International Journal on Engineering Technology (InJET)

www.kec.edu.np/kec-journal/
Volume 3, issue no. 1, Dec 2025, Page 100 - 108
Received Date: Jul 30, 2025
Accepted Date: Sep 24, 2025
ISSN: 3021-940X (Print)
DOI: *https://doi.org/10.3126/injet.v3i1.87015*

Predicting Student Academic Success through Explainable Machine Learning Models: A Comparative Study of BRF, XGBoost, and CatBoost

Prasish Timalsina[1*], Rajad Shakya[2]

[1]*Department of Electronics and Computer Engineering, Thapathali, Kathmandu, Nepal, prasishabde@gmail.com*
[2]*Department of Electronics and Computer Engineering, Thapathali, Kathmandu, Nepal, shakyarajad1@gmail.com*

Abstract

This study presents an explainable machine learning framework for predicting academic success among university students using Balanced Random Forest (BRF), XGBoost, and CatBoost. The dataset, collected from undergraduate computer science students in Bangladesh, encompasses academic, behavioral, and socio-demographic attributes. To handle class imbalance, SMOTE-based oversampling was applied, and model thresholds were optimized using the F1-score. Model interpretability was achieved using SHAP (SHapley Additive exPlanations), enabling transparent identification of influential features such as attendance, study habits, income, and academic progression. Among all models, CatBoost achieved the highest macro F1-score and ROC-AUC, demonstrating its robustness in handling heterogeneous educational data. The results highlight how interpretable ensemble learning can support early risk detection and data-driven academic interventions, thereby bridging the gap between predictive performance and educational transparency.

Keywords: Academic success prediction, Machine learning, XGBoost, Balanced Random Forest, CatBoost, SHAP

1. Introduction

The prediction and assessment of student academic performance have been some of the most serious research topics in educational data mining and learning analytics. With the increase in the number of higher education institutions and diversity of students, it is imperative to consider the vital factors affecting applicants' outcomes to formulate appropriate teaching strategies, personalized learning, and institutional decision-making.

Academic success may not be determined solely by cognitive ability or prior academic records, as it is affected by a complex and multivariate confluence of academic, behavioral, socio-economic, and psychological factors. In order to enable optimal planning, intervention, and timely support, these factors should be identified and studied at the earliest stages.

To address the growing need for transparent analytics in higher education, this study emphasizes the use of explainable machine learning models capable not only of generating accurate predictions but also of providing interpretable behavioral insights. Several theoretical perspectives—such as Cognitive Load Theory, which relates mental workload to learning performance; Human-Centered Design, which stresses actionable and understandable feedback for stakeholders; and Causal Inference Theory, which investigates feature–outcome relationships—highlight the importance of interpretability in educational settings. Guided by these principles, this work selects Balanced Random Forest, XGBoost, and CatBoost due to their superior performance on structured tabular data, their ability to handle class imbalance, and their compatibility with post-hoc explainability methods such as SHAP. This positions the study at the intersection of predictive performance and practical interpretability.

This research utilizes the *Students_Academic_Performance_Evaluation_Dataset*, collected from undergraduate students in the Department of Computer Science and Engineering at a private university in Bangladesh. The dataset contains a wide range of attributes, including demographic details, behavioral patterns, academic history, and personal challenges. Unlike traditional datasets that focus solely on grades or attendance, this dataset provides an inclusive view of each student, allowing for deeper insights into the factors that contribute to academic performance.

This study aims to use machine learning models to predict students' academic performance based on their Semester Grade Point Average (SGPA) and Cumulative Grade Point Average (CGPA). By integrating both nominal and quantitative features, the analysis intends to determine influential variables and evaluate the performance of different classification models. The results may contribute to the development of early alert systems and evidence-based institutional policies. Ultimately, this work seeks to support data-driven strategies that encourage student development and academic achievement.

2. Related Works

Predicting student academic success using machine learning has been explored through algorithms such as Random Forest, XGBoost, and deep neural networks. Ensemble-based techniques have consistently shown superior performance, particularly when combined with class balancing strategies like SMOTE or SMOTE-Tomek (Guanin-Fajardo, et al., 2024) (Mduma, 2023). XGBoost has demonstrated higher predictive accuracy than traditional models across structured academic datasets (Hakkal & Lahcen, 2024), while boosting models with Optuna-based hyperparameter tuning outperformed others in comparative evaluations (Villar & de Andrade, 2024).

Deep learning models have also shown promise. Attention-based Bi-LSTM architectures have improved GPA prediction and enabled interpretability through SHAP analysis (Kalita, et al., 2025). Similarly, deep neural networks have achieved up to 96% accuracy in early academic performance prediction, particularly in European institutions (Alnasyan, et al., 2024). Novel behavioral data approaches—such as analyzing campus movement and activity—have also enhanced predictive mode (Yao, et al., 2019).

Multimodal data fusion using LMS logs, psychological profiles, and academic metadata has further improved model performance in a blended learning environment (Chango, et al., 2024) (Orji & Vassileva, 2022). Random Forest has also proven robust for predicting academic major and student success across institutions (Beaulac & Rosenthal, 2019). Comparative studies of explainability tools, including LIME and SHAP, revealed interpretability variations across different models (Swamy, et al., 2022). Additionally, advanced tuning with CatBoost and ADASYN has yielded strong F1-scores in dropout prediction systems (Marcolino, et al., 2025).

Existing research has employed various model-agnostic explainability methods, including LIME, DeepLIFT, and Permutation Importance, to interpret student performance predictions. However, these approaches often lack consistency across feature distributions or fail to attribute contributions fairly at the instance level. SHAP, grounded in cooperative game theory, offers additive feature attributions and local instance reasoning, making it well-suited for educational intervention scenarios. Consequently, this study adopts SHAP as the primary explainability framework to ensure transparent and equitable feature influence assessment across heterogeneous student profiles.

Overall, these studies emphasize the importance of data balancing, explainable AI, and model comparison in building effective academic success prediction systems.

3. Related Theory

3.1. Balanced Random Forest (BRF)

Balanced Random Forest is a variant of the traditional Random Forest designed to handle imbalanced classification problems. It builds multiple decision trees using bootstrap samples, but for each tree, it performs random under-sampling of the majority class to balance the class distribution. The model aggregates predictions through majority voting: $\hat{y} = \text{mode}\big(f_1(x), f_2(x), \dots, f_T(x)\big)$.

Each tree is trained on a balanced subset to ensure that the minority class is fairly represented, which improves the model's sensitivity to rare classes.

3.2. Extreme Gradient Boosting (XGBoost)

XGBoost is an advanced implementation of gradient boosting that uses a second-order Taylor expansion to approximate the loss function and includes regularization to prevent overfitting. The model iteratively builds an ensemble of decision trees, minimizing the objective function:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t)$$

Where $g_i = \partial_{\left(\hat{y}_i^{(t-1)}\right)} l\left(y_i, \hat{y}_i^{(t-1)}\right) and h_i = \partial^2_{\left(\hat{y}_i^{(t-1)}\right)} l\left(y_i, \hat{y}_i^{(t-1)}\right)$ are the first and second derivatives of the loss. The regularization term is:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2$$

which controls the number of leaves *T* and the leaf weights w_j, promoting simpler trees.

3.3. CatBoost

CatBoost is a gradient boosting algorithm optimized for categorical data. Unlike traditional boosting methods, it uses ordered boosting and permutation-driven target statistics to reduce overfitting and prevent target leakage. The prediction is updated at each iteration *t* as: $\hat{y}^{(t)} = \hat{y}^{(t-1)} + \eta f_t(x)$ where f_t is the new tree and η is the learning rate. CatBoost processes categorical variables internally by converting them to numerical representations using combinations of category values and their statistics (e.g., target means). This improves performance on datasets with many non-numeric features.

From a methodological standpoint, this study extends conventional learning analytics frameworks by combining three complementary components: class rebalancing through SMOTE, decision threshold tuning using the F_1-score, and post-hoc model explainability via SHAP. This integration not only enhances predictive stability under class imbalance but also aligns with transparent and human-interpretable machine learning principles. Theoretically, it bridges the gap between statistical optimization and educational interpretability by leveraging ensemble learning theories (Bagging and Boosting) alongside Shapley value theory from cooperative game theory. Together, these elements form a reproducible and theoretically grounded framework that promotes equitable model reasoning and supports actionable educational insights.

4. Dataset Description

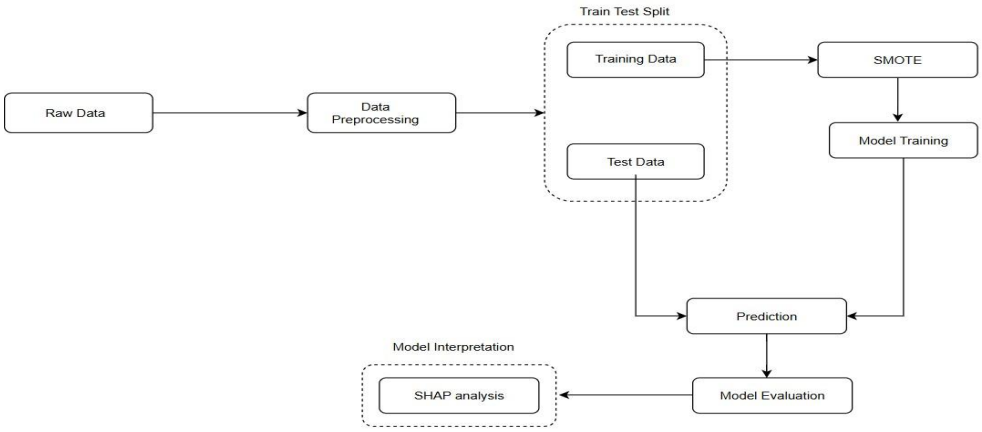
The dataset titled *"Students Academic Performance Evaluation Dataset"* is available from Mendeley Data, a well-known open-access research data repository. It contains academic, as well as non-academic details from students of the Department of Computer Science and Engineering at a private university in Bangladesh. An online survey form was used to collect the information which was then stored in an Excel file inside a folder named *"Students Performance Data Set"*. This file consists of responses from 1195 students-the number of rows and columns is 1195 and 31, respectively-giving 37,045 distinct data points. Each data point stands for a student and his or her report on academic background, behavioral patterns, as well as personal circumstances. The dataset consists of 31 features categorized into two types: nominal (categorical) and integer (quantitative) features. Table 1 summarizes them.

Table 1. Categorization of Dataset Features

Nominal (Categorical) Features	**Integer (Quantitative) Features**
Gender	HSC (Higher Secondary Certificate) passing year
Merit scholarship status	University admission year
Use of university transportation	Study hours per day
Learning mode (online/offline/hybrid)	Frequency of study per day
English language skills	Time spent on social media (daily)

Academic probation and suspension status	Class attendance percentage
Faculty consultancy participation	
Relationship status	
Participation in extracurricular activities	
Living status (with family, hostel, etc.)	
Political engagement Health issues	
Physical disability	
Semester Grade Point Average (SGPA)	
Cumulative Grade Point Average (CGPA)	

5. Methodology



6.

Figure 1. System Architecture

6.1. Data Preprocessing

The dataset was cleaned by standardizing columns and converting attendance data into numeric percentages. Income data was converted to numeric, categorized into income groups, and log-transformed. Important categorical features such as scholarship status, transportation use, probation, teacher consultancy, and English proficiency were encoded numerically. Living arrangements were grouped into standard categories. After transformations, original columns were dropped. Records with inconsistent or zero academic scores were removed, and a binary target variable was created where CGPA values of 3.0 or higher were labeled as 1 (success) and values below 3.0 as 0 (failure). These steps prepared the data effectively for classification modeling.

Distribution After Preprocessing

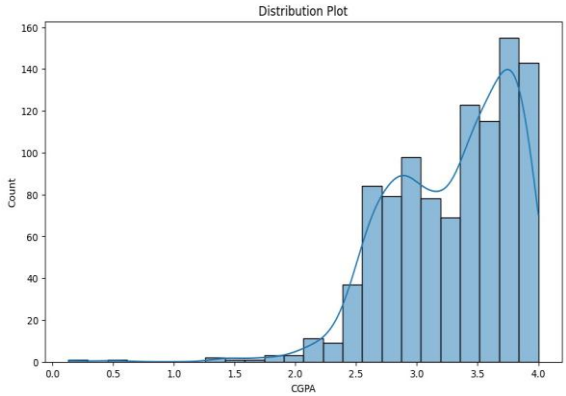


Figure 2. CGPA distribution

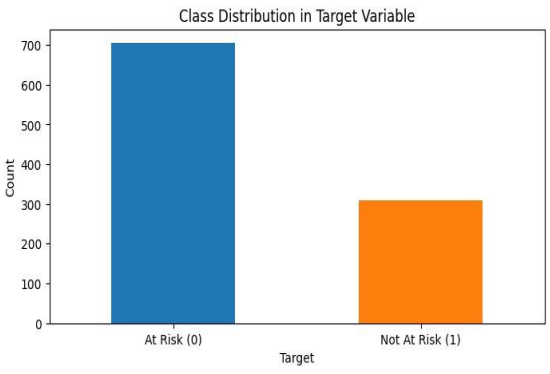


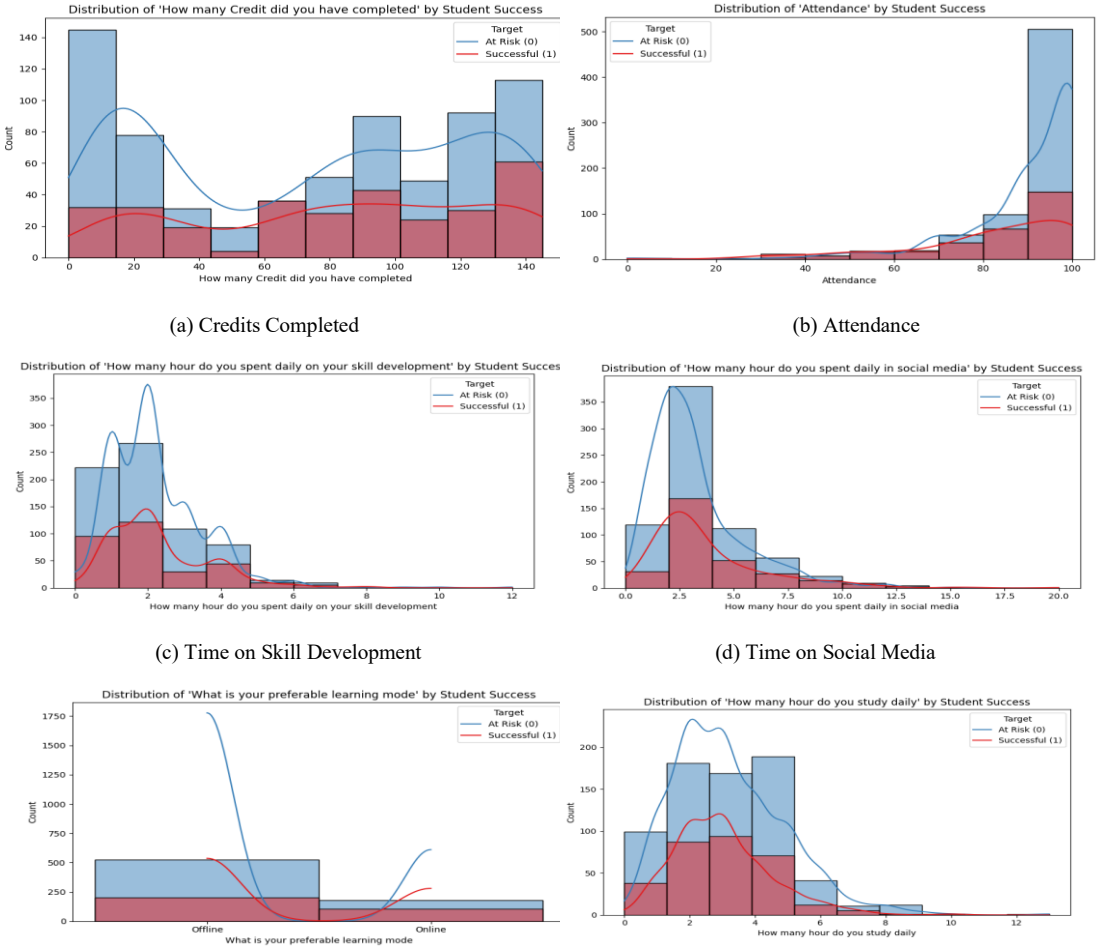
Figure 3. Class distribution

Figure displays the CGPA distribution after preprocessing. The plot indicates a slightly right-skewed distribution, with the majority of students having CGPA values above 3.0.

Figure shows the binary distribution of the 'Target' variable, where class '0' represents at-risk students and class '1' indicates not-at-risk students. As evident from the figure, the dataset is imbalanced, necessitating the use of SMOTE-based resampling.

Feature Distribution by Student Success

To better understand the relationship between student features and academic success, the distributions of six key variables were analyzed, divided by target label (0 = At Risk, 1 = Successful). The plots below show how these features vary across the two groups.



(a) Credits Completed

(b) Attendance

(c) Time on Skill Development

(d) Time on Social Media

(e) Learning Mode

(f) Time on Study

Figure 2. Distribution of selected features by student success (Target: 0 = At Risk, 1 = Successful)

6.2. Dataset Splitting and Handling Imbalance

The cleaned dataset was split into training and testing subsets using an 80-20 stratified split to preserve class distribution. To mitigate the issue of class imbalance in the training set, the Synthetic Minority Oversampling Technique (SMOTE) was applied. SMOTE generated synthetic samples for the minority class (at-risk students), thereby improving model sensitivity toward underrepresented outcomes and helping avoid bias toward the majority class.

6.3. Model Development and Configuration

Categorical variables were one-hot encoded, and numerical features were standardized. To address class imbalance, SMOTE was applied to the training data, generating synthetic samples for the minority classes. The Balanced Random Forest model was configured with 300 estimators and the sqrt feature selection strategy. For XGBoost, hyperparameters such as learning rate, max depth, gamma, subsample, and

colsample_bytree were optimized using GridSearchCV with 5-fold stratified cross-validation. The final model was retrained on the resampled dataset using the best parameters. Both models were trained on the scaled and resampled training data.

6.4. Model Evaluation

The models were evaluated using several standard classification metrics to assess performance across both the majority (successful) and minority (at-risk) student groups. The metrics used include:

- **Precision**: The proportion of correctly predicted positive cases among all predicted positives.
- **Recall**: The proportion of correctly predicted positive cases among all actual positives.
- **F1-Score**: The harmonic mean of precision and recall.
- **Macro Average F1**: Unweighted average F1-score across classes.
- **ROC AUC**: Area under the Receiver Operating Characteristic curve, measuring the model's ability to distinguish between classes.

7. Result and Discussion

Beyond evaluating model performance, this study provides a methodological contribution by combining SMOTE-based oversampling, F1-score–oriented decision threshold tuning, and SHAP instance-level interpretability into a unified workflow for academic risk prediction. While prior studies typically rely on accuracy-based evaluation or generic feature importance, this integrated approach prioritizes minority-class recall, supports early intervention scenarios, and offers transparent, student-specific explanations. Thus, the work contributes both practically and theoretically to learning analytics by demonstrating how interpretable ensemble models can be adapted to imbalanced educational datasets.

7.1. Model Performance Evaluation

In this study three machine learning models: Balanced Random Forest (BRF), XGBoost and CatBoost for predicting student academic success. The models were optimized using techniques such as SMOTE oversampling and threshold tuning, with a particular emphasis on identifying students at academic risk (denoted as class 0). Model performance was measured using metrics such as precision, recall, F1-score, and ROC-AUC.

Table 2. Performance Comparison of BRF, XGBoost, and CatBoost Models

Metric	Model	Class 0 (At Risk)	Class 1 (Successful)	Macro Avg
Precision	BRF	0.4478	0.7376	0.5927
	XGBoost	0.4746	0.7643	0.6195
	CatBoost	**0.4776**	**0.7800**	**0.6288**
Recall	BRF	0.4839	0.7376	0.6108
	XGBoost	0.4516	**0.7801**	0.6159
	CatBoost	**0.5161**	0.7518	**0.6340**
F1-Score	BRF	0.4651	0.7389	0.6080
	XGBoost	0.4628	**0.7710**	0.6174
	CatBoost	**0.4961**	0.7656	**0.6307**
ROC AUC	BRF	0.6465		

XGBoost	0.6499
CatBoost	**0.6810**

Table 2 summarizes the classification performance of Balanced Random Forest (BRF), XGBoost, and CatBoost across precision, recall, F1-score, and ROC AUC.

Among the three, CatBoost showed superior results, especially for the minority class (Class 0: At Risk), achieving the highest F1-score (0.4961) and macro averages across all metrics. XGBoost also performed competitively, outperforming BRF in most categories. While BRF provided balanced results, its effectiveness for the at-risk class was relatively lower.

Overall, CatBoost proved to be the most effective model in identifying at-risk students, with the highest ROC AUC (0.6810), suggesting better class separability.

7.2. SHAP-Based Feature Importance Analysis

Unlike prior work that relies solely on global feature rankings, this study employs SHAP waterfall plots to provide per-student explanatory profiles, enabling individualized reasoning behind predicted academic risk. This instance-level attribution supports targeted interventions rather than generic insights, representing a practical interpretive enhancement over previous studies.

To interpret the decisions made by the models, SHAP (SHapley Additive exPlanations) analysis was performed. It helps identify how much each feature contributes to the model's output.

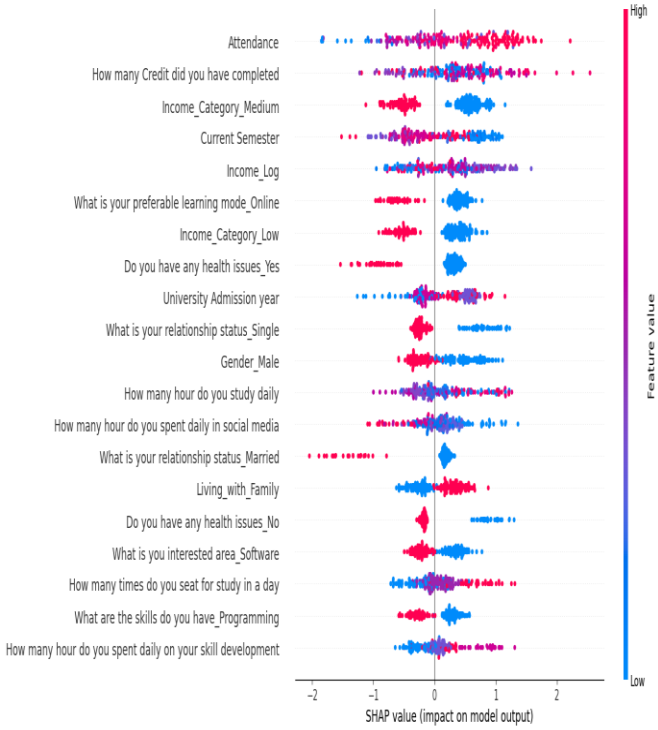


Figure 3. SHAP Summary Plot: XGBoost

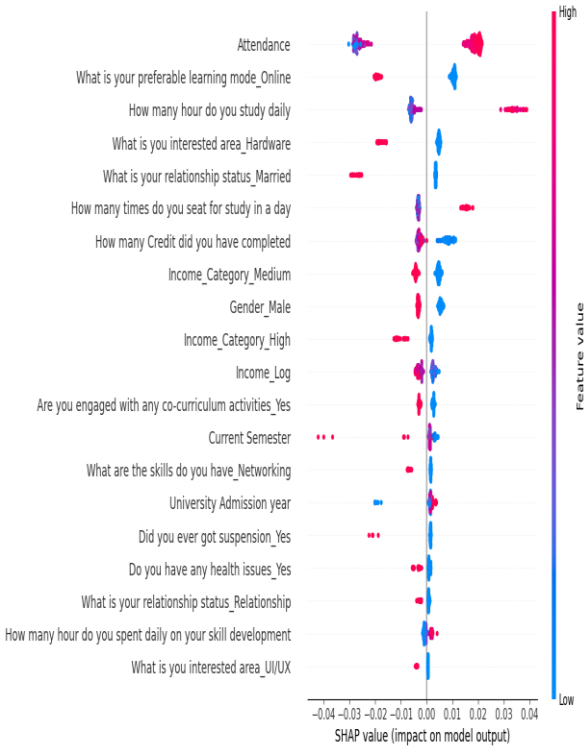


Figure 4. SHAP Summary Plot: BRF

Figure 5. SHAP Summary Plot: CatBoost

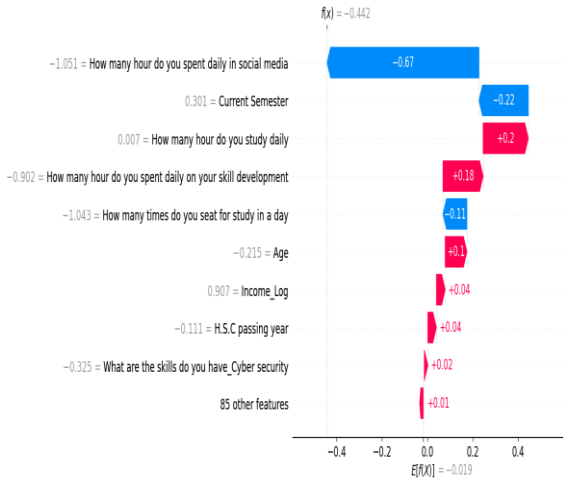


Figure 6. Interpretation of CatBoost prediction using SHAP

The SH AP analysis reveals Attendance as the most predictor across all models (XGBoost: Figure 3, Balanced Random Forest: Figure 4, CatBoost: **Error! Reference source not found.**), with XGBoost emphasizing **Completed Credits** and **Income Category**, Random Forest showing weaker but broader influences from **Learning Mode Preference** and **Study Hours**, and CatBoost demonstrating stronger dependence on **Current Semester** and **Income Log** with notably higher impact magnitudes compared to the other models.

All three models consistently identify attendance as the most critical behavioral factor, while differing in their treatment of academic progression and socioeconomic indicators. XGBoost and CatBoost show clearer feature differentiation compared to Random Forest.

Figure 6 illustrates a SHAP waterfall plot for a test instance classified by the CatBoost model. The base value $E[f(X)] = -0.019$ represents the average model output (log-odds) across all predictions. For this specific instance, the final prediction $f(x) = -0.442$ is obtained by summing the individual SHAP contributions of each feature. The most influential feature was **Social Media Time**, which contributed a strong negative impact (−0.67), driving the prediction toward the negative class (at-risk). Other negatively contributing factors include **Study Sessions**, **Cyber Security Skills**, and **Age**. Conversely, features like **Current Semester**, Skill **Development Time**, and **Income (Log)** provided positive SHAP values, pushing the prediction toward class 1 (successful).

This localized interpretability shows how even if certain features support success, dominant negative influences can drive the prediction toward the at-risk category.

8. Conclusion and Future Enhancement

This research implemented Balanced Random Forest, XGBoost and CatBoost classifiers to predict student academic success, with special focus on identifying at-risk students. Class imbalance was addressed using SMOTE, and models were optimized using threshold tuning based on the F_1-score. While all three models showed moderate performance, XGBoost slightly outperformed Balanced Random Forest in terms of macro-average F_1-score and ROC AUC. However, CatBoost achieved the best overall performance across most metrics. Specifically, CatBoost yielded the highest precision (0.4776), recall (0.5161), and F_1-score (0.4961) for the minority class (Class 0 – At Risk), as well as the highest macro-average F_1-score (0.6307) and ROC AUC (0.6810). These results suggest that CatBoost, with its ability to handle categorical features natively and robust regularization, was most effective in identifying at-risk students while maintaining balanced performance across both classes. SHAP analysis provided insights into the most influential features such as study time, credit completion, and learning mode.

Future work may incorporate larger and more diverse datasets spanning multiple institutions to improve generalizability. Additional resampling strategies, such as SMOTE-Tomek or ADASYN, can be explored to further mitigate noise near class boundaries. Deep learning architectures—including attention-based networks or graph neural models—could capture temporal and relational dependencies in student behaviors. Furthermore, causal inference techniques may help differentiate correlation from causal impact in educational

features. Finally, integrating explainability into user-centered academic dashboards could support proactive institutional interventions and improve stakeholder trust in AI-assisted decision-making.

9. Acknowledgement

I would like to express my sincere gratitude to Er. Rajad Shakya for his invaluable guidance, constructive feedback, and continuous support throughout the course of this research. His expertise and encouragement were instrumental in shaping the direction and quality of this study. Finally, I extend my appreciation to all individuals and colleagues who directly or indirectly contributed to the successful completion of this research.

References

Alnasyan, B., Basheri, M. & Alassafi, M., 2024. The power of Deep Learning techniques for predicting student performance in Virtual Learning Environments: A systematic literature review. *Computers and Education: Artificial Intelligence,* Volume 6, p. 100231.

Beaulac, C. & Rosenthal, J. S., 2019. Predicting university students' academic success and major using random forests. *Research in Higher Education,* 60(5), pp. 733--754.

Chango, W., Cerezo, R. & Romero, C., 2024. Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses. *arXiv preprint.*

Guanin-Fajardo, J. H., Guaña-Moya, J. & Casillas, J., 2024. Predicting academic success of college students using machine learning techniques. *Data,* 9(4), p. 60.

Hakkal, S. & Lahcen, A. A., 2024. XGBoost to enhance learner performance prediction. *Computers and Education: Artificial Intelligence,* Volume 7, p. 100254.

Kalita, E. et al., 2025. Predicting student academic performance using {Bi-LSTM} with attention and {SHAP}. *Frontiers in Education.*

Marcolino, M. R. et al., 2025. Student dropout prediction through machine learning: A CatBoost approach with ADASYN and multi-objective tuning. *Scientific Reports,* Volume 15, p. 1234.

Mduma, N., 2023. Data balancing techniques for predicting student dropout using machine learning. *Data,* 8(3), p. 49.

Orji, F. A. & Vassileva, J., 2022. Predicting students' academic performance and study strategies based on their motivation. *arXiv preprint.*

Swamy, V. et al., 2022. Evaluating the Explainers: Black-Box Explainable Machine Learning for Student Success Prediction in MOOCs. *arXiv preprint.*

Villar, A. & de Andrade, C. R. V., 2024. Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study. *Education and Information Technologies.*

Yao, H. et al., 2019. Predicting academic performance for college students: A campus behavior perspective. *arXiv preprint.*