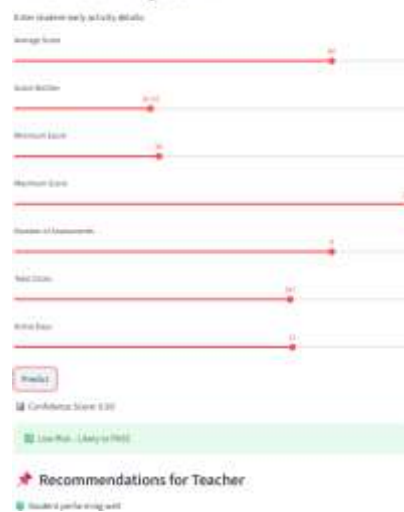


Fig. 2. Streamlit Dashboard for Real-Time Student Performance Prediction

score representing the confidence of the prediction made. From the dashboard the users can even determine the risk category for the student as shown, "Low Risk - Likely to PASS", and gain support on recommended action to help students based on teacher’s decision. The usability of this interface with such an intuitive visual presentation helps non-technical stake holders in understanding the predictions easily and thus helps teachers to gain insight and plan interventions for the students.

C. Feature Importance Analysis

Feature importance analysis was performed to investigate which features are the most important predictors of student performance. It was discovered that features associated with engagement, such as interaction frequency and activity level, along with initial assessment score, are important factors predicting student performance. Demographic features have some predictive value, while behavioral features are better indicators of student performance.

D. Explainability using SHAP

The Random Forest model’s predictions were explained using SHAP (SHapley Additive Explanations) to ensure model transparency. This includes global and local explanation and also visualizes how each feature contribute to a particular prediction. The model is made more explainable and educators are able to know which features are the most important for student success and thus more trust in the system.

E. Comparative Analysis

From the comparison of the implemented algorithms, Random Forest performs best in every aspect. Though models like Logistic Regression and Decision Tree give good results but

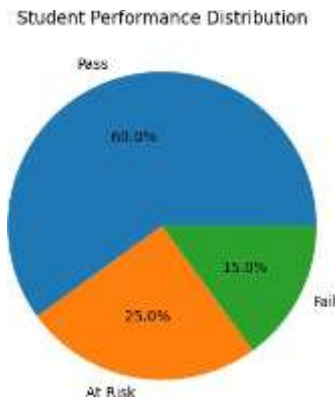


Fig. 3. Distribution of Student Performance Categories (Pass, At Risk, Fail)

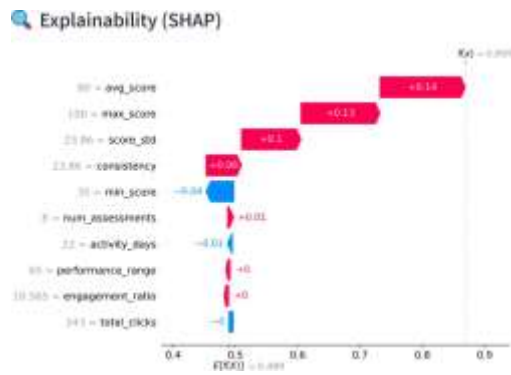


Fig. 4. Feature Contribution Analysis for Model Prediction Using SHAP

failed to get the underlying patterns of the dataset. SVM performed well but needs high computational power so Random Forest is the best algorithm to be selected for the present work.

F. Discussion of Findings

The results have showed that early prediction combined with explainable machine learning methods could potentially help to enhance the decision making process in the educational field. The predictive power of machine learning along with interpretability helps educators in gaining useful insights and formulating targeted interventions. Findings also have underlined the role of behavioral and engagement features in predicting student achievement.

V. CONCLUSION

This study proposed an explainable machine learning framework to predict students' academic performance in advance. The framework combines predictive modeling, explainability, and a human-friendly system. With the OULAD dataset, the proposed approach utilized the early features such as student

engagement, student attributes, and assessment information to make prediction earlier. In the experiment, Random Forest classifier shows a good performance to capture intricate patterns in data while keeping a high prediction accuracy.

SHAP-based explainability technique is another contribution in this work. By interpreting individual predictions with clear visualization, it shows the factors which contribute to the prediction results for each student. With this technique, teachers can easily understand the reasons why the student is at risk and provide timely interventions. Additionally, a web-based application which takes advantages of this framework into an interactive system makes it possible to build a system as a decision support tool in the real world.

In summary, this study showed that it is possible to enhance academic decision-making process and student support strategy by combining early prediction with explainable artificial intelligence. The proposed framework can provide a practical and scalable solution for schools to prevent student's dropout.

A. Future Work

Although the framework presented has been successfully shown to be effective for early prediction and explaining student academic success, there are a variety of areas that will be examined further in the future. Firstly, to allow the model to be generalized and robust, other datasets from various learning environments may be utilized.

To further advance the system, continuous updates can be made using data streams from learning management systems in real-time to monitor the students during the course. Machine and deep learning models can be incorporated to build on prediction abilities such as Gradient Boosting and Neural networks. Incorporating further psychological and behavioral traits such as motivation, learning style will also enhance student's predictions.

Recommendations of interventions can be recommended based on prediction and explainability. It can also be developed to better display information to users with an advanced GUI and dashboard and implemented on institutional academic systems for wider deployment and affect on students.

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*, vol. 40, no. 6, pp. 601–618, 2010.
- [2] R. S. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.
- [3] S. Kotsiantis, "Use of machine learning techniques for educational purposes: A decision support system for forecasting students' grades," *Artificial Intelligence Review*, vol. 37, no. 4, pp. 331–344, 2012.
- [4] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2009.



- [5] T. M. Mitchell, *Machine Learning*. New York, USA: McGraw-Hill, 1997.
- [6] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [7] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [8] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2011.
- [9] K. Koedinger, R. Baker, and K. Cunningham, “A data repository for the edm community: The pslc datashop,” in *Handbook of Educational Data Mining*, 2010.