ISSN No:-2456-2165

Machine Learning-based Students' Enrollment Analytics: A Case Study of Polytechnics in Kebbi State

Samira Kabir Nabade
Department of Centre for information and Technology
Kebbi State Polytechnic Dakin-gari, Dakin-gari,
862106, Nigeria

Anas Shehu
Department of Computer Science
Kebbi State Polytechnic Dakin-gari, Dakin-gari,
862106, Nigeria

Abstract:- Machine learning is a subfield of Artificial Intelligence (AI) that equips computers to learn from records to make inferences about the future. It has been employed in different fields such as medicine, agriculture, finance, and education to make explanatory data analyses and projections. Machine learning models have been used in making projections and planning. In this project, we have built a machine learning model to predict students' enrollment with the view to having better planning in the area of teaching and non-teaching staff recruitment, lecture halls and laboratories development, and student hostel construction. We have used support vector machine (SVM). SVM achieved a root mean square error (RMSE) and coefficient of determination (R2) of 0.61 and 0.54.

I. INTRODUCTION

Admission application into our tertiary institution is on the increase. Several applicants sent their requests to gain admission into our institutions every year. Very few of them succeed in getting admission due to many reasons. Some of the reasons are inadequate teaching and non-teaching staff, and inadequate infrastructure (offices, laboratories, hostels, etc.). Government and school administrators do not have concrete data that will enable them to objectively plan for their schools' future needs in terms of human capacity and infrastructure. We already have a huge volume of data in our schools. We only need to prepare and format Those data in a way that we will make useful insights from them. Machine learning has the potential to extract and make insightful analyses from a large data set so that we plan well for future students' enrollment needs.

The project aims to develop a machine learning-based predictive model to estimate future students' enrollment.
The specific objectives of the work are:

- To collect the students' enrollment records for the last 7 years for the training and testing of the machine learning model to be developed.
- To design the model for the prediction
- To train the model with 70% of the student's enrollment records collected
- To evaluate the performance of the model.

II. RELATED WORKS

Students' success is a central concern of many higher education institutions in recent times especially with budget cuts and increasing operational costs, academic institutions are paying more attention to sustaining students' enrollment in their programs without compromising rigor and quality of education [1]. Machine learning (ML) frameworks have the capability to derive knowledge from data [2] which can enhance planning with regard to student enrollment, infrastructure, and staff development. Machine learning helps in other aspects of planning such as tree planting planning with regard to the increase in the number of pedestrians in a given city [3]. Recently, predictive analysis has relied on Machine Learning to support business decision-making. Applications in finance, operations and risk management are good attestations of the relevance of Machine Learning research in various business functions [1].

Machine learning frameworks and tools have been employed in different aspects of human life. There is hardly any field of human endeavor that has not benefited from machine learning. Prediction of students' enrollment pattern is a very important issue in planning and management [4] for the attainment of sustainable education for all citizenry. In this respect, Nita et al. (2022) used machine learning models (GANs) to present the result of students' structure predictions and compare them against real data obtained from a registry system of a European public institution of higher education in economic sciences. The research attempt provided a wealth of knowledge and insight into practical skills related to the potential application of such solutions and revealed a number of problems associated with student structure prediction tasks. The experiments revealed that for 11 out of the 48 examined datasets – the PSI index was in excess of 75% but was decidedly lower for the remaining sets (with 18 sets assessed below the margin associated with this specific form of management.

Jaafaru and Agbelie [5] developed a machine-learning model that took into consideration the decision-maker's preferences for ranking bridges using the Multi-Attribute Utility Theory. The authors chose 19 bridges for maintenance based on budget and performance using a genetic algorithm model. The model was observed to improve project productivity, reduce downtime, and improve bridge inventory and planning conditions. In the area of city planning, there have been studies conducted to improve the efficiency of city design with regard to the volume of pedestrians and trees to be planted to enhance the well-being of the pedestrians. In this direction, Li and Ma [3] proposed a methodology framework called LightGBM with K-fold Max variance Semi-Supervised Learning and DeepLab v3+ (KMSSL-DL). KMSSL-DL combines machine learning and computer vision technology to estimate pedestrian volume with unlabeled data from high dimensional urban features, and extract tree crowns from satellite imagery in the Central Business District, City of Melbourne, Australia. KMSSL part achieved an excellent prediction effect (R^2 score = 0.8360, RMSE score = 0.2304). The authors also used DeepLab v3+ to recognize and extract street trees from Google Earth satellite imagery with good performance (mIoU = 84.37). They combined the two results to conduct a pattern analysis, enabling us to find four patterns between street trees and pedestrian volume: more trees – more pedestrians (MTMP), more trees - fewer pedestrians (MTFP), fewer trees - more pedestrians (FTMP), fewer trees - fewer pedestrians (FTFP).

In a similar effort, Zeineddine, et al. [1] proposed the use of Automated Machine Learning to enhance the accuracy of predicting student performance using data available prior to the start of the academic program.

Integrating Adaptive Production Planning and Prescriptive Maintenance (PsM) in future factories provides a novel perspective for flexibility, customization, and resilience of production plans. In this regard, Elbasheer, et al. [6] proposed a framework for developing an intelligent Decision Support Agent (DSA) for integrated PsM and production planning and control (PPC) based on Reinforcement Learning.

Timely and appropriate discharge placement for patients who have undergone radical cystectomy (RC) remains challenging. [Zhao, et al. [7]] attempted to improve the discharge planning process by creating a machine learning model that helps to predict the need for nonhome hospital discharge to a higher level of care. The authors used patients undergoing elective radical cystectomy for bladder cancer from 2014−2019 were identified in the ACS-NSQIP database. They trained a gradient boosted decision tree on selected predischarge variables to predict discharge location, dichotomized into home and non-home. They also used threshold-moving to calibrate model predictions and evaluated model performance on a testing set using receiver operating characteristic and precision recall curves. Model performance was further examined in subgroups of interest.

III. MATERIAL AND METHOD

The research collected students' admission data into the computer science program of Waziru Umaru Polytechnic Birnin Kebbi and Kebbi State Polytechnic Dakingari from the National Board for Technical Education (NBTE). The collected data were digitized, and normalized, and formatted them so that they were suitable for the machine learning model. We split the data into 2 sets: 70% for training the model and 30% for testing the model.

A. Model description

The model is built with predictor variables as follows:

- Study mode (fulltime, Part-time).
- Gender (male, male)
- ND or HND

The 2014-2015 to 2017-2018 data set was used to train the model 2016-2017 to 2017-2018 set was used for testing. A total of 5200 students' data were used from Waziri Umaru Polytechnic Birnin Kebbi and Kebbi State Polytechnic Dakingari. The datasets are summarized in Table 1.

Table 1:Students' enrolment data per academic session

	Training dataset				**Testing dataset**	
	2014-2015	**2015-2016**	**2016-2017**	**2017-2018**	**2016-2017**	**2017-2018**
WUP	591	603	650	680	703	712
KPO	250	237	245	276	291	345

Key: WUP: Waziri Umaru Polytechnic, KPO: Kebbi State Polytechnic

B. Model evaluation metrics

We measured the performance of our model using Root Mean Square Error (RMSE) and coefficient of determination (R^2). The formula for the RMSE is given in equation 1 [8].

$$RMSE = \sqrt{\frac{\Sigma(y_i - \hat{y}_i)^2}{N - P}} \qquad (1)$$

Where:
y_i is the actual value for the ith observation
$\hat{y}_i$ is the predicted value for the ith observation
N is the number of observations
P is the number of parameter estimates.
The R^2 is calculated using the formula in equation 2 [9].

$$R^2 = 1 - \frac{SSR}{SST} \qquad (2)$$

$$SSR = \sum_i (\hat{y}_i - \bar{y})^2$$

$$SST = \sum_i (y_i - \bar{y})^2$$

Where:
SSR is the Sum of Square Regression
SST is the Sum of Squared Total
$\bar{y}$ is the mean value of the y value.

IV. RESULT AND DISCUSSION

We used R software to compute the RMSE and R^2 and obtained 0.61 and 0.54 respectively. This indicates that the model performance was fair. Even though results are not excellent, they reveal room model refinement. This could be realized by increasing the number of predictor variables and size of the training datasets. This will enable the model to learn more effectively to achieve greater prediction accuracy.

V. CONCLUSION

This work presents a machine learning model for the prediction of students' enrollment into national diploma (ND) and higher national diploma (HND) programs in computer science. The model achieved appreciated performance despite the paucity of data. The model could be improved by adding more independent variables (e.g., students' previous academic records, the scores in the unified tertiary matriculation examination UTME, and so on). This is imperative because the accuracy of predictions of machine learning lies on the size of the training data, the amount of training conducted, and the quality and number of predictor variables.

REFERENCES

[1.] H. Zeineddine, U. Braendle, and A. Farah, "Enhancing prediction of student success: Automated machine learning approach," Computers & Electrical Engineering, vol. 89, p. 106903, 2021.

[2.] K. Büttner, O. Antons, and J. C. Arlinghaus, "Applied Machine Learning for Production Planning and Control: Overview and Potentials," IFAC-PapersOnLine, vol. 55, pp. 2629-2634, 2022.

[3.] Z. Li and J. Ma, "Discussing street tree planning based on pedestrian volume using machine learning and computer vision," Building and Environment, p. 109178, 2022.

[4.] B. Nita, K. Nowosielski, Z. Kes, O. Sidor, P. Oleksyk, E. Walaszczyk, et al., "Machine learning in the enrolment management process: a case study of using GANs in postgraduate students' structure prediction," Procedia Computer Science, vol. 207, pp. 1350-1359, 2022.

[5.] H. Jaafaru and B. Agbelie, "Bridge maintenance planning framework using machine learning, multi-attribute utility theory and evolutionary optimization models," Automation in Construction, vol. 141, p. 104460, 2022.

[6.] M. Elbasheer, F. Longo, G. Mirabelli, A. Padovano, V. Solina, and S. Talarico, "Integrated Prescriptive Maintenance and Production Planning: a Machine Learning Approach for the Development of an Autonomous Decision Support Agent," IFAC-PapersOnLine, vol. 55, pp. 2605-2610, 2022.

[7.] C. C. Zhao, M. A. Bjurlin, J. S. Wysock, S. S. Taneja, W. C. Huang, D. Fenyo, et al., "Machine learning decision support model for radical cystectomy discharge planning," in Urologic Oncology: Seminars and Original Investigations, 2022, pp. 453. e9-453. e18.

[8.] J. Frost. (2023, 01/08/2023). Root Mean Square Error. Available: https://statisticsbyjim.com/regression/root-mean-square-error-rmse

[9.] A. Tripathi. (2023, 02/07/2023). Data Science Duniya. Available: https://ashutoshtripathi.co,/2019/what-is-coefficient-of-determination-r-square/