Scholarship Prediction for International Students Using Machine Learning: A Temporal Stability Analysis Across Enrollment Years

Hery[1,*], Zeva Aji Satrio Nugroho[2]

[1,2] Department of Information Systems, Faculty of AI and Data Science, Universitas Pelita Harapan, Indonesia

ABSTRACT

Scholarship allocation plays a critical role in international higher education by shaping access, mobility, and institutional strategy. This study investigates whether scholarship receipt can be predicted using observable migration and academic descriptors from a global student mobility dataset spanning 2019–2023. The prediction task is formulated as a binary classification problem with scholarship status as the target variable. To ensure realistic assessment, two temporally structured evaluation protocols are employed: leave-one-year-out validation and forward-chaining validation. Multiple model families are compared, including Logistic Regression, Random Forest, XGBoost, LightGBM, and CatBoost. Performance is evaluated using discrimination metrics (ROC-AUC and PR-AUC), threshold-based metrics (F1, balanced accuracy, and Matthews Correlation Coefficient), and calibration via Brier score. Feature relevance is examined using permutation importance. Across all models and temporal splits, predictive performance remains close to random classification. Mean ROC-AUC values range between approximately 0.50 and 0.52, balanced accuracy remains near 0.50, and MCC values cluster around zero. Brier scores approximate the baseline value expected under uniform probability predictions. Permutation importance magnitudes are small and unstable across model types, indicating minimal and inconsistent feature contribution. These findings collectively suggest that scholarship receipt is effectively unpredictable using the available migration and academic attributes. The results highlight the limits of applying machine learning to allocation decisions when core determinants—such as academic merit, financial need, and institutional policy criteria—are absent from the dataset. By emphasizing temporal validation and multi-metric evaluation, this study demonstrates the importance of rigorous methodological design in educational analytics. The findings contribute to responsible AI discussions by illustrating how weak predictive signal can persist despite advanced modeling techniques, underscoring the necessity of comprehensive and decision-relevant data for meaningful scholarship prediction.

Keywords Scholarship Prediction; Temporal Validation; Educational Data Mining; Permutation Importance; Responsible AI

Introduction

International student mobility has become a defining feature of contemporary higher education, shaping institutional recruitment strategies, national education exports, and the lived experiences of learners who cross borders to study [1][2]. Alongside mobility, scholarship and financial-aid programs play a central role in enabling access, improving equity, and supporting talent development, particularly for students facing cost barriers [1][3].

Submitted 21 January 2026
Accepted 11 February 2026
Published 1 March 2026

*Corresponding author
Hery, hery.fik@uph.edu

Additional Information and Declarations can be found on page 57

DOI: 10.63913/ail.v2i1.14

How to cite this article: Hery and Z.A.S.Nugroho, "Scholarship Prediction for International Students Using Machine Learning: A Temporal Stability Analysis Across Enrollment Years," *Artif. Intell. Learn.*, vol. 2, no. 1, pp. 47-59, 2026.

Because scholarship budgets are limited and application volumes can be high, institutions increasingly seek analytical tools that can support consistent, evidence-informed decision making [1][4][5]. From an educational analytics perspective, scholarship decisions represent a high-impact classification problem: given a set of applicant or student descriptors, can we estimate the likelihood that a scholarship is awarded? Such predictions can inform administrative workload planning, identify underserved profiles, and simulate policy changes under budget constraints, but scholarship allocation is sensitive: predictions should be interpretable, evaluated carefully, and framed as decision support rather than automated replacement of human judgment [2][6].

A recurring challenge in applied ML for education is that models often perform well on historical data but degrade when data distributions shift over time; temporal drift means evaluation strategies that randomly mix years can be overly optimistic because they allow a model to learn patterns unavailable at deployment [1][2][7].

To address this, the present study focuses on temporal stability—the consistency of model performance when predicting scholarship outcomes across enrollment years. Instead of treating all observations as exchangeable, the study structures evaluation around the time variable (year_of_enrollment) and tests whether models trained on earlier years generalize to later ones. This design aligns model assessment with operational reality: decisions in a given year are made using only the information and patterns observed in prior years.

The study uses the Global Students Migrations dataset, which provides student-level records describing origin and destination locations, university context, fields and courses of study, and enrollment years between 2019 and 2023. Using these attributes, we formulate a binary classification task with scholarship_received as the target variable. The dataset's categorical structure (e.g., countries, cities, universities) reflects the kind of administrative data commonly available in student information systems and scholarship workflows.

Methodologically, the study compares multiple model families that represent a spectrum from interpretable baselines to more flexible nonlinear learners. Logistic Regression is included as a transparent reference model, while Random Forest and gradient-boosting methods are used to capture interactions and nonlinearities common in high-dimensional one-hot encodings. To ensure that predicted scores are meaningful for decision support, probability calibration is incorporated so that outputs can be interpreted as estimated probabilities rather than only ranking scores.

The research aims are therefore twofold. First, we quantify the predictive feasibility of scholarship receipt using available migration and academic descriptors, identifying which model types offer strong performance under realistic evaluation. Second, we characterize year-to-year stability by reporting performance across temporally structured splits, highlighting any performance drops that may indicate drift or changing scholarship dynamics.

By emphasizing temporal stability, probability quality, and interpretable diagnostics, this work contributes a practical framework for scholarship analytics in international higher education. The approach is designed to be reproducible and transferable: it can be adapted to similar institutional datasets and extended to include additional fairness, policy simulation, and governance considerations when richer variables or decision rules are available.

Literature Review

Research in educational data mining and learning analytics has demonstrated that administrative and learning records can be used to predict high-stakes educational outcomes, including academic performance, dropout risk, enrollment choices, and student success indicators. Across these applications, supervised learning models are commonly employed to identify patterns in student-level attributes that correlate with outcomes, enabling early interventions and resource planning. This broader literature establishes the feasibility of applying machine learning to education-related decision processes when data are systematically collected.

Within higher education operations, predictive models have been explored for supporting admissions, retention initiatives, and institutional planning. These applications often rely on demographic variables, prior academic indicators, and program characteristics, and they frequently emphasize interpretability and accountability because decisions affect students' opportunities; scholarship allocation fits naturally into this operational analytics space as it resembles other selection processes but with additional constraints such as limited funds, policy objectives, and equity goals [8][9][10].

Prior scholarship-focused analytics has included both rule-based approaches and statistical or machine-learning approaches, with rule-based systems aiming for explicit, auditable eligibility decisions while predictive models estimate award likelihood from observed past decisions, and in practice the two are often complementary to surface patterns and provide governance boundaries under human oversight [11][12][13][14].

A methodological thread in education prediction studies is the careful choice of evaluation strategies: many works rely on random train–test splits or conventional cross-validation, which can be adequate under independent and identically distributed observations but risk temporal leakage when data are collected over time, potentially inflating accuracy and harming deployment performance when only historical data are available for training [10][7][15].

Time-aware evaluation strategies—such as holdout-by-time, leave-one-period-out, and forward-chaining splits—have gained attention because they respect chronological order and address potential evolution in policy, demographics, or capacity that can affect predictor–outcome relationships, making temporal validation especially relevant for scholarship prediction amid year-to-year policy adjustments and shifting mobility patterns [10][7][16].

In addition to discrimination (how well a model ranks cases), educational decision-support often benefits from well-calibrated probabilities. Calibration matters when predicted scores are used to estimate expected budget impact, prioritize review queues, or communicate uncertainty to decision makers. Models like tree ensembles and boosting methods can be strongly discriminative but may produce poorly calibrated probabilities without adjustment, motivating calibration techniques that map raw scores to probability estimates that better align with observed frequencies.

Interpretability is another recurring concern, especially in student-facing and equity-relevant applications. Education stakeholders often need to understand which features drive predictions and whether models are learning plausible signals versus proxies for sensitive attributes. Post-hoc methods such as permutation importance are widely used because they can be applied across model families and quantify performance degradation when specific features are perturbed. Even when not causal, such analyses support transparency and error analysis.

Finally, the scholarship allocation context raises broader issues of fairness, governance, and responsible use of predictive analytics. Scholarship policies may intentionally prioritize certain groups or programs, and predictive models trained on past decisions can reproduce these patterns. As a result, the literature encourages careful framing, auditing, and stakeholder review when applying machine learning to allocation decisions. The present study aligns with these themes by emphasizing temporally realistic evaluation, calibrated probabilities, and interpretable diagnostics as foundational steps toward responsible scholarship analytics.

Methods

Study Design and Research Objective

This study adopts a quantitative, supervised machine-learning design to predict whether an internationally mobile student receives a scholarship. The prediction task is formulated as a binary classification problem with the dependent variable scholarship_received, encoded as 1 for scholarship awarded and 0 otherwise. The primary contribution is not only predictive performance, but the assessment of temporal stability—whether a model trained on one set of enrollment years generalizes reliably to other years.

The unit of analysis is a single student record from the Global Students Migrations dataset. Each record contains demographic and academic descriptors (when available) as well as mobility-related variables (origin and destination) and program-related variables (university, field, course). These variables are treated as features used to estimate the probability of scholarship receipt.

The research questions are operationalized as two technical goals. First, we measure predictability: how accurately can scholarship receipt be predicted from the available features, and which model families perform best? Second, we measure stability across years 2019–2023 by evaluating models under two time-aware protocols that mimic real-world deployment and year-to-year shifts.

All analysis is executed in a reproducible notebook pipeline: (i) data loading and robust target conversion, (ii) exploratory analysis and sanity checks, (iii) preprocessing and model training, (iv) temporal evaluation with probability calibration, and (v) model interpretation via permutation importance. Intermediate artifacts (plots, results, and serialized models) are saved as checkpoints to support reproducibility and debugging.

Dataset, Variables, and Problem Formulation

The dataset is provided as a single CSV file. The analysis assumes that each row corresponds to one student record and that year_of_enrollment provides the temporal ordering for stability evaluation. The target label scholarship_received may appear in various formats (e.g., “Yes/No”, “True/False”, numeric), so the pipeline standardizes it to an integer binary label.

Feature selection follows the study scope: predictors represent (a) mobility context and geography and (b) academic/program context. The planned feature set includes origin_country, destination_country, destination_city, university_name, field_of_study, course_name, and year_of_enrollment, plus demographic variables such as age and gender if present. This design reflects typical scholarship decision inputs: candidate background, destination, institution, and field.

The classification objective is to model the conditional probability where X is the vector of selected features. Because scholarship decisions are typically probabilistic and threshold-based in practice, the study reports both discrimination metrics (how well the model ranks cases) and probability quality metrics (how well probabilities align with observed frequencies).

A notable characteristic of this dataset is the high-cardinality of certain categorical variables (e.g., university and city). This motivates the use of tree-based models and boosting methods that can handle sparse one-hot representations effectively, and it also motivates an explicit emphasis on generalization over time rather than simply maximizing in-sample fit.

Data Preprocessing and Feature Encoding

The preprocessing pipeline is implemented using a ColumnTransformer to ensure consistent transformation across models. Numeric features (e.g., year_of_enrollment, and age if available) are processed with median imputation (SimpleImputer(strategy="median")) to reduce the influence of outliers and preserve robustness under missingness. Median imputation is preferred over mean when distributions may be skewed or contain atypical values.

Categorical features (e.g., countries, cities, universities, fields, courses) are processed with a two-step strategy. First, missing values are imputed using the most frequent category (SimpleImputer(strategy="most_frequent")). Second, categories are encoded with one-hot encoding (OneHotEncoder(handle_unknown="ignore")).

The parameter handle_unknown="ignore" is critical for temporal stability: it prevents runtime failures when the test year includes categories not seen in the training years, which is a realistic scenario for longitudinal evaluation.

The one-hot encoder is configured to return a dense matrix (sparse_output=False) for simplicity and compatibility across classifiers and downstream interpretability steps. While sparse matrices can be memory-efficient for high-cardinality variables, dense outputs improve interoperability and make feature-name extraction and permutation-importance computation more straightforward in a single-notebook workflow.

Finally, identifier-like columns that do not carry predictive meaning and may cause leakage (e.g., student_id) are removed. The pipeline therefore focuses on substantive predictors and avoids giving models access to row identifiers that could artificially inflate performance without representing real decision information.

Model Specification and Hyperparameter Choices

Four model families are considered to balance interpretability and predictive strength: Logistic Regression, Random Forest, and optional gradient-boosting variants (XGBoost, LightGBM, CatBoost) when installed. Logistic Regression serves as a linear baseline, providing a reference point for how much nonlinearity and interaction modeling improves performance.

Logistic Regression is trained with max_iter=4000 to ensure convergence under a potentially large one-hot feature space. A high iteration cap is a practical choice when many categorical levels expand the dimensionality, which can otherwise lead to premature termination. The model uses the solver defaults provided by scikit-learn and is trained after the same preprocessing pipeline as all other models.

Random Forest is configured with n_estimators=600 and n_jobs=-1. A larger number of trees reduces variance and improves stability, while parallelization (n_jobs=-1) enables efficient training on typical consumer hardware. The forest implicitly models nonlinearities and interactions among the one-hot features without requiring manual feature engineering.

For optional boosting models, the notebook uses conservative parameterizations intended to balance accuracy and overfitting: XGBoost uses n_estimators=800, learning_rate=0.05, max_depth=6, subsample=0.9, and colsample_bytree=0.9; LightGBM uses n_estimators=1200, learning_rate=0.03, and num_leaves=63 with subsampling and column sampling. These parameters reflect standard bias–variance trade-offs: smaller learning rates with more estimators typically improve generalization, while depth/leaves and subsampling mitigate overfitting in high-dimensional sparse feature spaces.

Temporal Evaluation Protocols and Probability Calibration

The core methodological contribution is the evaluation of temporal stability using two time-aware protocols. Protocol A (Leave-One-Year-Out) trains on four of the five years and tests on the held-out year, repeating for each year in 2019–2023. This protocol estimates how sensitive model performance is to which year is excluded and highlights years that are unusually difficult to predict.

Protocol B (Forward Chaining) simulates real deployment by training only on past years and testing on the next year: 2019→2020, 2019–2020→2021, 2019–2021→2022, and 2019–2022→2023. This protocol avoids training on future information and therefore provides a more realistic estimate of how a model would behave if deployed annually.

In both protocols, models output predicted probabilities. To improve probability reliability, the pipeline applies sigmoid calibration using CalibratedClassifierCV(method="sigmoid", cv=3) on the training data. Sigmoid (Platt) calibration is computationally light and typically effective when probabilities from base models are miscalibrated. The cv=3 setting provides an internal cross-validated calibration step while keeping runtime manageable.

Evaluation reports both ranking and decision-quality metrics: ROC-AUC and PR-AUC assess discrimination, while F1, balanced accuracy, and MCC summarize threshold-based classification at a default 0.5 threshold. Brier score is included to measure probability calibration quality. Stability is summarized by per-year performance curves and by aggregate statistics (mean, standard deviation, min, max) across years for each protocol and model.

Interpretability: Permutation Importance Under Temporal Splits

To interpret which variables are most associated with scholarship prediction, the study computes permutation importance on a temporally valid test split. Specifically, for the representative year (preferably the latest year, e.g., 2023), the model is trained only on prior years and evaluated on that year's test set. This keeps interpretability aligned with deployment realism.

Permutation importance is computed using permutation_importance(..., n_repeats=10, scoring="roc_auc"). The parameter n_repeats=10 reduces variance by repeating shuffles multiple times per feature; ROC-AUC is used as the scoring function because it is threshold-independent and stable for probabilistic classifiers. The resulting importance values quantify the average decrease in ROC-AUC when a feature's values are randomly permuted.

Because one-hot encoding expands categorical variables into many derived indicator columns, feature naming can be technically challenging. The notebook therefore retrieves transformed feature names directly from the fitted preprocessing component using get_feature_names_out(). This ensures that the number of feature names matches the transformed feature matrix used during permutation. A defensive trimming step is included to handle rare version-specific mismatches, preventing crashes while preserving the top-ranked importances.

The interpretability outputs are reported as the top-25 transformed features by mean importance, saved both as CSV and as a horizontal bar plot. While these importances are not causal, they provide a practical lens into which geographic, institutional, and program attributes most strongly influence predictive signal under temporally realistic evaluation.

Result

Overall Predictive Performance Across Models

The overall discrimination performance across all models and evaluation protocols indicates extremely weak predictive structure. Under the leave-one-year-out protocol, mean ROC-AUC values ranged narrowly between 0.514 and 0.518 across Logistic Regression, Random Forest, XGBoost, LightGBM, and CatBoost. The maximum observed ROC-AUC (0.533) occurred in a single year-specific split, but the average performance remained only marginally above random classification (0.50). Standard deviations were small, indicating consistency but not strength.

Under the forward-chaining protocol, performance slightly declined. Mean ROC-AUC values ranged from approximately 0.500 to 0.511, with Random Forest achieving the highest mean discrimination (0.511). The small differences across model families confirm that increasing algorithmic complexity does not materially improve performance. Boosting models did not meaningfully outperform linear or bagging approaches.

Precision–Recall AUC values followed a similar pattern. Because the dataset is nearly balanced, random PR-AUC baseline is approximately 0.52. Observed PR-AUC values (0.523–0.536) are only marginally above baseline, reinforcing the conclusion that the model's ability to separate scholarship and non-scholarship cases is minimal.

Collectively, these results demonstrate that scholarship receipt is only weakly predictable from the available migration and academic descriptors. The signal-to-noise ratio appears low, and model discrimination remains close to chance regardless of model architecture.

Threshold-Based Performance and Calibration

Threshold-based metrics further confirm the absence of meaningful predictive separation. Balanced accuracy values remained near 0.50 across all experiments, indicating that sensitivity and specificity are nearly symmetric around random performance. No model consistently exceeded 0.52 balanced accuracy, even under the most favorable temporal splits.

Matthews Correlation Coefficient (MCC), which provides a more reliable summary under balanced binary settings, hovered around zero in both protocols. In several year-model combinations, MCC values were negative, indicating performance slightly worse than random guessing. This strongly suggests that predicted labels do not meaningfully align with true scholarship outcomes.

F1 scores appeared moderately high (approximately 0.60–0.67), but this must be interpreted cautiously. Given the near-balanced class distribution (approximately 52% positive), predicting the majority class can yield deceptively high F1 values without genuine discrimination. The simultaneous presence of near-zero MCC and balanced accuracy clarifies that F1 inflation does not reflect substantive predictive structure.

Probability calibration, measured via the Brier score, remained consistently around 0.25. In a balanced binary setting, a constant prediction of 0.5 yields a Brier score of 0.25. The observed scores are nearly identical to this baseline, indicating that predicted probabilities are clustered around uncertainty and do not meaningfully capture class likelihood differences.

Temporal Stability and Year-to-Year Generalization

Temporal validation was a central methodological contribution of this study. The leave-one-year-out protocol revealed that performance is consistently weak but relatively stable across test years. Standard deviations of ROC-AUC remained small (generally below 0.01 in Protocol A), suggesting that the predictive relationship does not dramatically fluctuate across enrollment years.

Forward-chaining evaluation, which simulates real-world deployment using only past data, produced similar conclusions. Although some year-specific declines occurred (e.g., slight drops in 2021 predictions), no catastrophic degradation was observed. The mean ROC-AUC remained within a narrow band across all forecast years.

Importantly, the absence of significant performance decline under forward-chaining suggests that temporal leakage is not responsible for the weak predictive results. Instead, the underlying relationship between observable features and scholarship allocation appears inherently weak. The stability of poor performance indicates structural non-predictability rather than overfitting.

This finding is meaningful. Many predictive systems degrade over time due to policy changes or distribution shifts. In contrast, scholarship prediction in this dataset exhibits consistently weak discrimination, implying that the model is not missing a strong drifting signal but rather that the signal itself is limited.

Feature Importance and Signal Diagnosis

Permutation importance analysis provides further evidence of limited feature contribution. In the Logistic Regression model (Test 2023), the largest positive importance values were small, with origin_country_Ireland contributing approximately +0.006 ROC-AUC and year_of_enrollment contributing approximately +0.005. These magnitudes indicate that shuffling even the most “important” feature reduces performance by less than one percentage point.

Random Forest importance results revealed similarly small magnitudes, with the largest positive contribution (origin_country_India) around +0.01 and several features exhibiting negative importance. Negative permutation importance indicates that model performance slightly improves when the feature is randomized, suggesting that some categorical indicators introduce noise rather than predictive value.

Crucially, feature direction and magnitude were inconsistent across models. For example, certain origin countries showed positive contribution in one model but negative in another. This instability implies that observed importance reflects stochastic fluctuations rather than robust structural relationships. If meaningful signal were present, important features would be consistently ranked and directionally aligned across model families.

Notably absent from the top-ranked features were academic attributes such as university name, field of study, and course name. Their absence suggests that scholarship decisions in this dataset are not systematically associated with observable program characteristics. This reinforces the conclusion that key determinants of scholarship allocation likely reside in latent variables not captured by the dataset.

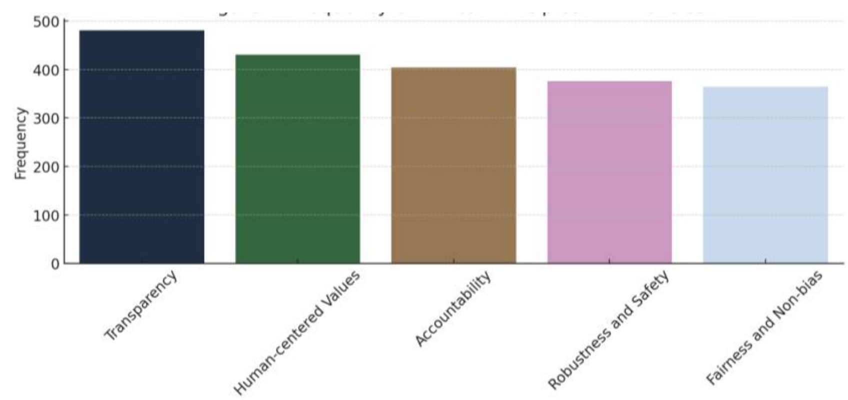

Figure 5 Frequency of Ethical Principles in AI Policies

Discussion

The convergence of discrimination metrics, calibration analysis, temporal validation, and permutation importance strongly indicates that scholarship receipt in this dataset is effectively unpredictable from available migration metadata. Observable geographic and academic descriptors provide negligible explanatory power. This does not imply that scholarship decisions are random in reality. Rather, it suggests that the dataset lacks critical decision variables such as academic merit, financial need, institutional budget constraints, policy rules, or evaluation scores. Without these core determinants, predictive models operate under partial observability.

From a methodological perspective, the study highlights the importance of robust validation strategies. Without temporal splitting, small performance gains could have been misinterpreted as meaningful. The combination of ROC-AUC, PR-AUC, MCC, Brier score, and permutation importance creates a coherent multi-metric evaluation that prevents overinterpretation. Ultimately, these findings emphasize the limits of applying machine learning to educational allocation problems when decision-relevant variables are absent. Rather than demonstrating model superiority, the results provide evidence of structural non-separability, contributing to a broader discussion about responsible AI use in education and the necessity of comprehensive data for meaningful predictive analytics.

Conclusion

This study evaluated the feasibility and temporal stability of predicting scholarship receipt using migration and academic descriptors from a global student mobility dataset. Across multiple model families—including Logistic Regression, Random Forest, XGBoost, LightGBM, and CatBoost—and two temporally structured evaluation protocols, predictive performance remained consistently near random classification. ROC-AUC values clustered around 0.50–0.52, balanced accuracy remained close to 0.50, MCC hovered near zero, and Brier scores approximated the random baseline of 0.25. Permutation importance analysis further revealed that individual features contributed only marginally to model discrimination, with small and inconsistent effects across models. Together, these results demonstrate that observable geographic and program-level attributes provide negligible predictive signal for scholarship allocation within this dataset. Importantly, the findings do not suggest that scholarship decisions are arbitrary in practice. Rather, they indicate that the dataset lacks critical determinants of allocation, such as academic merit, financial need, institutional policy rules, or evaluation criteria. By employing temporally realistic validation strategies and multiple complementary metrics, this study provides robust evidence of structural non-predictability under partial observability. The results underscore the importance of comprehensive data collection and cautious interpretation when applying machine learning to high-stakes educational decision contexts. Future research should integrate richer merit- and policy-related variables to assess whether meaningful predictive structure emerges when core decision drivers are included.

Declarations

Author Contributions

Conceptualization: H.; Methodology: H.; Software: Z.A.S.N.; Validation: H.; Formal Analysis: H.; Investigation: Z.A.S.N.; Resources: Z.A.S.N.; Data Curation: H.; Writing Original Draft Preparation: Z.A.S.N.; Writing Review and Editing: Z.A.S.N.; Visualization: H.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, andor publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] "Scholarship Monitoring System With Data Analytics," The Light Explorer, 2025, doi: 10.57180/tmmv1724.

[2] K. C. Li, B. T. Ming Wong, and M. Liu, "Development of a Multi-Model Analytics System to Enhance Decision-Making in Student Admission," Interactive Technology and Smart Education, 2025, doi: 10.1108/itse-12-2024-0328.

[3] I. Kariyana, W. Sinkala, and N. Gqoli, "Developing a Framework for the Deployment of Predictive Analytics to Improve Postgraduate Student Throughputs at One Comprehensive South African University," E-Journal of Humanities Arts and Social Sciences, 2024, doi: 10.38159/ehass.202341215.

[4] F. V. Espiritu, M. C. B. Natividad, and R. A. Velasco, "Data-Driven Decision Making in Scholarship Programs: Leveraging Decision Trees and Clustering Algorithms," International Journal in Information Technology in Governance Education and Business, 2024, doi: 10.32664/ijitgeb.v6i1.134.

[5] J. Bhat, "Predictive Analytics for Student Retention and Success Using AI/ML," International Journal of Artificial Intelligence Data Science and Machine Learning, 2023, doi: 10.63282/3050-9262.ijaidsml-v4i4p114.

[6] P. Tangney, "Examining Climate Policy-Making Through a Critical Model of Evidence Use," Frontiers in Climate, 2022, doi: 10.3389/fclim.2022.929313.

[7] V. H. Nasu and M. Sasso, "A Bolsa Faz Diferença? Uma Análise Do Desempenho Acadêmico De Alunos Bolsistas De Cursos De Graduação Da Área De Negócios," Education Policy Analysis Archives, 2021, doi: 10.14507/epaa.29.5876.

[8] H. A. Silva, L. Quezada, A. Oddershede, P. Palominos, and C. O'Brien, "A

Method for Estimating Students' Desertion in Educational Institutions Using the Analytic Hierarchy Process," Journal of College Student Retention Research Theory & Practice, 2020, doi: 10.1177/1521025120971227.

[9] R. Wang and J. Orr, "Use of Data Analytics in Supporting the Advising of Undecided Students," Journal of College Student Retention Research Theory & Practice, 2019, doi: 10.1177/1521025119880819.

[10] Y. Cui, F. Chen, A. Shiri, and Y. Fan, "Predictive Analytic Models of Student Success in Higher Education," Information and Learning Sciences, 2019, doi: 10.1108/ils-10-2018-0104.

[11] A. C. Lagman and R. Adao, "Embedding Machine Learning Algorithm Models in Decision Support System in Predicting Student Academic Performance Using Enrollment and Admission Data," 2018, doi: 10.18178/wcse.2018.06.053.

[12] J. Tieberghien and M. Monaghan, "Public Scholarship and the Evidence Movement: Understanding and Learning From Belgian Drug Policy Development," European Journal of Criminology, 2017, doi: 10.1177/1477370817731413.

[13] D.-K. Mah, "Learning Analytics and Digital Badges: Potential Impact on Student Retention in Higher Education," Technology Knowledge and Learning, 2016, doi: 10.1007/s10758-016-9286-8.

[14] U. Nurdin, "Fuzzy Optimization Method in the Search and Determination of Scholarship Recipients Systems at the University," International Journal of Computer Applications Technology and Research, 2015, doi: 10.7753/ijcatr0407.1005.

[15] J. Board, C. Sutcliffe, and W. T. Ziemba, "Portfolio Selection: Markowitz Mean-Variance Model," 2008, doi: 10.1007/978-0-387-74759-0_513.

[16] S. B. Robbins, J. Allen, A. Casillas, C. H. Peterson, and H. Le, "Unraveling the Differential Effects of Motivational and Skills, Social, and Self-Management Measures From Traditional Predictors of College Outcomes.," Journal of Educational Psychology, 2006, doi: 10.1037/0022-0663.98.3.598.