IJAMEC

INTERNATIONAL JOURNAL OF APPLIED METHODS IN ELECTRONICS AND COMPUTERS

www.ijamec.org

International Open Access

Volume 13
Issue 03

September, 2025

Research Article

Predicting School Dropout Among International Students in Türkiye Using Machine Learning: A Case Study of Yemeni Students

Anas ALHARDI [a,*], ***Selahattin ALAN*** [a]

[a] *Department of Computer Engineering, Selcuk University, Konya, Türkiye*

ARTICLE INFO

Article history:
Received 01 August 2025
Accepted 29 September 2025

Keywords:
Classification,
Data analysis,
International students,
Machine learning,
Prediction,
School dropout,

ABSTRACT

International students often encounter a range of challenges that can put their academic progress at risk, sometimes leading to dropouts. This study focuses on identifying the key factors that influence such outcomes, using Yemeni students as a case group to explore broader patterns among international students. A structured survey was conducted with 583 Yemeni students. After removing incomplete responses, 545 valid cases remained. From these, 268 students who were still enrolled were excluded (because students who were still attending school had neither graduated nor dropped out and were outside our scope of study), leaving 277 complete records (128 graduates and 149 dropouts) for the modeling phase. A total of fifteen supervised machine learning algorithms were applied, with training and evaluation carried out using an 80/20 split. Model performance was assessed through common classification metrics such as accuracy, precision, recall, and F1-score. Several important predictors were identified, including academic performance (GPA), proficiency in Turkish, satisfaction with their academic department, financial stability (e.g., access to scholarships, family income), and levels of psychological stress. Among the tested models, XGBoost performed best, achieving 91% accuracy and an F1-score of 0.92 for the dropout class. To illustrate the practical implications of this research, a prototype web application was also developed. Overall, the study demonstrates that machine learning can be a valuable tool for anticipating dropout risks among international students and highlights the importance of early, targeted support in academic, linguistic, financial, and psychological domains.

1. Introduction

Education serves as a critical foundation for individual development and societal advancement, equipping learners with essential knowledge, skills, and adaptive capacities in an era of rapid technological and social change [1, 2]. School dropouts are deprived of all the benefits that education can provide to the individual. Therefore, we care about school and school attendance [6]. However, educational institutions worldwide face a persistent challenge: student dropout -a complex phenomenon influenced by economic, social, academic, and personal factors [3, 4]. Student dropout rates are a major problem globally. School dropouts not only affect academic achievements but also have serious consequences on social development and economic growth. This problem is especially common among disadvantaged groups such as children from low-income families, ethnic minorities and immigrant students [28]. Globally, the UNESCO Institute for Statistics reported that 258.4 million children and young people were out of school in 2018, representing one-sixth of the world's population in that age group [5]. At the regional level, in Turkey, the early school leaving rate among young people aged 18–24 was 36% in 2017, compared to 9.6% in the European Union in 2022 [7-8].

International students, defined as individuals who travel abroad to pursue education, numbered over 6.4 million globally in 2023 [9, 10]. In Turkey, more than 350,000 students from 198 countries are enrolled in higher education, contributing to academic diversity and the

* Corresponding author. E-mail address: *anas.rajah1@gmail.com*
DOI: 10.58190/ijamec.2025.131

national economy [11]. While their presence enriches higher education, international students face unique academic, socio-cultural, and financial challenges that can increase the risk of dropout. These challenges include language barriers, economic pressures due to limited scholarships, social and cultural isolation, institutional bureaucratic hurdles, and difficulties in adapting to new teaching and assessment methods [12-15].

Among international student populations in Turkey, Yemeni students, numbering 8,198 in 2023 [11], exemplify many of these challenges. Political instability and economic hardship in Yemen, combined with pressures of cultural adaptation, can significantly affect their academic persistence. Common risk factors include insufficient language proficiency, financial strain, limited social integration, and inadequate institutional support. Addressing dropout within this population requires early identification of at-risk students and proactive interventions tailored to their specific circumstances.

Machine learning (ML) offers transformative potential in this context by analyzing complex, multidimensional datasets and identifying subtle patterns of risk. Ensemble methods such as Random Forests and XGBoost, as well as temporal models like LSTMs, have demonstrated high accuracy in predicting academic attrition based on performance, behavioral, and demographic indicators [16-18]. However, current models predominantly focus on domestic or MOOC-based cohorts, with limited application to international students in residential universities, particularly Arab students in Turkey. This gap constrains the ability of institutions to provide timely and targeted support.

This study addresses this critical gap by focusing on Yemeni international students in Turkey. Its objectives are twofold: first, to identify key academic, socio-economic, and psychological factors contributing to school dropout; and second, to develop and evaluate machine learning models capable of accurately predicting dropout likelihood. By integrating predictive analytics with practical interventions, this research aims to enhance early identification of at-risk students, improve retention rates, and support academic success among this vulnerable population.

2. MATERIAL AND METHODS

This study employed a quantitative, applied, and descriptive research design to investigate the factors influencing school dropout among international students in Türkiye, with a specific focus on Yemeni students. The methodology involved data collection through a structured survey, comprehensive data preprocessing, and the application of various machine learning algorithms for predictive modeling.

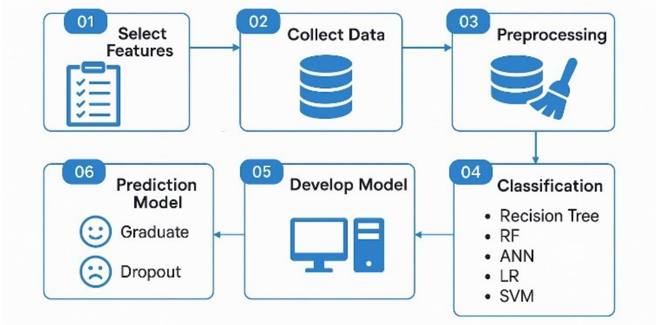

Figure 1. Model Design

This approach aligns with the established frameworks for educational data mining (EDM) and learning analytics, which have been successfully applied to similar educational challenges [19, 20].

2.1. Data Collection

The survey instrument was designed to capture a wide range of variables, including demographic information, academic background, financial conditions, and socio-cultural adaptation factors, resulting in a total of 15 variables used in the analysis. The data collection process involved several phases to ensure comprehensive coverage and participant engagement. Challenges encountered during data collection, such as accessibility and response rates, were addressed to maximize data quality.

2.2. Dataset

The study population comprised Yemeni students enrolled in Turkish higher education institutions. Data were collected from 583 Yemeni students through a structured survey. After removing incomplete or inconsistent responses, 545 valid cases remained. Among these, 268 students were still enrolled at the time of data collection and therefore excluded from the prediction modeling, as their final academic outcomes could not be determined. The remaining 277 students, comprising 128 graduates and 149 dropouts, formed the modeling dataset.

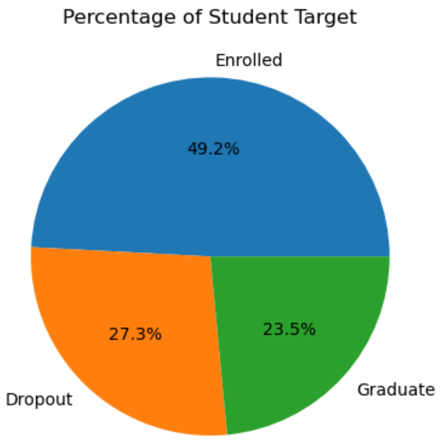

Figure 2. Distribution of Target Variable (Enrolled, Graduate, Dropout)

The distribution of students according to their academic status is as follows:

268 continuing students (49.2%)

128 graduated students (23.5%)

149 dropped out students (27.3%)

2.3. Data Preprocessing

Following data collection, a preprocessing pipeline was implemented. This involved several critical steps:

Data Cleaning: Initial raw data underwent cleaning to handle missing values, outliers, and inconsistencies. This step ensured the reliability and integrity of the dataset. The data cleaning process reduced the initial dataset from 583 responses to the final 545 records used in the analysis, ensuring a higher quality dataset for subsequent modeling. Subsequently, students who were still enrolled (N=268) were excluded from the modeling dataset because their final academic outcomes were unknown. This exclusion resulted in the final modeling dataset of 277 students (128 graduates and 149 dropouts) used for the binary classification task.

Data Translation: Given the diverse linguistic backgrounds of the participants, relevant data points were translated to ensure uniformity and accuracy for analysis.

Data Encoding: Categorical variables were transformed into numerical formats suitable for machine learning algorithms using appropriate encoding techniques.

Data Normalization/Standardization: Numerical features were normalized or standardized to prevent features with larger scales from dominating the learning process, thereby improving model performance.

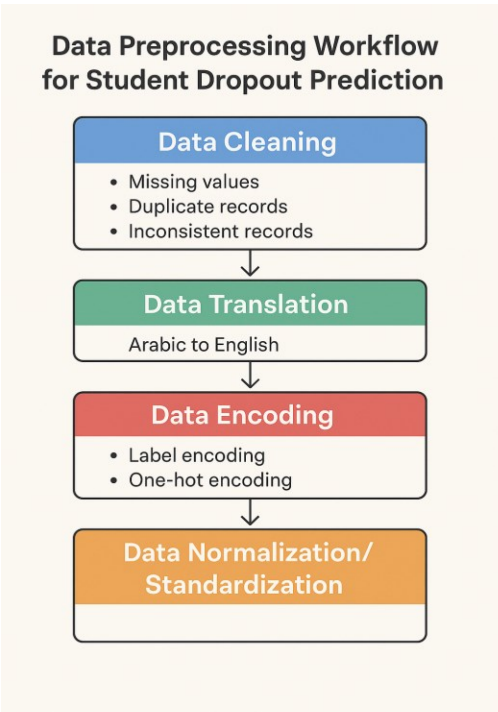

Figure 3. Data Preprocessing Workflow for Student Dropout Prediction

As described in Section 2.2, the final modeling dataset consisted of 277 students (128 graduates and 149 dropouts).

2.4. Prediction Using Machine Learning Algorithms

A total of fifteen supervised machine learning algorithms were employed to predict school dropout. These algorithms were chosen due to their established effectiveness in classification tasks and their applicability to educational data mining. The applied algorithms included, but were not limited to, Logistic Regression, Random Forest, Support Vector Machines (SVM), XGBoost, Decision Tree, Naïve Bayes, LightGBM, Linear Discriminant Analysis (LDA), AdaBoost, Ridge Classifier, Extra Trees, and Quadratic Discriminant Analysis (QDA).

2.5. Model Evaluation and Split

To develop and evaluate the predictive models, the cleaned dataset consisted of 545 valid records. Students who were still enrolled in their studies (N=268) were excluded, as their data would not contribute meaningfully to training the model. The remaining dataset included 277 students, comprising 128 graduates and 149 dropouts. An 80/20 split was employed, allocating 80% of the data (N=221 instances) for model training and reserving the remaining 20% (N=56 instances) as an unseen test set for final model evaluation. This partitioning approach is widely adopted in machine learning applications and has proven effective in educational data mining studies with similar sample sizes [21, 22].

Model performance was assessed using standard classification metrics, defined as follows:

Accuracy: The proportion of correctly classified instances

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision: The proportion of true positive predictions among all positive predictions

Precision = TP / (TP + FP)

Recall: The proportion of true positive predictions among all actual positive instances

Recall = TP / (TP + FN)

F1-score: The harmonic mean of precision and recall, providing a balanced measure of a model's performance

F1 = 2 × (Precision × Recall) / (Precision + Recall)

Here, TP (true positives), TN (true negatives), FP (false positives), and FN (false negatives) represent the outcomes of the confusion matrix.

2.6. Software Used

All data preprocessing, model training, and evaluation were performed using Python programming language, leveraging popular machine learning libraries such as scikit-learn, pandas, and NumPy. XGBoost was implemented using its dedicated library. The prototype web application was developed using standard web development frameworks.

2.7. Web Application for Prediction

A web-based application was developed to deploy the best-performing machine learning model for real-time prediction of student dropout risk. The model, trained and serialized using PyCaret, was integrated with a FastAPI backend, Pydantic for input validation, and an HTML/CSS/JavaScript front end that supports both individual and batch predictions. Deployment was achieved through GitHub (version control), Netlify (front-end hosting), Render (server hosting), and a custom domain. Ethical safeguards were carefully implemented throughout the study: the research protocol received approval from the Selçuk University Ethics Committee; the predictive variables used in the application were designed to exclude any information that could directly identify students (such as name, surname, or student ID); individual student login data and prediction histories were not permanently stored on the web application's server; the predictive model was intended to function as a Decision-Support tool rather than a Decision-Maker; it provided probabilistic rather than definitive outcomes; and any intervention based on the results was required to be carried out by a qualified human expert, ensuring responsible and ethically sound use in student retention strategies.

The operational workflow of the system is structured as follows:

Data Input: Users enter student information through the web interface.

Data Transmission: Input data is sent to the API via an HTTP request.

Model Processing: The trained machine learning model processes the input data.

Prediction Output: The model generates a predicted outcome, indicating either graduation or dropout risk.

Result Display: Predictions are presented on the web interface in an interpretable format for the user.

This end-to-end workflow demonstrates the practical application of predictive modeling in educational contexts, providing a decision-support tool that can assist institutions in early identification of at-risk students and implementation of timely retention interventions.

3. RESULTS

3.1. Exploratory Data Analysis

Prior to model development, a comprehensive Exploratory Data Analysis (EDA) was conducted to gain insights into the dataset characteristics and the relationships between student attributes and academic outcomes. The cleaned dataset comprised 545 valid responses of Yemeni international students in Turkey. For the binary classification task of predicting dropout versus graduation, a refined subset of 277 students with recorded final outcomes was utilized. For the binary classification task of predicting dropout versus graduation, a refined subset of 277 students with recorded final outcomes was utilized. This modeling dataset included 15 predictor variables, encompassing demographic details (e.g., age, gender), academic background (e.g., last GPA), Turkish language proficiency, socio-economic indicators (e.g., scholarship status, family economy, parental education), and self-reported measures of program satisfaction and psychological stress.

The distribution of the target variable within the refined dataset (N=277) revealed 149 instances classified as 'Dropout' (approximately 53.8%) and 128 instances classified as 'Graduate' (approximately 46.2%). This distribution was considered reasonably balanced, mitigating concerns regarding class imbalance potentially skewing model training.

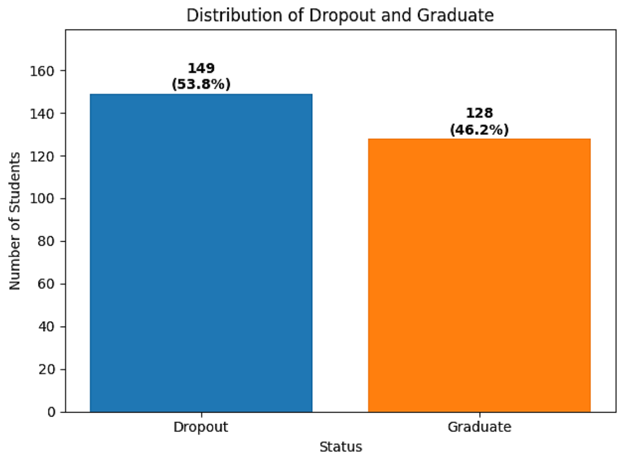

Figure 4. Distribution of Dropout and Graduate.

Further analysis using Kernel Density Estimate (KDE) plots demonstrated a distinct separation in the distribution of the last recorded GPA between the 'Dropout' and 'Graduate' groups. Students who dropped out were predominantly concentrated in the lower GPA ranges (e.g., 0-69), whereas graduates largely occupied the higher GPA ranges (e.g., 70-100).

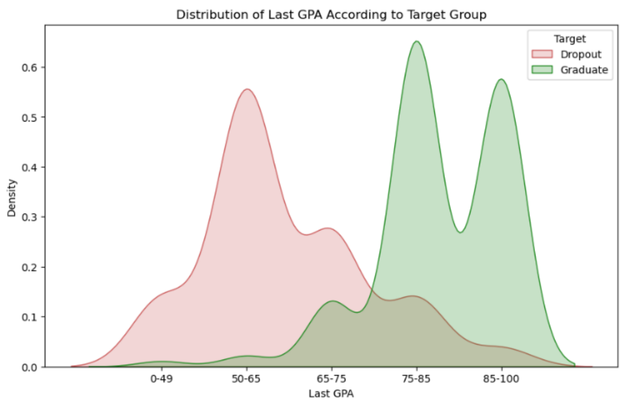

Figure 5. KDE plot showing distribution of Last GPA

This finding strongly suggests that recent academic performance is a critical indicator of persistence.

Similarly, analysis indicated a clear association between Turkish language proficiency ('Weak', 'Good', 'Excellent') and academic outcomes, with 'Weak' proficiency being linked to higher dropout rates.

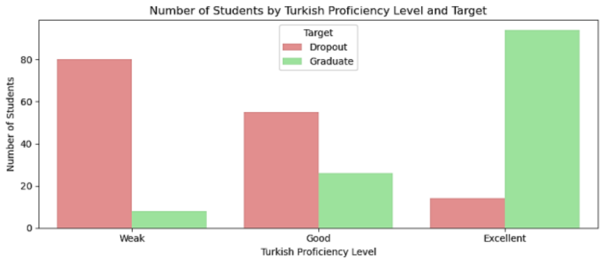

Figure 6. Chart showing distribution of Turkish Proficiency Level

A Cramér's V correlation plot was created to examine the strength and direction of relationships between categorical variables.

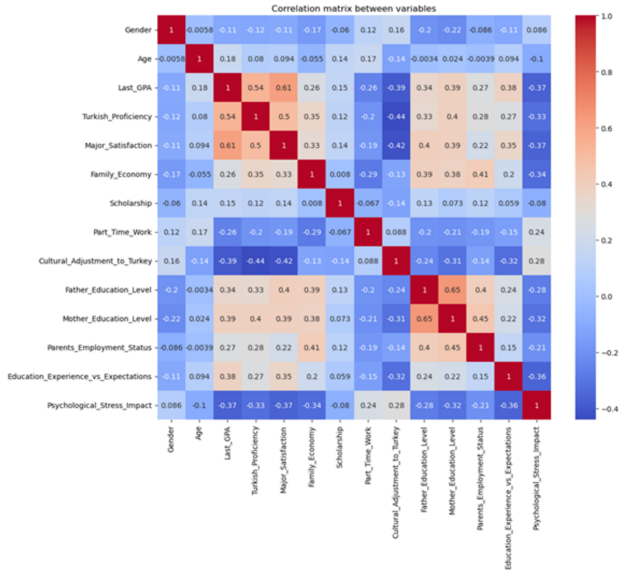

Figure 7. Correlation heatmap of predictor variables

Table 1. Performance Metrics of Machine Learning Models

Model	Accuracy	Dropout Precision	Dropout Recall	Dropout F1-Score	Graduate Precision	Graduate Recall	Graduate F1-Score
XGBoost	0.911	0.97	0.88	0.92	0.85	0.96	0.90
QDA	0.875	0.93	0.85	0.89	0.81	0.91	0.86
Gradient Boosting	0.857	0.90	0.85	0.88	0.80	0.87	0.83
LDA	0.857	0.96	0.79	0.87	0.76	0.96	0.85
Ridge Classifier	0.857	0.96	0.79	0.87	0.76	0.96	0.85
AdaBoost	0.857	0.96	0.79	0.87	0.76	0.96	0.85
LightGBM	0.857	0.93	0.82	0.87	0.78	0.91	0.84
Random Forest	0.839	0.93	0.79	0.85	0.75	0.91	0.82
Decision Tree	0.839	0.90	0.82	0.86	0.77	0.87	0.82
Extra Trees	0.821	0.90	0.79	0.84	0.74	0.87	0.80
Naive Bayes	0.821	0.90	0.79	0.84	0.74	0.87	0.80
Logistic Regression	0.786	0.92	0.70	0.79	0.68	0.91	0.78
Neural Network	0.786	0.89	0.73	0.80	0.69	0.87	0.77
K-NN (k=5)	0.625	0.69	0.67	0.68	0.54	0.57	0.55
SVM	0.518	0.59	0.61	0.60	0.41	0.39	0.40

3.2. Model Implementation and Performance Comparison

Fifteen different classification algorithms were implemented and rigorously evaluated based on their predictive performance on unseen test data, using an 80% training and 20% testing split. The models were assessed using standard evaluation metrics including accuracy, precision, recall, and F1-score, along with confusion matrices to provide a detailed breakdown of classification performance.

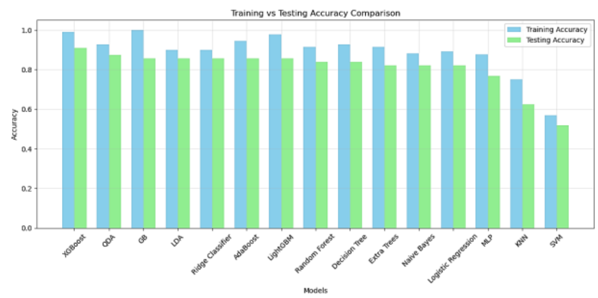

Figure 8. Performance Comparison of Machine Learning Algorithms

The tuned XGBoost model consistently demonstrated the highest predictive accuracy, achieving 91.1% on the hold-out test set (N=56). For the 'Dropout' class, XGBoost exhibited a precision of 0.97, a recall of 0.88, and an F1-Score of 0.92. The confusion matrix for XGBoost indicated its effectiveness, correctly classifying 29 out of 33 actual dropouts and 22 out of 23 actual graduates, with only 5 misclassifications in total.

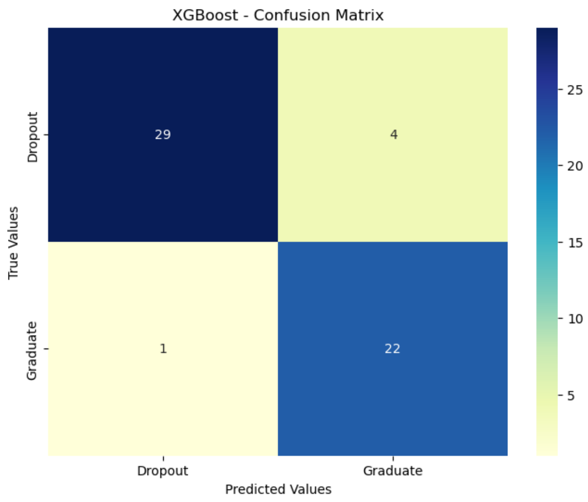

Figure 9. Confusion Matrix for XGBoost Model

Other ensemble methods, such as Quadratic Discriminant Analysis (QDA), also showed strong performance with an accuracy of 87.5%, while Gradient Boosting, AdaBoost, LightGBM, Linear Discriminant Analysis (LDA), and Ridge Classifier all achieved promising results with accuracies exceeding 85%. These findings underscore the superior capacity of ensemble and advanced tree-based methods to capture complex, non-linear relationships within educational data compared to simpler models.

3.3. Feature Importance Analysis

To identify the factors most significantly contributing to the prediction of student dropout, a feature importance analysis was conducted using the mean decrease in impurity (Gini impurity) criterion, as calculated by the Extra Trees algorithm. While XGBoost was the top performer for overall prediction, Extra Trees was specifically chosen for feature importance analysis due to its robustness in feature evaluation, reducing variance in importance estimates and mitigating bias towards high-cardinality features. The analysis revealed a distinct hierarchy among the predictors:

Academic Performance: Consistent with established student retention literature [23, 24], the student's most recent Grade Point Average (Last_GPA) emerged as the single most influential predictor (Importance: 0.138). This highlights the critical role of academic success in the persistence of Yemeni international students in Turkey.

Language Proficiency: Turkish_Proficiency ranked second (Importance: 0.106), underscoring the significant challenge posed by language barriers for international students, a factor widely recognized in previous studies [25, 26].

Program Satisfaction: Satisfaction with the chosen major (Major_Satisfaction) was the third most important feature (Importance: 0.104). This aligns with theories emphasizing academic integration and the fit between student and institution as crucial for retention [27].

Financial Factors: Both Scholarship status (Importance: 0.086) and perceived Family_Economy (Importance: 0.073) ranked within the top five predictors. This emphasizes the substantial impact of financial resources and stability on the ability of these international students to continue their studies, a particularly pertinent issue given the context of Yemen.

Psychological Well-being: The reported impact of Psychological Stress Impact on education ranked sixth (Importance: 0.071), confirming the relevance of mental health and coping mechanisms in navigating the challenges of international study [Smith & Khawaja, 2011].

Table 2. Feature Importance Rankings from Extra Trees Model

Rank	Feature	Importance Score
1	Last_GPA	0.138
2	Turkish_Proficiency	0.106
3	Major_Satisfaction	0.104
4	Scholarship	0.086
5	Family_Economy	0.073
6	Psychological_Stress_Impact	0.071
7	Mother_Education_Level	0.067
8	Cultural_Adjustment_to_Turkey	0.060
9	Part_Time_Work	0.058
10	Parents_Employment_Status	0.057
11	Age	0.051
12	Father_Education_Level	0.049
13	Gender	0.041
14	Education_Experience_vs_Expectations	0.039

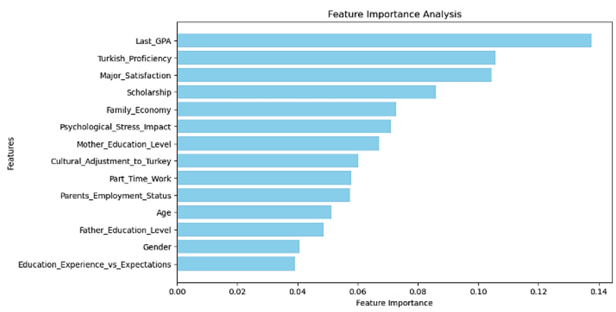

Figure 10. Chart illustrating Feature Importance scores

Other factors such as cultural adjustment, parental background (education and employment), part-time work status, age, and gender exhibited progressively lower importance scores. While still contributing to the model, their predictive power for this specific cohort appeared less dominant compared to academic, linguistic, satisfaction, and financial factors. These results strongly suggest that interventions aimed at improving dropout rates among Yemeni international students in Turkey should prioritize academic support, Turkish language assistance, measures to enhance program satisfaction and integration, and strategies to alleviate financial burdens and psychological stress.

3.4. Model Validation

The robustness and potential applicability of the tuned XGBoost model were further validated through its performance on both aggregate and individual-level unseen data. The strong performance on the hold-out test set provided substantial confidence in the model's ability to generalize its learned patterns to new, similar student profiles within the population from which the sample was drawn. To explore the model's practical utility as a proactive tool, it was applied to a distinct dataset comprising 268 Yemeni international students who were actively enrolled at the time of data collection. This forward-looking application scenario identified 66 students (24.6%) as being at a heightened risk of dropout, while 202 students (75.4%) were predicted to graduate.

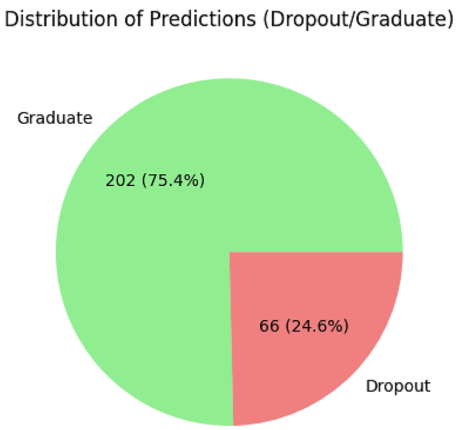

Figure 11. Distribution of Predicted Outcomes for Enrolled Students

This demonstrates the model's potential as an early warning system, capable of guiding targeted interventions such as enhanced academic advising, language support, financial aid counseling, or mental health resources, thereby potentially mitigating risk factors and improving retention rates. It is crucial to interpret these outputs as risk indicators, guiding the allocation of support resources rather than determining student fate.

3.5. Web Application for Prediction

The development of a web application serves as a significant outcome of this research, operationalizing the best-performing XGBoost model into a tangible and accessible tool. This application is designed as a decision-support system for university administrators, academic advisors, and student support staff, enabling real-time assessment of dropout risk for individual students. The user-friendly interface allows for the input of the 14 key predictor variables, and upon submission, the backend processes the data using the trained XGBoost model to generate a prediction.

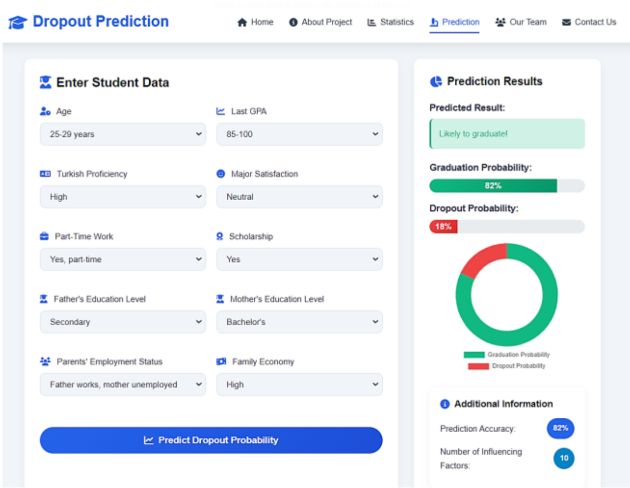

Figure 12. Individual prediction interface – Single student data input and prediction

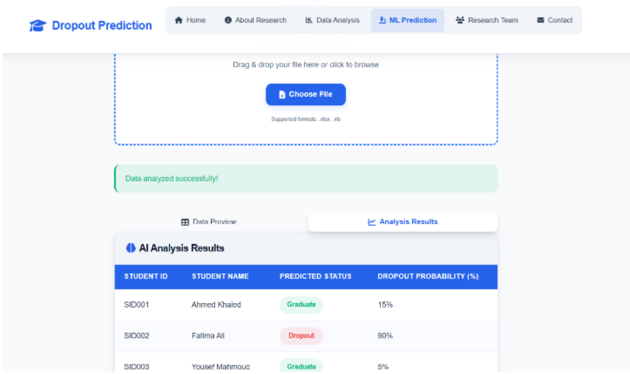

Figure 13. Batch prediction interface – Excel file upload for multiple student predictions

This practical application demonstrates how data-driven insights can be transformed into tools that support proactive and targeted interventions aimed at enhancing international student success and retention.

4. CONCLUSIONS

This study examined the factors influencing dropout among international students, with a particular focus on Yemeni students in Turkish universities and evaluated the potential of machine learning techniques to predict attrition risk. A structured survey was conducted with 545 Yemeni students, collecting data on demographics, academic background, financial conditions, and social adaptation; following rigorous data cleaning and preprocessing, 277 complete responses (128 graduates and 149 dropouts) were used for model development. Fifteen algorithms were tested, with XGBoost achieving the strongest predictive performance. Feature importance analysis revealed that academic performance, Turkish language proficiency, and satisfaction with the chosen major were the most powerful predictors, followed by financial constraints and psychological stress, highlighting the multidimensional drivers of dropout. By validating robust predictive models and operationalizing them into a web-based application, the study provides both empirical insights and a practical early-warning tool that universities can deploy to identify and support at-risk students. These findings underscore the need for institutions and policymakers to enhance academic and linguistic support, expand financial aid, and strengthen social and cultural integration initiatives to foster international student retention. Nonetheless, the study's focus on Yemeni students limits its generalizability; future research should extend to more diverse cohorts, employ longitudinal designs to capture evolving risk factors, and integrate qualitative data and behavioral indicators to enrich understanding. Incorporating advanced temporal modeling techniques could further improve predictive capacity. Overall, this research demonstrates the value of data-driven approaches in informing targeted interventions and shaping policies that create more inclusive and supportive environments for international students. Future studies can extend this work in several ways. For instance, generalizability could be enhanced by applying k-fold cross-validation and testing the models on larger and more diverse populations of international students. In addition, incorporating longitudinal data may help capture dynamic changes in dropout risk over time. The integration of qualitative data (e.g., student interviews) and behavioral indicators (e.g., attendance, learning management system usage) could further enrich the predictive capacity of the models. Finally, employing advanced temporal models such as LSTMs or transformers may provide deeper insights into early dropout prediction.

Declaration of Ethical Standards

This study adhered to ethical principles for human research; all participants provided informed consent and confidentiality was maintained. The protocol was approved by the Selçuk University Faculty of Technology Research Ethics Committee (Decision No. 2025/07, Document No. E-29202782-100-1030873).

Credit Authorship Contribution Statement

Anas ALHARDI: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Visualization.

Selahattin ALAN: Resources, Writing – Review & Editing, Supervision, Project Administration.

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.

Funding / Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

The data that supports the findings of this study are available on request from the corresponding author, upon reasonable request.

References

[1] M. Lövdén, L. Fratiglioni, M. M. Glymour, U. Lindenberger, and E. M. Tucker-Drob, "Education and Cognitive Functioning Across the Life Span,"*Psychological Science in the Public Interest,* vol. 21, no. 1, pp. 6-41, 2020.

[2] E. A. Hanushek and L. Wößmann, "The Role of Education Quality for Economic Growth," *World Bank policy research working paper,* no. 4122, 2007.

[3] D. R. Entwisle, K. L. Alexander, and L. S. Olson, "Temporary as Compared to Permanent High School Dropout," *Social forces,* vol. 82, no. 3, pp. 1181-1205, 2004.

[4] J. J. McWhirter, B. T. McWhirter, A. M. McWhirter, and E. H. McWhirter, *At-risk Youth: A Comprehensive Response*. Thomson Brooks/Cole Publishing Co, 1993.

[5] U. I. f. S. UIS. "New Methodology Shows that 258 Million Children, Adolescents and Youth Are Out of School." UNESCO Institute for Statistics. https://uis.unesco.org/sites/default/files/documents/new-methodology-shows-258-million-children-adolescents-and-youth-are-out-school.pdf (accessed July 3, 2025).

[6] M. Aydoğmuş, H. Serçe, and M. Çalışkan, New Trends and Promising Directions in Modern Education School Dropout. May 2022, ISBN: 978-625-8414-19-6.

[7] S. E. Eurostat, "Early Leavers from Education and Training," *Analysis by Degree of Urbanisation,* 2018.

[8] S. Kartal and F. E. Ballı, "Okul Terki Araştırmaları: Sistematik Bir Analiz Çalışması," *Kırıkkale Üniversitesi Sosyal Bilimler Dergisi,* vol. 10, no. 2, pp. 257-278, 2020.

[9] R. A. Smith and N. G. Khawaja, "A Review of the Acculturation Experiences of International Students," *International Journal of Intercultural Relations,* vol. 35, no. 6, pp. 699-713, 2011.

[10] W. M. R. IOM. "International Students." https://worldmigrationreport.iom.int/what-we-do/world-migration-report-2024-chapter-2/international-students (accessed 15 February 2025.

[11] YÖK, "Yükseköğretim Bilgi Yönetim Sistemi - Uyruğa Göre Öğrenci Sayıları Raporu," 24 May 2024. [Online]. Available: https://istatistik.yok.gov.tr

[12] G. Kurum and O. Erdemli, "Challenges Encountered by Academicians in International Students' Education: Turkey Case," *Eurasian Journal of Educational Research,* vol. 91, pp. 145-172, 2021.

[13] J. Jung and Y. Kim, "Exploring Regional and Institutional Factors of International Students' Dropout: The South Korea Case," *Higher Education Quarterly,* vol. 72, no. 2, pp. 141-159, 2018.

[14] K. Eynullayeva, M. Gökalp, and B. Y. Hatunoglu, "Investigation of the Turkish Cultural Adaptation of International Students Living in Turkey," *European Educational Researcher,* vol. 4, no. 2, pp. 155-169, 2021.

[15] M. Özoğlu, B. S. Gür, and İ. Coşkun, "Factors Influencing International Students' Choice to Study in Turkey and Challenges They Experience in Turkey," *Research in Comparative and International Education,* vol. 10, no. 2, pp. 223-237, 2015.

[16] S. Lee and J. Y. Chung, "The Machine Learning-Based Dropout Early Warning System for Improving the Performance of Dropout Prediction," *Applied Sciences,* vol. 9, no. 15, p. 3093, 2019.

[17] M. Fei and D.-Y. Yeung, "Temporal Models for Predicting Student Dropout in Massive Open Online Courses," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, 2015: IEEE, pp. 256-263.

[18] W. Romsaiyud *et al.*, "Predictive Modeling of Student Dropout Using Intuitionistic Fuzzy Sets and XGBoost in Open University," in *Proceedings of the 2024 7th International Conference on Machine Learning and Machine Intelligence (MLMI)*, 2024, pp. 104-110.

[19] R. S. Baker, T. Martin, and L. M. Rossi, "Educational Data Mining and Learning Analytics," *The Wiley Handbook of Cognition and Assessment: Frameworks, Methodologies, and Applications,* pp. 379-396, 2016.

[20] C. Romero and S. Ventura, "Educational Data Mining and Learning Analytics: An Updated Survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 10, no. 3, p. e1355, 2020.

[21] S. Cho, T. J. Jun, B. Oh, and D. Kim, "Dapas: Denoising Autoencoder to Prevent Adversarial Attack in Semantic Segmentation," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020: IEEE, pp. 1-8.

[22] C. Márquez-Vera, A. Cano, C. Romero, and S. Ventura, "Predicting Student Failure at School Using Genetic Programming and Different Data Mining Approaches with High Dimensional and Imbalanced Data," *Applied intelligence,* vol. 38, pp. 315-330, 2013.

[23] J. P. Bean, "Dropouts and Turnover: The Synthesis and Test of a Causal Model of Student Attrition," *Research in higher education,* vol. 12, pp. 155-187, 1980.

[24] V. Tinto, "Dropout from Higher Education: A Theoretical Synthesis of Recent Research," *Review of Educational Research,* vol. 45, no. 1, pp. 89-125, 1975.

[25] M. S. Andrade, "International Students in English-Speaking Universities: Adjustment Factors," *Journal of Research in International education,* vol. 5, no. 2, pp. 131-154, 2006.

[26] M. Sherry, P. Thomas, and W. H. Chui, "International Students: A Vulnerable Student Population," *Higher Education,* vol. 60, pp. 33-46, 2010.

[27] V. Tinto, *Leaving College: Rethinking the Causes and Cures of Student Attrition*. University of Chicago Press, 2012.

[28] S. A. Sulak and N. Koklu, "Predicting Student Dropout Using Machine Learning Algorithms," *Intelligent Methods in Engineering Sciences*, vol. 3, no. 3, pp. 91–98, 2024.