Journal of Basics and Applied Sciences Research (JOBASR)
ISSN (print): 3026-9091, ISSN (online): 1597-9962
Volume 4(3) May 2026
DOI: https://dx.doi.org/10.4314/jobasr.v4i3.7

Academic performance and Graduation time Predictions for Students from Vulnerable Insurgency Zone using Machine Learning Algorithms

Awuza Abdulrashid Egwa[1*], Emmanuel Baldwin Mbaya[2], Sikirat Kehinde Aina[3], Abubakar Karabade[4] & Abdullahi Musa Bello[5]
[1,2,3,4,5]Department of Computer Science, Federal University Gashua, Yobe State, Nigeria
[*]Corresponding Author Email: awuza@fugashua.edu.ng

ABSTRACT
The timeliness of graduation is a critical performance indicator for the quality and effectiveness of higher education systems. Students enrolled into universities aiming to lay the foundation in pursuance of a decent career. Although large numbers of students enrol in universities each year, many drops out or fail to graduate within stipulated period. Understanding student's graduation pattern can better assist universities lecturers and managers to serve student populations better in achieving their learning objectives. In the North-Eastern Nigerian context, this issue extends beyond institutional efficiency and represents a socio-technical challenge. Insurgency-related disruptions introduce unique stressors**,** necessitating data-driven monitoring and intervention strategies. An appropriate step is establishing a mechanism capable of monitoring and predicting graduation timelines correctly. Therefore, this study evaluates the academic performance and graduation likelihood of 1266 undergraduate students from insurgency-prone areas (IPA) at Federal University Gashua. By applying Random Forest, Logistic Regression, SVM, and ANN across imbalanced, SMOTE, and ADASYN-balanced datasets, the analysis shows that Random Forest (RF) consistently outperformed all other models with 91% precision and 97% AUC for imbalance data, 92% precision and 98% AUC for SMOTE-balance data, and 92% precision and 99% AUC for ADASYN-balance data. The results further showed that CGPA is the strongest predictor of timely graduation, while IPA status is the most influential factor associated with the likelihood of delayed graduation. Applying prediction models to timely graduation have diverse implications for both students and management thereby; harnessing preventive solution to those with likelihood of not graduating on time is of the essence.

Keywords:
Student performance,
Graduation time,
Machine learning,
Insurgents,
Insurgency-affected Regions

INTRODUCTION

University education plays an essential role in enhancing a nation's human resource capacity. While improving their abilities and knowledge, it also ensured stability in economic growth of a country. University's main goal is to provide quality education to its students (Tampakas et al., 2019). Students enrolled into universities hoping to lay the foundation or to learn specialised skills in pursuance of a decent career. Although large numbers of students enrol in universities each year, many drops out or fail to graduate within the stipulated period (Orion et al., 2014). Understanding student's graduation pattern can better assist universities lecturers and managers to serve student populations better in achieving their learning objectives (Aiken et al., 2020). Universities should develop robust mechanisms for monitoring students' progression and identifying key determinants of academic performance for educational quality assurance (Tampakas et al., 2019) and timely graduation.

Graduation on time (GOT) as a concept is when students finish or complete their studies within the stipulated time frame or better still, a state where students accomplished their studies within courses duration set by the university. This student's graduation time is of uttermost important for universities locally here in Nigeria and worldwide. This is because a good yearly rate of graduation is key matrix for ranking university position in the education industry and also influences significantly in its annual operation costs,

How to cite this article: Awuza A. E., Emmanuel B. M., Sikirat K. A., Abubakar K. & Abdullahi M. B. (2026). Academic performance and Graduation time Predictions for Students from Vulnerable Insurgency Zone using Machine Learning Algorithms. *Journal of Basics and Applied Sciences Research*, *4*(3), 66-73. https://dx.doi.org/10.4314/jobasr.v4i3.7

(Mohammad Suhaimi et al., 2019; Suhaimi et al., 2019). In Nigeria, the national University commission (NUC) in collaboration with universities, are mapping out initiative strategizing towards ensuring timely graduation of both undergraduates and postgraduates' student within their stipulated program time. These problems are now more worrisome mostly for some part of the country (North-east specifically). Even though the core factors for this predicament may vary, sometimes graduation delay is unavoidable, but a careful monitory of the situation might yield beneficial outcomes (Mohammad Suhaimi et al., 2019). Unfortunately, universities and others higher institutions of learning accumulate large volumes of data about their students' attributes and performances over the years. Essential information can be extracted from these piled of historical data to gain more insights for better decision making. Although, Machine learning has achieved a human-level performance in many domains, including image classification, speech recognition and machine translation (Munkhdalai et al., 2019). However, in educational domain, machine learning models for predicting student's performances and graduation time is recently gaining popular attention from researchers across the globe with little research made here in Nigeria. A considerable body of research has examined the prediction of students' academic outcomes, particularly their time to graduation. Within the field of Educational Data Mining (EDM), this topic has remained a persistent challenge, and classification-based approaches are among the most widely employed methods for addressing it (Thakar et al., 2015). In another work, Mohammad Suhaimi et al. (2019) built a predictive model for determining undergraduate student's graduation status using five ML algorithms; DT, RF, NB, and two variants of SVM. The SVM (PolyKernel) demonstrated the highest predictive accuracy based on multiple evaluation metrics. Hadi & Muhamad (2019) identified factors that shape postgraduate research performance through exploratory factor analysis and structural equation modelling applied to 112 valid responses. Positive associations between student performance to institutional, personal, and supervisor related factors were established, with personal factors exerting the strongest influence on graduation.

A key factor militating against performance and graduation time of students where researchers have not pay attention to is the insurgency which have be-devilled the nation education system in the past few years. The Boko Haram insurgence, primarily culminating unbridled attacks on educational facilities, students and their tutors-alike rendering thousands of citizens homeless in their determination to eradicate 'western education', (Olakulehin & Ojo, 2010). This has cause apprehension and fear in the minds of students residing in those regions studying in various institutions of learning in the country. Especially institutions domiciled within the region (north-east).

Academic activities are disrupted intermittently across primary, post primary schools arising from incessant attacks on educational facilities. This phenomenon compelled government to shut down schools in order to protect the students and their teachers. The Boko Haram attacks also culminate in poor student's performance because learning is characterized by threat in the school environment of the north-east, whereas it is an accomplished fact that learning thrives mostly in an environment devoid of threat. Etebu et al. (2011) asserts that "any society characterized by any form of violence will not be conducive for any social interaction in form of teaching and learning".

Although predicting student performance and graduation time has been extensively studied in literature, little to non-research has been conducted on data from Nigerian Universities focusing on students from vulnerable insurgent's area. Due to some peculiarities, models build for use in other countries may not effectively apply well in another country (Bako et al., 2023)

Therefore, there is a need for data-driven monitoring and predictive systems to identify students at risk of delayed graduation and enable timely interventions. This study proposes a machine learning-based model to track and predict the academic performance and graduation time of students from insurgency-affected areas at Federal University Gashua, comparing them with students from non-insurgency regions. For effective model performance comparison, Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM) and Artificial Neural Network (ANN) are employed. They are employed because Complex, nonlinear patterns that conventional models frequently miss can be captured by machine learning (ML) techniques ranging from Linear Regression to sophisticated models like SVM, RF, and ANN (B. I. et al., 2025; Sada et al., 2025).

This paper is organized as follows: Section II presents the Methodology, Results and Discussion is presented in Section III, recommendations in section IV and Conclusion in Section V.

MATERIALS AND METHODS

Data Collection

Demographic and academic records of 1266 undergraduate students studying in Federal University Gashua (FUGA) were collected covering 2015/2016 to 2021/2022 academic sessions. This dataset comprises a wide range of information such as CGPA, age, sex, nationality, state of origin, settlement, sponsorship, marital status, year of entry, year of graduation, SSCE grades, school absenteeism, tutorial attendance, employment status, mode of entry, family income, insurgency-prone area (IPA), and the target variable. Some communities within the state are mark as insurgent vulnerable zone where students from such communities are the primary target of this research work. They include Dapchi, Gaidam, Yunusari, Baimari, Gujba, Gulani, Potiskum, Sabon gida, Baimari, Bun'aji, Damaturu, Bama, Maiduguri, Chibok, among others as shown in Table 1.

Table 1: Dataset Description

S/N	VARIABLE	DESCRIPTION	TYPE
1	Sex	Male = 1, Female = 0	Categorical (Numeric)
2	Age	Age of students	Continuous (Numeric)
3	Mar_St	Marital Status: 1=Married, 0= Single	Categorical (Numeric)
4	Trb	Tribe of the students	Categorical (String)
5	DisA	Disability: Yes =1, No = 0	Categorical (Numeric)
6	Loc.	Locality	Continuous (String)
7	Setlm	Settlement: 1 = Urban, 0 = Rural	Categorical (Numeric)
8	IPA	Insurgent Prone Area: Yes = , No = 0	Categorical (Numeric)
9	Fam_Inc	Family Incomes: 1= High, 0= Low	Categorical (Numeric)
10	Spons.	Sponsorship: 1 = Sponsored, 0 = Self	Categorical (Numeric)
11	O'L_Grade	O'Level Grade: 1 = High, 0 = Low	Categorical (Numeric)
12	MOE	Mode of Entry: 1= UTME, 2 = Remedial, 3 = DE	Nominal (Numeric)
13	Entry_S	Entry Score: 1 = High, 0 = Low	Categorical (Numeric)
14	Lect_A	Lecture Attendance: 1 = High, 0 = Low	Categorical (Numeric)
15	Tut	Tutorial: 1 = Yes, 0 = No	Categorical (Numeric)
16	CGPA	Cumulative Grade point Average	Continuous (Numeric)
17	Class	Gr = graduate, Ng = Not graduate	Target (Numeric)

Data Preparation

Pre-processing is a critical stage in preparing data for machine learning applications. In this study, the dataset was examined for quality, consistency, missing values, and outliers**.** Since the data originated from a single source (the Department of Computer Science) issues such as inconsistent data types or mixed categorical values (e.g., "male" versus "man") were minimal. Missing values were handled using appropriate imputation techniques, and outliers were carefully assessed. To ensure inconspicuousness, every particular identifiable variable was removed leaving only relevant ones required for the analysis.

Feature Selection (FS)

Feature selection improves model efficiency and interpretability by reducing the dataset dimension, burst model training, minimizes model intricacy, improves predictive rating and reduces risk of over-fitting. FS is typically applied independently of specific machine learning algorithms. Ahmed et al. (2017) showed FS as an act of removing irrelevant features and retaining ones that contribute meaningful to prediction outcome. In this study, Recursive Feature Elimination with Cross-Validation (RFECV) approach was used for FS as in (Chandramohan, 2022).

Data Resampling

Dataset class imbalance is often a common challenge faced in ML predictions athwart various domains, which often leads to biased model predictions (Brennan, 2019). When models are built with an imbalanced dataset, classifiers algorithms incline to favour the majority class, thereby ignoring minority cases. Although many previous studies have applied the Synthetic Minority Oversampling Technique (SMOTE) for addressing imbalance (Ünal et al., 2019), SMOTE can be less effective with high dimensional datasets. For this reason, both SMOTE and ADASYN were utilized in this study. ADASYN an extension of SMOTE, generate synthetic samples in regions of the minority class with difficulties to learn pattern. It uses a density-based criterion to determine how many synthetic samples should be spawned for each minority instance, thereby accentuating harder-to-learn cases. This approach contrasts with SMOTE, which creates the same number of synthetic samples for each minority data point. Using both methods enable a comparative evaluation of their impact on model performance and helps identify the most effective approach for balancing the dataset.

Data Splitting

The entire dataset was divided into training and test sets at a roughly 2:1 ratio, and the models were validated using 5-fold cross validation (5-CV) on the training set. One slice was reserved for testing for each of the five folds, and the other four slices were utilized to train the model.

F. Models Parameter Settings
In this case, the combination of hyper-parameters was optimized using grid search. We supply the search space for these algorithms as follows since grid search is an exhaustive search of a pre-defined hyper-parameter space: Tree_num in RF models was searched between 100 and 1000, min_samples_leaf between 3 and 10, and max_depth between 5 and 25; log2c and log2g in SVM models were searched between −10 and 10 using the RBF kernel; hidden_nodes in ANN models were searched between 10 and 500 and log(alpha) between (8,8,8). Other parameters not mentioned were set as default. Python software was used for implementing the models.
G. Models Evaluation Matrices
A series of performance measures were employed, including Precision, Recall, F-measure rates and AUC values for ROC.

Precision
Precision is to measure the proportion of truly positive samples in samples that are predicted as positive samples by the classification model.

$$Precision = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)+False Positive (FP)}} \quad \text{i}$$

Recall
The recall is calculated by dividing the number of successfully predicted positive products by the total number of correctly predicted positive goods and wrongly anticipated negative words.

$$Recall = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)+False Negative}} \quad \text{ii}$$

F1-Score
It is possible to interpret the Precision and Recall requirements jointly as opposed to separately. In order to achieve this, we took into account the F-Measure values provided by the Precision and Recall harmonic mean, since the harmonic mean yields the average of two distinct factors generated per unit.

$$F1\ Score = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision+Recall}} \quad \text{iii}$$

RESULTS AND DISCUSSION

The performance evaluation of four machine learning algorithms, namely RF, ANN, SVM, and LR, was conducted on three types of datasets: the imbalanced dataset, and balanced datasets generated using SMOTE and ADASYN. The performance evaluation metrics used include Precision, Recall, F-Score, and ROC-AUC**.** The complete results are presented in Table 2.

Table 2: Models Performance Results across all Data categories

Data_Class	**Algorithms**	**Precision**	**Recall**	**F_Score**	**ROC_AUC**
Imbalanced	RF	0.9058	0.9057	0.9054	0.9812
Imbalanced	ANN	0.8150	0.8113	0.8120	0.9333
Imbalanced	SVM	0.8678	0.8679	0.8675	0.9667
Imbalanced	LR	0.8901	0.8868	0.8872	0.9710
SMOTE	RF	0.9245	0.9245	0.9245	0.9855
SMOTE	ANN	0.8315	0.8302	0.8306	0.9116
SMOTE	SVM	0.8901	0.8868	0.8872	0.9638
SMOTE	LR	0.9123	0.9057	0.9061	0.9652
ADASYN	RF	0.9245	0.9245	0.9245	0.9841
ADASYN	ANN	0.8691	0.8679	0.8682	0.9435
ADASYN	SVM	0.8691	0.8679	0.8682	0.9667
ADASYN	LR	0.9123	0.9057	0.9061	0.9652

Performance Evaluation on the Imbalanced Dataset
In all models, the Random Forest Classifier was found to produce the best results on the imbalanced dataset with a precision of 0.9058, recall of 0.9057, F-score of 0.9054, and ROC-AUC of 0.9812. This result implies that the RF Classifier inherently deals with class imbalance problems to a considerable extent. This may be because the model is an ensemble model.
On the other hand, the SVM model was found to produce moderately high results on the imbalanced dataset with an F-score of 0.8675 and a ROC-AUC of 0.9667. These results are better than those produced by the ANN and LR models in terms of recall but are still low compared to the RF Classifier.
The LR model was found to produce an F-score of 0.8872 and a ROC-AUC of 0.9710 on the imbalanced dataset. These results are better than those produced by the ANN model but are still low compared to the SVM and RF Classifier models.
The ANN model was found to produce the worst results on the imbalanced dataset with an F-score of 0.8120 and a ROC-AUC of 0.9333. Therefore, the results on the imbalanced dataset show a clear ranking of the models

from the best performing model to the worst model. The ranking is RF Classifier > SVM > LR > ANN.

Performance after SMOTE Oversampling

Applying SMOTE resulted in overall performance improvement**.** RF again performed best, with Precision, Recall, and F-Score increasing to 0.9245, while the ROC-AUC curve also increased to 0.9855. This improvement may be attributed to the fact that Random Forest supports synthetic oversampling.

Likewise, LR significantly improved using SMOTE, with an F-Score of 0.9061 and an ROC-AUC curve of 0.9652, surpassing its previous performance on the unbalanced dataset. Conversely, SVM did not show significant improvement, with an F-Score of 0.8872 and a slightly reduced ROC-AUC curve to 0.9638.

On the other hand, ANN showed marginal improvement, with an F-Score of 0.8306 while the ROC-AUC curve reduced from 0.9333 to 0.9116. This suggests that SMOTE did not significantly improve the separability of the decision boundary of ANN.

Therefore, using SMOTE, the performance of the algorithms improves significantly, with RF remaining the best, followed by LR, then SVM, while ANN remains the worst.

C. Performance after ADASYN Oversampling

Exploring ADASYN for class imbalance, RF again has the finest performance, attaining stout consistent values across all metrics with F-Score of 0.9245 and ROC-AUC of 0.9841. This indicates robustness to various resampling approaches.

Captivatingly, both ANN and SVM achieved identical values; Precision of 0.8691, Recall of 0.8679 and F-Score of 0.8682. ADASYN enhanced ANN more significantly compared to SMOTE, with increases in both F-Score and ROC-AUC of 0.9435. Likewise, SVM shows boosted F-Score performance paralleled to the imbalanced dataset but had unaffected ROC-AUC of 0.9667.

LR upheld stable performance under ADASYN, reflecting its SMOTE results to 0.9061F-Score and 0.9652 ROC-AUC values. This shows steady behaviour across resampling methods.

Overall, ADASYN improved ANN and SVM more than SMOTE did while upholding high performance for Random Forest and Logistic Regression.

Cross-Method Comparison

Crosswise all three dataset variants, RF steadily outperformed all other algorithms, showing the highest Precision, Recall, F-Score, and ROC-AUC. This highlights Random Forest's robustness to class imbalance and its ability to influence synthetic samples effectively.

LR exhibited stable and reliable performance, benefiting moderately from resampling. SVM demonstrated improved balanced performance with ADASYN, while ANN remained the most sensitive to dataset imbalance and benefitted the most from ADASYN relative to SMOTE. These results are in consistent with the work of (Ahmed et al., 2017; Aiken et al., 2020; Bako et al., 2023; Ünal et al., 2019)

Summary of findings

1. Random Forest model is the top performing model across all metrics and all dataset variants.
2. SMOTE and ADASYN both improved model performance, but ADASYN offered better gains for ANN and SVM.
3. ANN is most affected by class imbalance, showing the lowest scores before resampling.
4. Logistic Regression shows stable, predictable improvements with resampling.
5. ADASYN provides more adaptive minority sample generation, reflected in the bigger performance boost for ANN and SVM.

Features Important Ranking

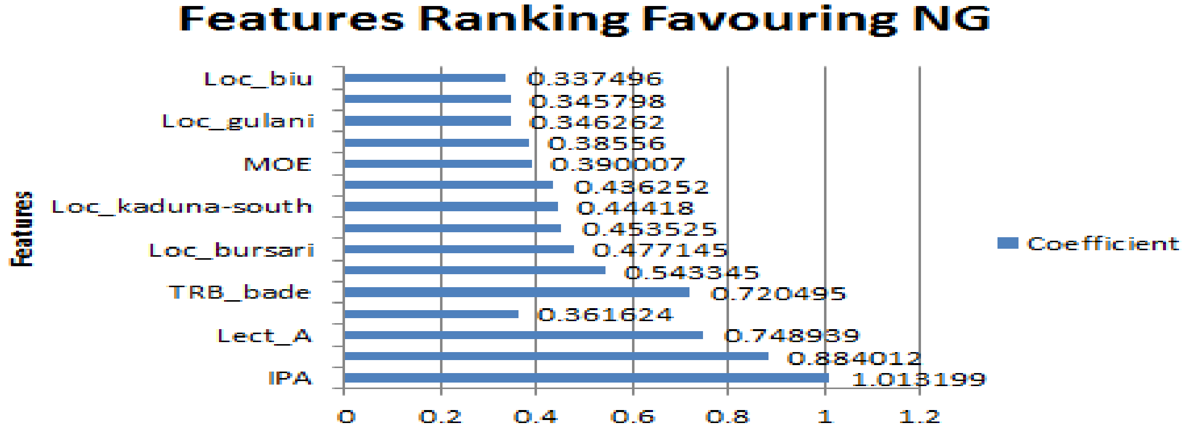

Figure 1: Features Ranking Favouring Ng

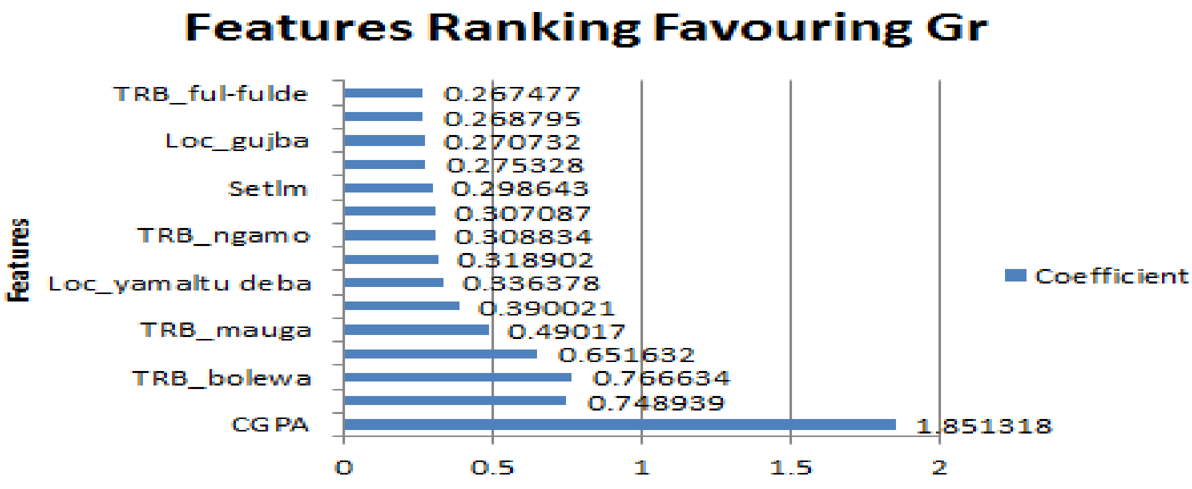

Figure 2: Features Ranking Favouring Gr

One of the objective of this research work is to determine graduation pattern (Graduate Gr; Not graduate Ng) of students from Insurgents Prone Area (IPA) studying in Fuga. Using RF and LR being the best performing algorithms, we ranked coefficients of features contributing to outcomes favouring the target variables (Gr and Ng), allowing a comparative assessment of the relative influence of each variable as shown in Figure 1 and Figure 2. The coefficients reflect the magnitude and direction of each feature's contribution, with larger positive values indicating stronger predictive importance.

As can be seen in Figure 1 above, the strongest influencer among the predictors is clearly IPA, whose coefficient is approximately 1.01. This suggests the strongest positive relationship with the Ng-outcome. The next strongest are Lect_A at approximately 0.88 and TRB_bade at approximately 0.72. These features are also positive influencers of the prediction and are likely significant features in the model.

The middle tier of influencers comprises Loc_bursari at approximately 0.54, Loc_kaduna-south at approximately 0.44, and MOE at approximately 0.39. These features are influencers of the prediction in a steadier and less dominant manner than the top influencers. They appear to be playing a supporting role in influencing the Ng-favouring outcomes.

The lower-tier features, such as Loc_biu, Loc_gulani, and Loc_bursari, have coefficients that fall in the range of 0.33 to 0.36, indicating a weaker but positive influence in the outcome of the model's prediction results.

In conclusion, the ranking of the features, as depicted in Figure A, indicates that the features have a stratified influence in the outcome of the model's results, with a small number of features dominating the influence, while others make a more modest contribution to the results.

Consequently, Figure 2 indicates the homogeneous coefficients derived from the predictive model that seeks to analyze the factors that most influence the outcome variable, i.e., "Gr." This figure indicates the ranking of the features in ascending order, thereby providing a clearer visualization of the contribution of each feature to the outcome results. Among the features, it is evident that the feature, i.e., CGPA, is the most influential, with a significantly higher coefficient of 1.851318 compared to the other predictors in the model, indicating a positive correlation between the feature and the outcome results, i.e., the graduation results for the student, thereby indicating that the most important feature in the outcome results is the student's performance in school, i.e., CGPA.

The next set of salient predictors includes TRB_bolewa, which has a coefficient of 0.748939 and TRB_mauga, which has a coefficient of 0.49017. The values indicate a moderate positive effect. It is evident that the tribal/locational characteristics of the categories have a positive effect, although this effect is considerably lower compared to the effect of CGPA. Moreover, Loc_vamatlu deba and TRB_ngamo have coefficients of 0.386378 and 0.308834, respectively. The values indicate a weak positive effect.

At the lower end of the spectrum, TRB_ful-fulde has a coefficient of 0.267477, whereas Loc_gujba and Setlm have coefficients of 0.268795 and 0.278528, respectively. The low coefficient values indicate that although the features have a positive effect, the effect is considerably low compared to the effect of CGPA.

In conclusion, the feature ranking as depicted in Figure B indicates a clear hierarchy in which academic performance takes precedence over demographic/territorial factors. This evidence confirms that academic performance remains a vital factor in the predictive and classification framework for similar groups.

CONCLUSION

Understanding student's graduation pattern can better assist universities lecturers and managers to serve student populations better in achieving their learning objectives. Develop robust mechanisms for monitoring students' progression and identifying graduation and non-graduation determinants is key for educational quality assurance and timely graduation.
This study uses academic records of 1266 undergraduate students of the department of computer science ranging from 2015/2016 to 2021/2022 academic sessions to evaluated the academic performance and graduation likelihood of undergraduate students from insurgency-prone areas at Federal University Gashua using machine learning techniques. By applying RF, LR, SVM, and ANN across imbalanced data, SMOTE-balanced and ADASYN balanced datasets, the analysis established that RF steadily outperformed all other models in precision, recall, F-score, and ROC-AUC. The results further indicate that CGPA is the strongest predictor of timely graduation, while insurgent-prone area status (IPA) is the most influential factor associated with the likelihood of not graduating on time. Other academic engagement factors, such as lecture attendance, tutorial participation, and entry scores, also contributed significantly to model predictions; however, demographic factors exhibited weaker influences. The results shed more light on the complex interrelation between academic performance indicators and socio-environmental challenges faced by students from regions that are prone to violence. In addition, it is important to note that the research proves that it is possible to apply machine learning tools to identify such students at an early stage.

This work will impact curriculum design for programs and education policy design in general. Serving as Early-Warning Monitoring System for universities, Targeted Academic Support tools for monitory Students from insurgency-prone areas, Strengthening Engagement Factors, Psychosocial and Security Related Support and Data Driven Decision Making for universities. Future work should include collecting more dataset from all department of the university for a generalized result interpretation. Also, extending the performance prediction to elective courses and using the prediction results to recommend courses to students.

ACKNOWLEDGEMENT

This work was supported by federal university Gashua through its Institutional Based Research grand (IBR). The authors would like to thanks the Institution for the opportunity and the department of computer science for given us the data used for this research.

CONFLICT OF INTEREST

The authors declare that no conflict of interests that could have impacted the study exist.

REFERENCE

Ahmed, S., Mahbub, A., Rayhan, F., Jani, R., Shatabda, S., & Farid, D. Md. (2017). Hybrid Methods for Class Imbalance Learning Employing Bagging with Sampling Techniques. *2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, 1–5. https://doi.org/10.1109/CSITSS.2017.8447799

Aiken, J. M., De Bin, R., Hjorth-Jensen, M., & Caballero, M. D. (2020). Predicting time to graduation at a large enrollment American university. *PLOS ONE*, *15*(11), e0242334. https://doi.org/10.1371/journal.pone.0242334

B. I., M., E. S., C., & M. I., U. (2025). Comparative analysis of machine learning algorithms for stock price prediction. *Journal of Basics and Applied Sciences Research*, *3*(6), 243–251. https://doi.org/10.4314/jobasr.v3i6.24

Bako, H. S., Ambursa, F. U., Galadanci, B. S., & Garba, M. (2023). Predicting timely graduation of postgraduate students using random forests ensemble method. *Fudma journal of sciences*, *7*(3), 177–185. https://doi.org/10.33003/fjs-2023-0703-1773

Brennan, J. (2019, December 10). *Dealing with Imbalanced Data*. Digital Catapult. https://medium.com/digital-catapult/ dealing-with-imbalanced-data-8b21e6deb6cd

Chandramohan, R. M. (2022). *Stroke Detection and Prediction Using Deep Learning Techniques and Machine Learning Algorithms*.

Etebu, C. E., Buseni, J. A., & Amamieyenimighan, C. (2011). Hostage taking in the Niger delta: Implications on Educational Development in Bayelsa state of Nigeria. In *Psychology and Anthropology in Practice* (Vol. 3, Number 1).

Hadi, N. U., & Muhamad, B. (2019). Factors Influencing Postgraduate Students' Performance: A high order top down structural equation modelling approach. *Educational Sciences: Theory & Practice, 19*. https://doi.org/10.12738/estp.2019.2.004

Mohammad Suhaimi, N., Abdul-Rahman, S., Mutalib, S., Abdul Hamid, N. H., & Md Ab Malik, A. (2019). Predictive Model of Graduate-On-Time Using Machine Learning Algorithms. In *Soft Computing in Data Science* (pp. 130–141). https://doi.org/10.1007/978-981-15-0399-3_11

Munkhdalai, L., Munkhdalai, T., Namsrai, O.-E., Lee, J. Y., & Ryu, K. H. (2019). An Empirical Comparison of Machine-Learning Methods on Bank Client Credit Assessments. *Sustainability*, *11*(3), 699. https://doi.org/10.3390/su11030699

Olakulehin, F. K., & Ojo, O. D. (2010). Factors Influencing the Completion of Dissertations by Students of Post-Graduate Diploma in Education (PGDE) by Distance Learning in South-western Nigeria. *Ανοικτή Εκπαίδευση: Το Περιοδικό Για Την Ανοικτή Και Εξ Αποστάσεως Εκπαίδευση Και Την Εκπαιδευτική Τεχνολογία*, *4*(1), 37. https://doi.org/10.12681/jode.9722

Orion, H. C., June Forosuelo, E. D., & Cavalida, J. M. (2014). *Factors Affecting Students' Decision To Drop Out of School* (Vol. 2, Number 1).

Sada, I., G. N., Prof. O., & Abubakar, M. (2025). Machine learning-based framework for predicting user satisfaction in e-Learning systems. *Journal of Basics and Applied Sciences Research*, *3*(2), 78–85. https://doi.org/10.4314/jobasr.v3i2.9

Suhaimi, N. M., Abdul-Rahman, S., Mutalib, S., Hamid, N. H. A., & Malik, A. M. A. (2019). Review on Predicting Students' Graduation Time Using Machine Learning Algorithms. *International Journal of Modern Education and Computer Science*, *11*(7), 1–13. https://doi.org/10.5815/ijmecs.2019.07.01

Tampakas, V., Livieris, I. E., Pintelas, E., Karacapilidis, N., & Pintelas, P. (2019). Prediction of Students' Graduation Time Using a Two-Level Classification Algorithm. In *Communications in Computer and Information Science* (pp. 553–565). https://doi.org/10.1007/978-3-030-20954-4_42

Thakar, P., Mehta, A., & Professor, A. (2015). Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue. In *International Journal of Computer Applications* (Vol. 110, Number 15).

Ünal, Y., Sağlam, A., & Kayhan, O. (2019). Improving classification performance for an imbalanced educational dataset example using SMOTE. *European Journal of Science and Technology*, 485–489. https://doi.org/10.31590/ejosat.638608