



Predicting Student Graduation Using Multilayer Perceptron.

¹S. K Alaramma, ¹B. I Ya'u, ²A. A. Habu, ¹M. Maishanu, ¹M. K. Dauda, ¹S. W. Mustapha, ¹M. A. Musa

¹Department of Mathematical Sciences,
Abubakar Tafawa Balewa University, Bauchi, Nigeria.

²University of St Andrews,
Jack Cole Building,
North Haugh, St Andrews, KY16 9SX, UK.

ABSTRACT

The prediction of student graduation holds significant value for higher institutions as it addresses the issue of student dropout. While the number of graduating students typically exceeds that of dropouts in a given cohort, there has been a recent increase in the number of students discontinuing their university studies. Various factors, including marital status, age, and prior academic performance, contribute to this problem. University management and other stakeholders are concerned about this alarming situation and recognize the need for urgent action. To tackle this challenge, the researchers present a novel dataset comprising five key parameters: gender, age, results from ordinary level examinations (such as WAEC, NECO, or NABTEB), Post Unified Tertiary Matriculation Examination (PUTME) score, and first-semester Grade Point Average (GPA). We employ a Multi-Layer Perceptron (MLP) model to predict students' graduation at a university. The dataset encompasses 2,512 records collected over an 8-year period. For model training and testing, we utilized cross-validation with three k-fold variations, specifically 2, 5, and 8. The study reveals that the MLP model achieves an impressive accuracy of 97.3%, with a recall of 92.1% and precision of 94.7%. When compared to other state-of-the-art models, the MLP-based approach demonstrates superior accuracy. Predictive models like the one developed in this study serve as valuable tools for both students and management in addressing the dropout issue.

ARTICLE INFO

Article History

Received: February, 2023

Received in revised form: April, 2023

Accepted: May, 2023

Published online: June, 2023

KEYWORDS

Machine Learning, MLP, Prediction, Retention, Student Success.

INTRODUCTION

The use of Machine Learning (ML) and Artificial Intelligence (AI) in predictive tasks have in recent years shown a considerable increase. A number of algorithms have been used to build models for evaluating student performance in schools. Universities have been encouraged to adopt the use of data-driven approach to reduce student dropout, increase retention and improved quality of education (Nuutila et al., 2018; Gabriele et al., 2019).

A serious nation needs to have a large share of its citizen educated and enlightened with the requisite skills for progress and development. One of the responsibilities of government world over is to provide its populace with quality education needed to produce employable high skilled workers that can drive the nation forward. Students that enrol into the university do so in order to become

graduate so that their lives can improve but unfortunately, some students do not make it, they ended up dropping out from the university and government is concern about the menace as it keeps increasing every year (Asif et al., 2017).

On the part of the university, good graduation rate is an indicator of a healthy institution and is among the factors used in measuring the position of universities in the education sector (Ojha et al., 2017). In addition, government and some donor agencies linked funding to graduation rate (Tsui et al., 2017).

The challenging issue of student quitting their educational pursuit in universities is very worrisome and literature revealed that some of the factors responsible include gender (Dayio, 2015), learning environment, marital status and age (Amuda et al., 2016), family issue (Asif et al., 2017) and previous academic records

Corresponding author: S. K. Alaramma

✉ skalarammane@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi

© 2023. Faculty of Tech. Education, ATBU Bauchi. All rights reserved



(Herath, 2018). When there are many issues that student need to battle with, focus on study have a tendency to reduce and once things get worst, they tend to voluntarily withdraw from the university to face the challenging situation squarely.

This paper explores the predictive power of seven (7) supervised machine learning algorithms and multilayer perceptron with a dataset consisting of 1956 observations to predict student graduation. The objective of the study is to determine which classification algorithms performs better as a predictive model.

The research work starts by investigating the factors responsible for predicting students' performance in university that several researches used in conducting their studies. A number of researches used age as factor in analysing students' performance. Some physiological changes in our brain as we grow older make us experience difficulties remembering things that leads to inefficiency in our learning pursuits. This inadequacy often caused some student to drop out (Zamani et al., 2018). However, assertion in (Zamani et al., 2018) have some under serious criticism and doubt. It has been reported in (Najafabadi et al., 2017) that age has no known role in determining student's graduation or quitting. Apart from age, gender has been used extensively as a predictor in determining students' performance. The research work conducted by (Shariff et al., 2016) to predict the graduation status of some students. The results show that male tend to drop out of school more than female and the finding was evidenced with the fact that female tends to have better self-discipline and more focus on their study or work compare to men (Potokri, 2011).

Grade Point Average (GPA) have also been used as a factor for predicting graduation in a number of studies. Some students' dropout of school due to very poor first semester GPA out of embarrassment or thinking that the chances of getting their desired class of degree has already been defeated.

According to Gopalakrishnan et al., (2018); Ahuja and Kankane, (2018); Salim et al., (2018) students' academic records is one of the most vital reasons why student drop out and it us used as feature in predicting student graduation. Previous student performance they

argued is what determine if student can make it or otherwise.

Marital status plays an important role in determining academic performance and retention. Potori, (2011) stated that married students have difficulties meeting up with their studies because they have other responsibilities to discharge. Completing assignment and preparation for test or examination are always traded with marital responsibilities and when they become enormous, some students opt to drop their studies.

Other factors that contribute to dropout include city, family occupation and income, job status, faculty, study mode, disability and even course of study.

Now that we have seen factors that are responsible for dropout, the study will shift to investigate the supervised machine learning and deep learning algorithms used in this type of prediction task in previous studies.

During the past decade, a number of researches have attempted to study student academic data to predict persistence rates, enrolment rates, and/or graduation rates. A number of statistical methods were used including factor analysis (Allen, 2017), multiple regression (Salim et al., 2018; Cubeta, et al., 2019), path analysis (Elkins, 2020), and discriminant analysis (Roweton, 1994). Each of these approaches produced various levels of success but in terms of discovering previously undetected patterns they are not effective. In addition, other researches in previous studies (Zamani et al., 2013; Guleria, et al., 2014; Ma and Zhao, 2018; Shahiri et al., 2015; Halinka et al., 2015; Hamadneh et al., 2018) employed the use of supervised machine learning and deep learning approaches to predict student performance in schools. The most widely used algorithms are: Decision Tree (DT), Naïve Bayes (NB), Artificial Neural Network (ANN), K-Nearest Neighbours (K-NN) and Support Vector Machines (SVM).

1. SVM: SVM was used in a number of studies to predict student performance by constructing a hyperplane in multidimensional space to separate class labels. Research work conducted by Ma and Zhao, (2018) analysed important features in determining pass rate and used the features with SVM and DT to predict student pass rate. The

Corresponding author: S. K. Alaramma

✉ skalarammane@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi

© 2023. Faculty of Tech. Education, ATBU Bauchi. All rights reserved

- performance of the built models was recorded. Finding from the research revealed that SVM has the highest accuracy. When the size of the dataset is large, SVM don't perform quite better than other algorithms as observed in Pushpa et al., (2017) and this observation was further established by Shahiri et al., (2015) where they based their findings and concluded that the size of the dataset has impact on the performance of SVM models as they perform better with s small dataset.
2. Naïve Bayes: this is among the simplest and easiest to use classification algorithms based on Bayesian theorem. The research work carried out by Pujianto et al., (2017) used NB to predict the academic performance of secondary school students on their respective programmes. The features used for building the model were derived from their academic records and classified as either pass or fail groups. WEKA tool was used as a platform to carry out the experiment and the result of the study was a model built with NB with an accuracy of 87%. Similarly, the findings from the work of Kamal and Ahuja, (2019) also conclude that NB is an easy algorithm for predicting students' performance and work well with large dataset. In addition, the work claimed that NB even though very simple to use algorithm, it has outperformed more complex algorithms in the experiments they carried out. Naïve Bayes tends to converge faster and therefore, takes smaller time to predict and perform better than other models. This is observation is among the findings of the research conducted by (Al-Shehri, 2017).
 3. DT: Decision Tree is among the most used ML algorithms in classification tasks. It leverages a tree structure and was used in the work of Guleria et al., (2014) to predict student academic result and the findings from the study claimed DT is helpful in the task and easy to interpret. The same claim has also been made in Paul and Kamal, (2015). Another interesting work that used DT in carrying out a similar task was the study conducted by Pushpa et al., (2017). The study performed set of experiments on students that are probably going to fail in their study. The results from the experiments show that DT has an accuracy of 73.92% in the prediction task while NB has 68.6% indicating higher accuracy achieved by DT.
 4. NN: A typical neural network has a system of interconnected neurons that has three layers; the input, hidden and output layer. As a machine learning algorithm, a number of research work employed the use of this approach in solving a lot of classification problems (Halinka et al., 2017). All the interaction between features used in NN from the input layer down to the output layer can be revealed by the network (Shahiri et al., 2015). Several studies have been carried out using NN to predict the performance of students in universities. Research work conducted by [56] analysed the impact of different features used with different machine learning algorithms including NN. The results of the finding concluded that NN has the highest accuracy of 78.6% when all features were included but when behavioural features were excluded, the performance of the model dropped to 64.5%. This indicates that the behavioural features were relevant to the prediction task. The work of De-Albuquerque et al., (2016) used features such as grade, length of study, previous academic records to train a NN to predict the performance of students. The result of the research work shows that NN has accuracy of 95%. Another research work conducted by Barker et al., (2019) reported the use of SVM and NN to predict student graduation rate in a 4-year degree programme. They conducted set of experiments with dataset that have 59 features consisting of students' academic records, demography, and attitudinal information. The results of the study show that NN has best accuracy of 67.5% and SVM has accuracy of 66.1%.
 5. KNN: K-Nearest Neighbour is another classical supervised machine learning classification algorithm that has been used extensively in educational data mining. In student performance prediction tasks, the works carried out by Jain (2016) and Sheetal et al., (2016) employed the use of KNN to conduct a set of experiments and in both studies, the models built have achieved accuracy of more than 95%. However, KNN has been known to have a

Corresponding author: S. K. Alaramma

✉ skalarammane@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi

© 2023. Faculty of Tech. Education, ATBU Bauchi. All rights reserved



serious drawback bothering on high computation time. This is because, the algorithm has to perform a one-to-one checking between the test data and the training data. This shows that the more the training data the more distance calculations needed.

It is evidently clear that numerous researches were carried out to predict the performance of students in higher institution. In this paper, we proposed a novel student dropout prediction education data mining model that will predict student dropout using a dataset having 9 features with Support Vector Machine, Logistic Regression, Naïve Bayes, Multilayer Perceptron and Random Forest Machine Learning classification algorithms with parameter tuning.

METHODOLOGY

This section describes the method used in carryout this research work and it is divided into two parts. The first part presents the educational data mining dataset description. The second part is the model description where the Artificial Neural Network is explained.

Dataset Description

This study used randomly selected data of students enrolled at Abubakar Tafawa Balewa University, a federal university of technology located in the Bauchi State, during a 5-year period (2014 to 2021) which contains 3,817 records and contains nine features, namely gender, age, GPA of first semester level one, OLevel Results, and Post UTME score, marital status, disability, faculty, Cumulative GPA (CGPA). See Table 2 below. These nine features are used as input variables and features for the machine learning algorithms. Each record has a class label, 'success' (denoted by 1) or 'failure' (denoted by 0). Number of records with the label 'success' (1) are 3604 pieces, and the remaining 213 records labelled 'failure' (0). Success here means that the student has successfully graduated while failure indicate that the student has voluntarily withdrawn from the institution. Note that the data contains some missing values on some features as summarized in Table 1. All data containing missing value is converted to zero. The feature 'Gender' contains a score of 1 (male) or 0 (female) while the GPA and CGPA parameters in the data are continuous numbers that are in the range [0, 5].

Table 1: Description of each feature in the academic data used

S/N	Parameter	Type	Number of incomplete field values
1	Gender	Discrete	8
2	Age	Integer	15
3	GPA	Float	0
4	PUTME Score	Integer	0
5	OLevel Score	Integer	2
6	CGPA	Float	3
7	Marital Status	Discrete	0
8	Disability	Discrete	0
9	Faculty	Integer	0

Table 2: Parameters in each student record

	Parameter	Value	Explanation
1	Gender	Boolean (1 digit)	Code denoting gender
2	Age	Integer (2 digits)	Student's age at the time of applying to college
3	GPA	Float	First semester GPA
4	PUTME Score	Integer (3 digits)	Student Post UTME Score
5	OLevel Score	Integer	Student Ordinary Level results Score
6	CGPA	Float	Cumulative GPA
7	Marital Status	Boolean	Student's Marital Status
8	Disability	Boolean	If Student has disability like blindness

Corresponding author: S. K. Alaramma

✉ skalarammane@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi

© 2023. Faculty of Tech. Education, ATBU Bauchi. All rights reserved

Parameter	Value	Explanation
9	Faculty	Integer
10	Success	Boolean (1 digit)

At any given time, the university maintains a variety of information on each student such as academic, personal and demographic information. Considering only the pre-enrolment information (the student's information at the time of seeking admission into the University) and first semester GPA, and subsequent CGPA, the study created a 9-parameter record for each of the students. Table 1 lists the parameters included in each record. In total 3, 817 profiles were collected. Randomly, 3,054 (80%) of the records in the dataset were selected and used for training the models (training set) and the remaining 763 (20%) for testing (testing set).

Model Description

The success of developing an accurate machine learning model partly depend on the choice of the underlying algorithm.

Among the prominent ones is the Artificial Neural Network (ANN) employed in solving regression and classification problems. Through a learning algorithm, ANN is able to learn patterns from a set of data. A number of researches in various fields such as the area of voice recognition, fraud detection, face recognition etc were conducted using ANN. This leads to its rapid development in the last decade.

Multilayer Perceptron (MLP) which is an implementation of ANN is one of the widely used model based on its reliability in the detection task. The model consists of three layers; the input layer, the hidden layer and the output layer. The smallest element in ANN is the neuron and in MLP, there is an interconnection between each neuron in each layer with the neuron in the layer above it. The MLP architecture is shown in Figure 1 below.

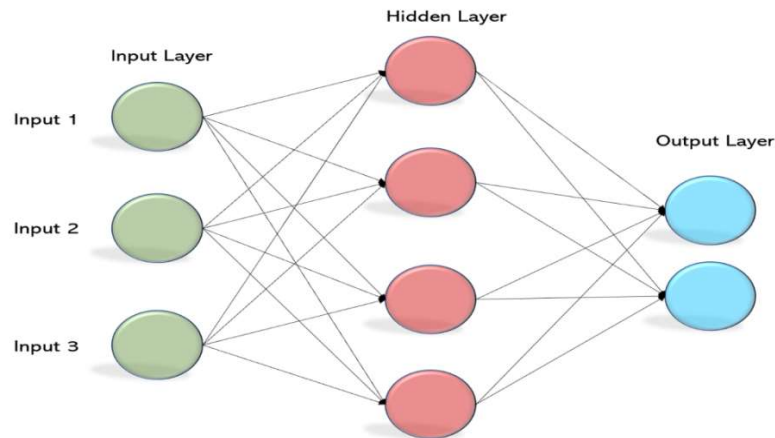


Figure 1 Multi-Layer Perceptron Architecture (MLP) with three layers

The Model Architecture

Three-layered perceptron was developed to harness the computational capabilities of a multilayer perceptron as a general pattern classifier. Since MLP consists of three layers, the first layer which is at the input level, comprised of 10 neurons (processing elements) - one for each profile parameter minus the "successful" graduation parameter.

The third layer which is at the output level, comprised of 2 neurons - one for denoting "successful" graduation and the other "unsuccessful". To establish the ideal number of processing elements in the second layer which is the hidden level, a sequence of tests was conducted. The result of the tests carried out indicated that 4 processing elements from the set of neurons produced the least error. The

connection among the various neurons is in such a way that each neuron is fully connected to every neuron in the subsequent layer. Regardless of the layer, each neuron accumulates input from the neurons in the previous layer and offers output to neurons in the layer just above it. To calculate each neuron's output we use the formula below:

$$y_i(t) = f(x_i(t)) \quad (1)$$

Where $y_i(t)$ is the output of neuron i at time t , $x_i(t)$ is an accumulation of input activity from the neurons at the lower layer.

In this network for f , hyperbolic tangent function (TanH) was used as a shifted and scaled sigmoid function. The function therefore, binds the range of each neuron to between -1 and 1, which often results in increased training efficiency (Heykin, 2019; W.S. Stornetta et al., 1987; S. I. Gallant, 1993).

Training

In educational data mining tasks, the desired response was usually known and therefore, the perceptron was trained with error correction learning. Let $y_i(n)$ denote the response of the system at iteration n on processing element i , and $d_i(n)$ for the desired response at element i on iteration n . Then, for a given input, an error $e_i(n)$ is defined by the formula below:

$$e_i(n) = d_i(n) - y_i(n) \quad (2)$$

The weight in network has to be adapted in compliance with the principle of gradient descent. Each weight will be corrected by a value that is proportionate to the present input and error at the weight. The formula below depicts the relationship:

$$w_{ij}(n+1) = w_{ij}(n) + \eta e_i(n) x_j(n) \quad (3)$$

Where η is a small constant, called the learning rate parameter, $w_{ij}(n)$ is the weight between processing elements i and j at iteration n , and $x_j(n)$ is the input value at processing element j at iteration n .

The weights in the network need to be updated by the use of an enhancement to the

straight gradient descent principle with a memory term which is the past increment to the weight in the network.

Equation (4) described equation:

$$w_{ij}(n+1) = w_{ij}(n) + \eta e_i(n) x_j(n) + \alpha (w_{ij}(n) - w_{ij}(n-1)) \quad (4)$$

Where α is a small constant, called the momentum.

The use of batch learning was employed in the training were all the patterns that describe the student profiles, were first presented then accumulated the weight updates, and at the end the training average weight update was used to update weight. Training typically take place over a number of epochs. An epoch is formed when weight is updated with the contribution of all patterns.

The training starts by allocating small random values for each. The weight in the hidden layer was updated by setting η to 1.1 and α to 0.6. In the output level η and α were set to 0.2 and 0.8 respectively.

RESULTS AND DISCUSSION

The ANN model used in this research is Multi-Layer Perceptron (MLP) model with 12 neurons in hidden layer. After some trials and errors using different scenarios, the number of neurons for the middle layer were obtained from the previous preliminary observation. Cross validation was done in testing and training processes on the students' records datasets. K-fold cross validation system with 2, 5, and 8 was used for the general performance of the system. The results of the test revealed that the accuracy for $k=2, 5$, and 8 were 96.4%, 98.3%, and 97.2% respectively.

One of the problems which classification problems get entangled with is the issue of over generalization which usually happen when the network has to simply memorise the training set and finds it difficult or unable to generalise the problem. With a little number of errors, the indication is that the network is over trained. To avoid over generalization, a series of tests need to be carried out to find an ideal number of training epochs. The need for finding this epoch was terminate the training session at an epoch where the maximum generalization of the

network was recorded. Based on the results of the tests conducted, the network produced the best predictive power after 6000 epochs.

The network predictive power was tested on the training dataset after training the network within the 6000 epochs with optimised network structure. The mean square error was

0.18. Out of 2,902 successful graduates in the dataset, 2,675 were classified correctly by the model that is about 92.2%. On the other hand, out of 152 unsuccessful graduates, 127 were classified correctly about 83.55%. Table 3 represents the network's performance when tested with the training dataset.

Table 3 Performance when tested with training data

<i>Performance</i>	<i>Successful</i>	<i>Unsuccessful</i>
MSE	0.17	0.17
NMSE	1.22	1.22
MAE	0.36	0.36
Min Abs Error	0.00	0.00
Max Abs Error	1.05	1.05
R	0.15	0.15
Percent Correct	92.2	83.55

When the predictive power of the network was tested with the testing dataset consisting of 763 observations, the mean square error was 0.22 on both the successful and unsuccessful graduates. The number of successfully classified graduates by the

network stands as 612 out of 681 that is 89.87%. On the other hand, 71 out of the 82 unsuccessful graduates were classified correctly representing 86.6%. Table 4 shows the network's performance when the testing dataset was used.

Table 4. Performance when tested with testing data.

<i>Performance</i>	<i>Successful</i>	<i>Unsuccessful</i>
MSE	0.21	0.21
NMSE	2.21	2.21
MAE	0.26	0.26
Min Abs Error	0.00	0.00
Max Abs Error	1.05	1.05
R	0.21	0.21
Percent Correct	83.09	85.04

Table 5. Data of system test result by Cross Validation method

<i>Measurements</i>	<i>2-fold</i>	<i>5-fold</i>	<i>8-fold</i>	<i>Average</i>
Accuracy	96.4%	98.3%	97.2%	97.3%
Precision	90.8%	92.3%	93.2%	92.1%
Recall	93.9%	94.4%	95.8%	94.7%

When cross validation method was employed, the model built achieved a good performance in the classification task with both recall and precision exceeding 90%. At 8-fold cross validation, precision and recall are 93.2% and 95.8% respectively. The amount of data used in the training in each in the k-folds increases with the value of k and in addition, accuracy, precision and recall also tend to

increase as well. This result is depicted in Table 5 above.

As a form of comparison, other models were created using various machine learning algorithms earlier used in previous studies, such as k-NN, Naive Bayes, SVM, and DT. Table 6 summarizes the comparison of accuracy, precision and recall between MLP's used in this study with other classifiers.

Table 6 Comparison of test results between several classification methods

Methods	Measurements	2-fold	5-fold	8-fold	Average
MLP	Accuracy	96.4%	98.3%	97.2%	97.3%
	Precision	93.8%	92.3%	90.2%	92.1%
	Recall	93.9%	94.4%	95.8%	94.7%
KNN	Accuracy	93.3%	92.4%	91.7%	92.5%
	Precision	98.2%	96.4%	94.2%	95.3%
	Recall	92.6%	95.1%	93.9%	93.9%
SVM	Accuracy	92.8%	93.8%	94.2%	93.6%
	Precision	93.2%	94.1%	94.4%	93.9%
	Recall	91.6%	93.8%	94.2%	93.2%
NB	Accuracy	81.3%	84.1%	81.5%	82.3%
	Precision	82.7%	87.8%	82.1%	84.2%
	Recall	81.1%	83.4%	81.2%	81.9%
DT	Accuracy	92.2%	90.1%	93.8%	92%
	Precision	92.8%	90.4%	92.3%	91.8%
	Recall	90.6%	89.8%	92.7%	91.0%

Experimental results have shown that MLP outperforms the other algorithms with an average accuracy of 97.3% and an average precision and recall of 92.1% and 94.7% respectively. Closely related in terms of performance was K-NN ($k = 3$) with even a slightly better precision (96.6%). However, to classify every single test data, K-NN requires a considerable computation. In addition, Naïve Bayes achieved the lowest performance when compared with the other machine learning algorithms used in this study. Typically, when feature has no correlation to the class label, Naïve Bayes tend to have no sensibility to such feature and that is one of its weaknesses. Among the features used in this study, it was observed that 'age' and 'gender' have less significant role in predicting graduation or dropout. Finally, SVM and DT have good accuracy although still lower than MLP.

CONCLUSION

Machine Learning algorithms have been employed in several researches to predict the performance of students in institutions of learning. In this paper, models to predict students' graduation have been developed from selected features in students' records. These features are: gender, age, first GPA, O'Level Results, and Post UTME score, marital status, disability, faculty, and CGPA. ANN and specifically MLP has been selected to build model for this prediction. The MLP network with

three-layered perceptron has been trained and tested using backward propagation. A dataset with 3,817 records of university students divided into two sets of 3,054 and 763 was used for training and testing respectively. In the model built using MLP has achieved good performance with accuracy of 97.3%, a recall of 92.1% and precision of 94.7%. This show that MLP is a reliable classifier and suitable for educational data mining prediction task.

More work can be done in the future to improve on thus findings. First, increasing the number of features for each student in the dataset can enhance the performance of the system. Future direction will look into that aspect. An equally important follow up of this work is on the determination of the significance of the contribution each feature has in the prediction task. This can help in two ways. First, the school management can tackle the issue of students' dropout since the reason is known beforehand. In addition, it can help in finetuning the model by reducing noise through elimination of features that least contribute to the success of the prediction and thus optimise the network.

REFERENCES

- Ahuja, R. and Kankane, Y. (2018). "Predicting the probability of student's degree completion by using different data mining techniques," 2017 4th Int. Conf. Image Inf. Process. ICIIIP



- 2017, vol. 2018–Janua, pp. 474–477, 2018.
- Allen, D. (2017). The Hunger Factor in Student Retention: An Analysis of Motivation. AIR 2017 Annual Forum Paper. Paper presented at the 37th Annual Forum of the Association for Institutional Research, Orlando, FL., May 2017. (ERIC Document Reproduction Service No. ED 410 865)
- Al-Shehri H. (2017). "Student performance prediction using Support Vector Machine and K-Nearest Neighbor," Conference Electr. Comput. Eng., pp. 17–20, 2017.
- Amuda, B. G., Bulus, A. K., and Joseph, H. P. (2016). "Marital Status and Age as Predictors of Academic Performance of Students of Colleges of Education in the Nort- Eastern Nigeria," *Am. J. Educ. Res.* 4, 2016, Pages 896-902, vol. 4, no. 12, pp. 896–902, 2016.
- Asif, R., Merceron, A., Ali, S. A. and Haider, N. G. (2017). "Analyzing undergraduate students' performance using educational data mining," *Comput. Educ.*, 113, pp. 177–194, 2017.
- Cubeta, J. F., Travers, N. L., and Scheckley, B. G. (2019). Predicting the Academic Success of Adults from Diverse Populations. *Journal of College Student Retention*, 2(4), 2019, P. 295-311.
- Dayio, M. (2015). "Gender Differences in Academic Performance in a Large Public University in Gender Differences in Academic Performance in a Large Public University in Turkey Department of Economics Serap Türüt-A ik," no. April, 2015.
- De-Albuquerque, R. M., Bezerra, A. A., De Souza, D. A., and Nascimento, D. C. (2016). "Using neural networks to predict the future performance of students," 2015 Int. Symp. Comput. Educ. SIIE 2015, pp. 109–113, 2016.
- Elkins, S. A, Braxton, J. M., and James, G. W. 2020. Tinto's Separation Stage and its Influence on First-semester College Student Persistence. *Research in Higher Education*, 41(2), 2020, P. 251-268.
- Gabriele, A.J., Joram, E., & Park, K.H. (2019). Elementary mathematics teachers' judgment accuracy and calibration accuracy: Do they predict students' mathematics achievement outcomes? *Learning and Instruction*, 45, 49–60.
- Gopalakrishnan, A., Kased, R., Yang, H., Love, M. B., Graterol, C., and Shada, A. (2018). "A multifaceted data mining approach to understanding what factors lead college students to persist and graduate," Proc. Comput. Conf. 2017, vol. 2018–Janua, no. July, pp. 372–381, 2018.
- Guleria, P., Thakur, N., and Sood, M. (2014). "Predicting Student Performance Using Decision Tree Classifiers and Information Gain," pp. 126–129, 2014.
- Halinka, A., Rzepka, P., and Szablicki, M. (2015). "Agent model of multi-agent system for area power system protection," Proc. - Int. Conf. Mod. Electr. Power Syst. MEPS 2015, pp. 1–4, 2015.
- Hamadne, N. N., Khan, W. S., and Khan, W. A. (2018). "Prediction of thermal conductivities of polyacrylonitrile electrospun nanocomposite fibers using artificial neural network and prey predator algorithm," J. King Saud Univ. - Sci., 2018.
- Haykin, S. (2019). *Neural Network and Learning Machines – Third Edition*. New Jersey: Pearson Prentice Hall.
- Herath, D. (2018). "Educational Data Mining to Investigate Learning Behaviors: A Literature Review," *Indian J. Sci. Technol.*, 8, July, pp. 1–12, 2018.
- Jain, R. (2016). Application of KNN-Genetic Algorithm for Analyzing Student Learning in Educational Data Mining Paradigm. *International Journal of Innovative Research in Computer and Communication Engineering*, 4(6), 10319-10323. doi: 10.15680/IJIRCCE.2016. 0406011.
- Kamal, P., and Ahuja, S., Academic Performance Prediction Using Data Mining Techniques: Identification of

Corresponding author: S. K. Alaramma

✉ skalarammane@gmail.com

Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi

© 2023. Faculty of Tech. Education, ATBU Bauchi. All rights reserved



- Influential Factors Effecting the Academic Performance in Undergrad Professional Course, vol. 741. Springer Singapore, 2019.
- Ma, X., and Zhou, Z. (2014). "Student pass rates prediction using optimized support vector machine and decision tree," 2018 IEEE 8th Annu. Comput. Commun. Work. Conf. CCWC 2018, vol. 2018–January, pp. 209–215, 2018.
- Najafabadi, M. K., Mohamed, A. H., and Mahrin, M. N. (2017). "A survey on data mining techniques in recommender systems," *Soft Comput.*, pp. 1–28, 2017.
- Nuutila, K., Tuominen, H., Tapola, A., Vainikainen M.P., & Niemivirta, M. (2018). Consistency longitudinal stability and predictions of elementary school students' task interest success expectancy and performance in mathematics. *Learning and Instruction*, 56, 73–83.
- Ojha, T., Heileman, G. L., Martinez-Ramon, M., and Slim, A. (2017). "Prediction of graduation delay based on student performance," 2017.
- Paul S., and Kamal, S. (2015). "Data Mining Techniques for Predicting Student Performance," *IEEE*, no. March, pp. 0–2, 2015.
- Potokri, O. C. "The Academic Performance Of Married Women Students In Nigerian Higher Education Doctor Of Philosophy (Phd) In Management And Policy Studies University Of Pretoria, South Africa Promoter : Prof . Venitha Pillay," 2011.
- Pujianto, U., Azizah, E. N., and Damayanti, A. S. (2017). "Naive Bayes to Predict Students' Academic Performance at Faculty of Literature," *Med. Image Comput. Comput. Interv. –MICCAI 2017*, pp. 163–169, 2017.
- Pushpa, M., Mrunal, M., Singh, H., and Suhas, C. (2017). "Class Result Prediction using Machine Learning," *Int. Conf. Intell. Human Machine Syst. Cybern.*, 2017 vol. 6, pp. 1208–1212, 2017.
- Salim, A., Heileman, G. L., Hickman, M., and Abdallah, C. T. (2018). "A geometric distributed probabilistic model to predict graduation rates," 2017 IEEE SmartWorld Ubiquitous Intell. Comput. Adv. Trust. Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People Smart City Innov. SmartWorld/SCALCOM/UIC/ATC/CB DCom/IOP/SCI 2017, pp. 1–8, 2018.
- Shahiri, A. M., Husain, W., and Rashid, N. A. (2015). "A Review on Predicting Student's Performance Using Data Mining Techniques," *Procedia Comput. Sci.*, vol. 72, pp. 414–422, 2015.
- Shariff, S. S. R., Rodzi, N. A. M., Rahman, K. A., Zahari, S. M., and Deni, S. M. (2016). "Predicting the 'graduate on time (GOT)' of PhD students using binary logistics regression model," *AIP Conference Proceedings*, 2016.
- Tsui, P., Murdock, T., and Mayer, L. (2017). *Trend analysis and enrolment management*. AIR 2017 Annual Forum Paper. Paper presented at the 37th Annual Forum of the Association for Institutional Research, Orlando, FL., May 2017. (ERIC Document Reproduction Service No. ED 410 887)
- Zamani, N. W., Shaliza, S., and Khairi, M. (2018). "A Comparative Study on Data Mining Techniques for Rainfall Prediction in Subang," *Proceeding Int. Conf. Math. Eng. Ind. Appl. 2018 (ICoMEIA 2018)*,