Using Supervised and Unsupervised Machine Learning Models to Analyze Students Academic Performance

Osondu Everestus Oguike, Emmanuel Chukwudi Ukekwe, Gabriel Abiodun Elufidodo

***Abstract:** Examination result repositories generated by most universities can serve as machine learning datasets for training various models to gain insights from the data. These datasets can train multiple linear regression models to determine a student's cumulative grade point average (CGPA), or the score that a student will get in specific courses. Additionally, classification-based supervised machine learning models can use these datasets to provide insights into the class result that a student will obtain. These insights can be invaluable for academic advising and early intervention. Moreover, these datasets can train clustering-based unsupervised machine learning models, such as the K-means clustering model, to understand how student results are grouped into various clusters. This information can be crucial for planning and evaluating the quality of the university. This paper uses the dataset of undergraduate students' examination results from the Department of Computer Science at the University of Nigeria, Nsukka, to train three supervised machine learning models and one unsupervised machine learning model, utilizing Jupyter Notebook as the Python IDE. The training results showed acceptable accuracies of 91.5% for the Naïve Bayes model and 95.1% for the Decision Tree model. The linear regression model demonstrated a negligible root mean square error of 8.23×10−18, while the K-means clustering model exhibited an acceptable Silhouette metric of 0.12.*

***Keywords:** Naïve Bayes model, decision tree model, K-means clustering model, linear regression, students' academic performance.*

I. INTRODUCTION

The grade point grading system has generated extensive records of students' results in university examination units. Traditionally, these results are only used to prepare transcripts for students when needed. Machine learning, a branch of artificial intelligence, offers new methods for uncovering hidden insights in accumulated data. Consequently, repositories of students' examination results data can be transformed into machine learning datasets, allowing various machine learning models to reveal these insights, [18].

Manuscript received on 24 July 2024 | Revised Manuscript received on 05 August 2024 | Manuscript Accepted on 15 September 2024 | Manuscript published on 30 September 2024.
*Correspondence Author(s)

Osondu Everestus Oguike*, Department of Computer Science, University of Nigeria, Nsukka, Nigeria. Email ID: osondu.oguike@unn.edu.ng, ORCID ID: 0000-0002-4833-5278

Emmanuel Chukwudi Ukekwe, Department of Computer Science, University of Nigeria, Nsukka, Nigeria. Email ID: emmanuel.ukekwe@unn.edu.ng

Gabriel Abiodun Elufidodo, Department of Computer Science, University of Nigeria, Nsukka, Nigeria. Email ID: gabriel.elufidodo@unn.edu.ng

In addition to determining the class of degree and CGPA, a student achieves, these datasets can be used to predict the grades or scores a student might receive in specific courses to attain a particular class of degree. This information can assist academic advisers in setting realistic goals for students and providing early intervention. Additionally, the data can be used to cluster students' performance into different groups, which can aid in planning and evaluating the quality of the university or department.

A. Statement of the Problem

Using machine learning models to gain insights from students' examination performance results can solve many problems. Uninformed academic advising, due to lack of insight from this study will hurt the student. Such negative impacts include poor academic performance, late intervention in guiding the students, etc. Therefore, this paper addresses the following problems:

- Poor student academic performance due to uninformed academic advising.
- Difficulty in knowing when and how to intervene to improve a student's academic performance.
- Difficulty in making effective plans due to a lack of insights from trends in students' performance.
- Difficulty in rating the quality of a department or university based on the overall performance of students over many years.

B. Aim and Objectives

This study aims to analyze students' examination results in the Department of Computer Science at the University of Nigeria, Nsukka, using a dataset of their results. This analysis will provide insights for proper academic advising and planning. The specific objectives of this study are:

- To prepare a machine learning dataset from the available repositories of student results in the Department of Computer Science, University of Nigeria, Nsukka.
- To train the following machine learning models using the dataset of student results: Naïve Bayes, Decision Tree, Multiple Linear Regression, and K-means clustering.
- To analyze, predict, and visualize the student results dataset.
- To evaluate the trained models and tune their parameters to improve performance.

II. LITERATURE REVIEW

Different studies have utilized various datasets with different attributes to train machine learning algorithms for predicting student performance. This section summarizes and synthesizes these studies.

A. Analysis Based on Supervised Machine Learning Models

The available literature reveal that classification-based supervised machine learning models have been widely used to predict students' academic performance in tertiary institutions, using secondary school CGPA data for early intervention [1], [2], [9], [10] [12], [14], [15]. For instance, [1] utilized various classification-based supervised machine learning algorithms, such as Naïve Bayes, KNN, Support Vector Machine, XGBoost, and Multi-Layer Perceptrons, to predict student performance, while [2] employed only logistic regression to predict whether a student would be successful or not. Similarly, [9] used Random Forest, Naïve Bayes, Multi-Layer Perceptron (MLP), and Decision Tree (J48) to predict student performance, whereas [9] applied Decision Trees (DT), K-Nearest Neighbors (KNN), Naive Bayes (NB), Linear Discriminant Analysis (LDA), and LogitBoost (LB). While [1] and [2] used secondary school data before entry into tertiary institutions, [9] used datasets of university graduates collected online, and [10] utilized two real educational datasets, enhancing the quality of predictions through feature selection methods like the Enhanced Whale Optimization Algorithm, Sine Cosine Algorithm, and Logistic Chaotic Map. To improve the accuracy of classification-based supervised machine learning algorithms such as Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), Naïve Bayes (NB), and K-nearest neighbor (KNN), ensemble classifiers like Extreme Gradient Boosting (XGB), Random Forest (RF), and Heterogeneous Ensemble Method (HEM) were used in [5] to predict engineering students' performance. Their dataset included demographic, cognitive, and non-cognitive attributes. Additionally, a hybrid of two ensemble classifiers—random forest and simulated annealing—was used in [7] to predict student performance. These classifiers were combined as the Improved Random Forest Classifier (IRFC). The dataset included attributes such as individual basic information (birthplace, gender, place of entrance examination, and direct contact), individual education information (student grade level, teaching semester, teaching classroom, student major, faculty members, and course scores), and individual behavior information. Regarding accuracy, a review study [6] [23] asserted that the limited research in various machine learning approaches contributes to inaccuracies in predicting student performance despite the large volume of educational data. They identified six commonly used models: support vector machine (SVM), linear regression (LinR), decision tree (DT), artificial neural networks (ANNs), K-nearest neighbor (KNN), and Naive Bayes (NB). The common attributes used in predictions included academic, demographic, internal assessment, and family/personal attributes. Given that performance depends on dynamic knowledge, [8] used recurrent neural networks (RNN) suitable for dynamic time series. They utilized six public datasets and generated a seventh dataset, all related to mathematical problem examinations. While most reviewed literature focused on predicting examination performance, [11] determined the importance of different factors affecting student performance using classifiers such as random forest (RF), support vector machines (SVM), logistic regression (LR), and artificial neural network (ANN). The factors considered in the dataset included behavioral information, individual information, and student scores collected from teachers through one-to-one surveys and online data, the results obtained showed that ANN had the best performance on the training data. Among the reviewed studies, only a few, such as [13] and [16], used deep learning to predict academic performance/results. [13]. used course data, demographic data, socio-economic information, and student course grades, while [16] employed CGPA datasets to train deep learning models, reporting high accuracy after evaluation.

B. Analysis Based on Unsupervised Machine Learning

A review article, [17] reported that both machine learning and clustering are useful in predicting students' academic performance. The study highlighted that clustering is particularly effective for categorizing students according to their performance. The most common unsupervised machine learning models used to analyze student results are clustering models, as demonstrated in [3] and [4]. In [3] [19][20][21][22], the K-means clustering model was used to group students' academic performance into various clusters. In [4], IFCM was utilized to cluster students' performance data during the COVID-19 lockdown. Both studies visualized the clusters obtained, which helped understand the general performance of students over many years, aiding in planning and decision-making.

C. Identified Gaps in Literature

Although numerous studies have explored the use of different machine learning models to predict students' performance, most focused on predicting overall performance for early intervention, typically using CGPA as the target/class label. To our knowledge, none has attempted to predict students' specific performance in particular courses to achieve a certain CGPA, i.e., using a course score as the target/class label. Furthermore, while various datasets have been used in the literature to predict academic performance, none have utilized the private examination result dataset from the Department of Computer Science, University of Nigeria, Nsukka. This study aims to address these gaps.

III. METHODOLOGY

A. Description of Dataset

The dataset used to train various machine learning models consists of students' examination results and CGPA from the Department of Computer Science, University of Nigeria, Nsukka. It covers all levels from the 2013/2014 to 2019/2020 academic sessions and includes only students with a minimum of forty-one examination scores. The dataset contains forty-four attributes (columns) and 1786 instances (rows). Of the forty-four attributes, forty-one attributes represent the various scores, one attribute represents the student ID, another attribute represents the CGPA, and the last attribute represents the class of the result. Scores for a student in a particular course range between 0 and 100, while CGPA values range between 0 and 5. The dataset is structured with the following attributes:

[ID, SC1, SC2, SC3, SC4, SC5, SC6, SC7, SC8, SC9, …, SC41, CGPA, Degree_Class]

Using multiple linear regression to predict the score that a student will obtain in a particular course, the score for that course will be the class/target attribute, while to predict the CGPA, the CGPA attribute will be the class/target attribute. Furthermore, using any of the classification models to predict the class of degree a student will obtain, the Degree_Class will be the class/target attribute, which can take the values: 'First Class', 'Second Class Upper', 'Second Class Lower', 'Third Class', or 'Pass'

B. Multiple Linear Regression Model

Multiple linear regression is a supervised machine learning algorithm that predicts a continuous variable as the class/target attribute, like CGPA or course score. It assumes a linear relationship between the target attribute and each of the non-target attributes. This assumption can be verified by estimating the correlation coefficient between the target attribute and each non-target attribute. After training the multiple linear regression model that predicts the score that a student will obtain in a particular course, the learner's output, which is the multiple linear equation used for prediction, takes the form shown in Equation (1).

SC41=A0+A1*SC1+A2*SC2+…+Acgpa*CGPA (1)

A0, A1, A2, …, Acgpa are the regression coefficients, which will be determined after the training of the linear regression model, using the dataset. To predict the CGPA, the regression equation obtained after training is shown in Equation (2).

CGPA=A0+A1*SC1+A2*SC2+…+A41*SC41 (2)

The metrics for evaluating multiple linear regression models include Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and Mean Absolute Error (MAE).

C. Naïve Bayes Model

The Naïve Bayes model is a classification-based machine learning model used to predict a class or category as the target attribute. After training the Naïve Bayes model with an appropriate dataset, the learner's output is a table of conditional probabilities based on Bayes' theorem, as given in Equation (3).

$$\mathrm{P(H\,|\,E)} = \frac{P(E\,|\,H)}{P(E)} P(H) \qquad (3)$$

P(H|E) is called the posterior probability, while P(H) is called the prior probability and $\frac{P(E\,|\,H)}{P(E)}$ is called the likelihood ratio. Bayes' theorem can be rephrased as "Posterior probability equals the likelihood ratio multiplied by the prior probability."

D. Decision Tree Model

The Decision Tree model is another classification-based machine learning model that predicts class or category as the target attribute. After training with a specific dataset, the learner's output is a tree diagram, which can be traversed from various nodes to the leaf of the tree. The nodes represent the non-target attributes of the dataset, while the edges or branches represent the various values the attribute can assume. The model uses the concept of Information Gain (IG) to determine the attribute for each node. The attribute with the highest information gain forms the node, as computed using Equations (4) and (5).

IG = Entropy(attribute) – Weighted(Entropy(Children)) (4)

The children are the various values of an attribute.

$$Entropy(attribute) = \sum_{i=1}^{k} - p_i \log_2 p_i \qquad (5)$$

E. K-Means Clustering Model

K-Means clustering is a widely used clustering-based unsupervised machine learning model that groups a dataset into various clusters of similar data. The model splits a dataset into K arbitrary clusters. Each cluster has a central value, called the centroid, which changes in each iteration of the algorithm. The initial choice of K clusters and the assignment of data into these clusters can be done arbitrarily. Data items are then reassigned based on the cluster that has the minimum Euclidean distance from the data point to the centroids. This process continues until the current assignment is the same as the previous assignment. The Silhouette metric can be used to evaluate the K-Means clustering model.

F. Training the Machine Learning Models

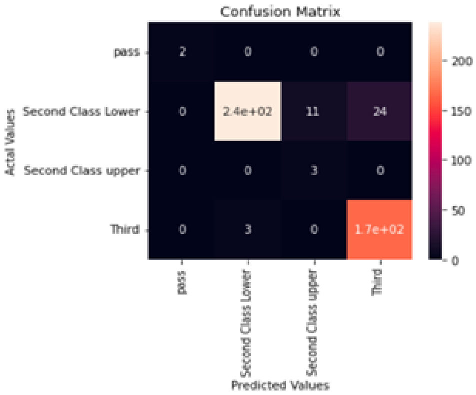
Training the models started with data cleaning using the Python library pandas. The cleaning phase began with checking for missing values. Columns with more than 80% missing values were discarded. For the remaining columns, missing values were replaced with the mean. After cleaning, the dataset distribution was viewed using histograms to check for skewness. For classification-based supervised machine learning algorithms, the categorical data in the class column, Degree_Class, which can have any of the following values, (First Class, Second Class Upper, Second Class Lower, Third Class, and Pass) were encoded into numerical values using an ordinal encoder. The dataset was then divided into feature variables (X) and target variable (Y) and split into training and testing datasets in a 75:25 ratio. Both training and testing features were scaled using the standard scaler module to reduce patterns and disparities in the data. The scaled dataset was used to train the various machine learning models, and the testing dataset was used to evaluate the performance of the models. For classification models, the confusion matrix and accuracy score were used for evaluation. For regression analysis, the root mean square error was used. For the K-Means clustering algorithm, Principal Component Analysis (PCA) was used to form relevant attributes. Using 80% of the cumulative explained variance, 24 components were created from the 43 attributes of the dataset, as shown in Figure 1.

IV. RESULTS AND DISCUSSION

From the analysis, the Naive Bayes classification model produced an accuracy of 91.5%, implying that for every 100 instances of the data, the model can predict at least 91 instances correctly. However, this accuracy alone is not sufficient to fully justify the performance of the model.

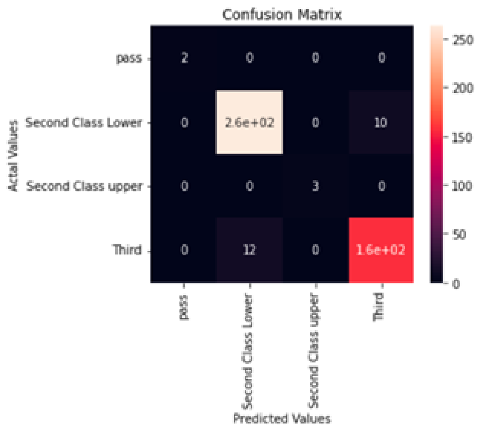
Therefore, another evaluation metric called the Confusion Matrix was used, and the results are shown in Table I below.

Table I. Confusion Matrix for the Naïve Bayes Model



The confusion matrix illustrates the number of correctly and incorrectly predicted values. It indicates that 238 instances of the "Second Class Lower" grade were correctly predicted, while 35 were incorrectly predicted. Additionally, 166 instances of the "Third Class" grade were correctly predicted, with only 3 being incorrectly predicted. All other grades were correctly predicted. In comparison, the decision tree classifier achieved an accuracy score of 95.1%. The confusion matrix for this classifier is shown in Table II. This matrix reveals a slight difference from that in Table I: 10 instances of "Second Class Lower" were incorrectly predicted, and 12 instances of "Third Class" were also wrongly predicted. All other classes were correctly predicted.

Table II. Confusion Matrix for the Decision Tree Model



The performance of the regression analysis was evaluated using the Root Mean Squared Error (RMSE), which measures the closeness of predicted values to actual values. The RMSE value for the model was 8.23*10-18, indicating that the predicted values are nearly identical to the actual values. For the K-Means clustering model, using the 24 components obtained from the Principal Component Analysis, as shown in the Cumulative Explained Variance Against Number of Components of Figure 1. The K-Means elbow method determined that 5 clusters were the optimal number of clusters, as shown in Figure 2. This naturally divides the dataset into five classes of results: First Class, Second Class Upper, Second Class Lower, Third Class, and Pass. The Silhouette metric for the K-Means clustering was 0.12. The number of data items in the five clusters is 594, 202, 464, 310, and 216, respectively, as visualized in Figure 3.

V. CONCLUSION

This paper analyzes the academic performance of students from the Department of Computer Science at the University of Nigeria, Nsukka, using three supervised machine learning models—Multiple Linear Regression, Naïve Bayes, and Decision Tree—along with one unsupervised machine learning model, K-Means Clustering. These models provided various insights into the students' academic performance, including predictions of course scores, CGPA, and degree classification, as well as the clustering of student performance into different groups. These insights can be useful for early intervention aimed at improving student performance and for assessing the department and university.

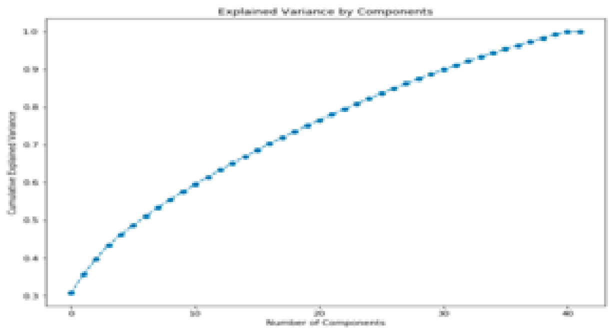


Figure 1. Cumulative Explained Variance Against Number of Components

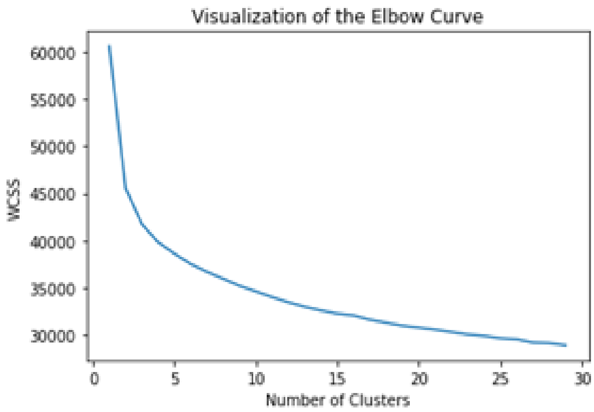


Figure 2. The Elbow Curve that Determines Optimal Number of Clusters

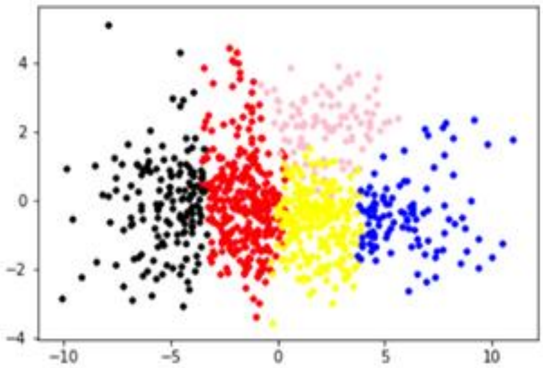


Figure 3. Visualization of the Clusters

Retrieval Number: 100.1/ijsce.D36401404092
DOI: 10.35940/ijsce.D3640.14040924
Journal Website: www.ijsce.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Financial Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted with objectivity and without any external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Authors Contributions:** The authorship of this article is attributed equally to all participating authors.

REFERENCES

1. Khalil Ahammad, Partha Chakraborty, Evana Akter, Umme Honey Fomey, Saifur Rahman, "A Comparative Study of Different Machine Learning Techniques to Predict the Result of an Individual Student using Previous Performances," International Journal of Computer Science and Information Security (IJCSIS), Vol. 19, No. 1, January 2021. DOI:10.5281/zenodo.4533374
2. Mohammed Nasiru Yakubu, "A. Mohammed Abubakar, Applying machine learning approach to predict students' performance in higher educational institutions," Kybernetes © Emerald Publishing Limited 0368-492X DOI 10.1108/K-12-2020-086, https://www.emerald.com/insight/0368-492X.htm
3. Revathi Vankayalapati, Kalyani Balaso Ghutugade, Rekha Vannapuram, Bejjanki Pooja Sree Prasanna, "K-Means Algorithm for Clustering of Learners Performance Levels Using Machine Learning Techniques," Revue d'Intelligence Artificielle, Vol. 35, No. 1, February, 2021, pp. 99-104 Journal homepage: http://iieta.org/journals/ri https://www.researchgate.net/publication/350550593_K-Means_Algorithm_for_Clustering_of_Learners_Performance_Levels_Using_Machine_Learning_Techniques https://doi.org/10.18280/ria.350112
4. K.P. Prakasha , and K Selvakumari, "An Intelligent Clustering Technique for Analysing the Performance of Students during Lockdown Period of Covid-19," Turkish Journal of Computer and Mathematics Education Vol.12 No.9 (2021), 2499– 2512 DOI: https://doi.org/10.17762/turcomat.v12i9.3733
5. A'zraa Afhzan Ab Rahim, Norlida Buniyamin, "Predicting Engineering Students' Academic Performance using Ensemble Classifiers- A Preliminary Finding," JOURNAL OF ELECTRICAL AND ELECTRONIC SYSTEMS RESEARCH https://doi.org/10.24191/jeesr.v20i1.013
6. Yazan A. Alsariera, Yahia Baashar, Gamal Alkawsi, Abdulsalam Mustafa, Ammar Ahmed Alkahtani and Nor'ashikin Ali, "Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance," Hindawi Computational Intelligence and Neuroscience Volume 2022, Article ID 4151487, 11 pages https://doi.org/10.1155/2022/4151487
7. Shaohai Huang1 and Junjie Wei, "Student Performance Prediction in Mathematics Course Based on the Random Forest and Simulated Annealing," Hindawi Scientific Programming Volume 2022, Article ID 9340434, 9 pages https://doi.org/10.1155/2022/9340434
8. Marina Delianidi, and Konstantinos Diamantaras, "KT-Bi-GRU: Student Performance Prediction with a Recurrent Knowledge Tracing Neural Network," IEEE TRANS. LEARNING TECHNOLOGIES, May 26, 2022 DOI:10.36227/techrxiv.20055545.v1https://doi.org/10.36227/techrxiv.20055545.v1
9. M. P. R. I. R. Silva, R. A. H. M. Rupasingha, B. T. G. S. Kumara, "A Comparative Study of Predicting Students' Academic Performance Using Classification Algorithms," 978-1-6654-0741-0/22/$31.00 ©2022 IEEE, https://www.researchgate.net/publication/359964811_A_Comparative_Study_of_Predicting_Students%27_Academic_Performance_Using_Classification_Algorithms?enrichId=rgreq-b5c4b78c687c517d8b646bae15a7c134-XXX&enrichSource=Y292ZXJQYWdlOzM1OTk2NDgxMTtBUzoxMTQ2NjIzNTg1NDYwMjI3QDE2NTAzODc3NzQ2NzY%3D&el=1_x_2&_esc=publicationCoverPdf
10. Thaher, T.; Zaguia, A.; Al Azwari, S.; Mafarja, M.; Chantar, H.; Abuhamdah, A.; Turabieh, H.; Mirjalili, S.; Sheta, A., "An Enhanced Evolutionary Student Performance Prediction Model Using Whale Optimization Algorithm Boosted with Sine-Cosine Mechanism," Appl. Sci. 2021, 11, 10237. https://doi.org/ 10.3390/app112110237 https://doi.org/10.3390/app112110237
11. Zafari, M., Sadeghi-Niaraki, A.; Choi, S.-M.; Esmaeily, A., "A Practical Model for the Evaluation of High School Student Performance Based on Machine Learnin,". Appl. Sci. 2021, 11, 11534. https://doi.org/10.3390/app112311534 Academic Editor: Mayank Kej https://doi.org/10.3390/app112311534
12. Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A.A., Alsariera, Y.A., Ali, A. Q., Hashim, W., Tiong, S.K., "Toward Predicting Student's Academic Performance Using Artificial Neural Networks (ANNs).," Appl. Sci. 2022, 12, 1289, 2022, https://doi.org/10.3390/app12031289
13. Nida Aslam, Irfan Ullah Khan, Leena H. Alamri, Ranim S. Almuslim, "An Improved Early Student's Performance Prediction Using Deep Learning, International Journal of Emerging Technologies in Learning (iJET), 16(12), pp. 108–122. " https://doi.org/10.3991/ijet.v16i12.20699 , http://www.i-jet.org
14. Mukesh Kumar, Chetan Sharma, Shamneesh Sharma, Nidhi, Nazrul Islam, "Analysis of Feature Selection and Data Mining Techniques to Predict Student Academic Performance," 2022 International Conference on Decision Aid Sciences and Applications (DASA), https://www.researchgate.net/publication/359520060_Analysis_of_Feature_Selection_and_Data_Mining_Techniques_to_Predict_Student_Academic_Performance?enrichId=rgreq-928653ef18577a9777570376e2b471fb-XXX&enrichSource=Y292ZXJQYWdlOzM1OTUyMDA2MDtBUzoxMTM4ODY4NzQ5MzY5MzQ0QDE2NDg1Mzg4NzcwNDg%3D&el=1_x_2&_esc=publicationCoverPdf
15. Alboaneen, D., Almelihi, M., Alsubaie, R., Alghamdi, R., Alshehri, L., Alharthi, R., "Development of a Web-Based Prediction System for Students' Academic Performance. Data," 2022, Education Data Mining, 7, 21. https://doi.org/10.3390/ data7020021 https://doi.org/10.3390/data7020021
16. Ayon Roy, Md. Raqibur Rahman, Muhammad Nazrul Islam, Nafiz Imtiaz Saimon, M Aqib Alfaz, and Abdullah-Al-Sheak Jaber, "A Deep Learning Approach to Predict Academic Result and Recommend Study Plan for Improving Student's Academic Performance," International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS 2021) At: Erode, India. https://www.researchgate.net/publication/349735414_A_Deep_Learning_Approach_to_Predict_Academic_Result_and_Recommend_Study_Plan_for_Improving_Student%27s_Academic_Performance
17. Muhammad Sudais, Muhammad Sudais, Danish Asa, Student's Academic Performance Prediction – A Review, https://doi.org/10.21203/rs.3.rs-1292468/v1
18. Esmael Ahmed, "Student Performance Prediction Using Machine Learning Algorithms," Hindawi Applied Computational Intelligence and Soft Computing Volume 2024, Article ID 4067721, 15 pages https://doi.org/10.1155/2024/4067721
19. Shaik, I., Nittela, S. S., Hiwarkar, Dr. T., & Nalla, Dr. S. (2019). K-means Clustering Algorithm Based on E-Commerce Big Data. In International Journal of Innovative Technology and Exploring Engineering (Vol. 8, Issue 11, pp. 1910–1914). https://doi.org/10.35940/ijitee.k2121.0981119
20. Gupta, M. K., & Chandra, P. (2019). MP-K-Means: Modified Partition Based Cluster Initialization Method for K-Means Algorithm. In International Journal of Recent Technology and Engineering (IJRTE) (Vol. 8, Issue 4, pp. 1140–1148). https://doi.org/10.35940/ijrte.d6837.118419
21. Maheswari, Dr. K. (2019). Finding Best Possible Number of Clusters using K-Means Algorithm. In International Journal of Engineering and Advanced Technology (Vol. 9, Issue 1s4, pp. 533–538). https://doi.org/10.35940/ijeat.a1119.1291s419
22. Patravali, S. D., & Algur, Dr. S. P. (2023). COVID-19 Sentiment Analysis using K-Means and DBSCAN. In International Journal of Emerging Science and Engineering (Vol. 11, Issue 12, pp. 12–17). https://doi.org/10.35940/ijese.l2558.11111223
23. Jain, N., & Kumar, R. (2022). A Review on Machine Learning & It's Algorithms. In International Journal of Soft Computing and Engineering (Vol. 12, Issue 5, pp. 1–5). https://doi.org/10.35940/ijsce.e3583.1112522

AUTHORS PROFILE

Osondu E. Oguike, received a B.Sc. degree in Mathematics and Statistics, from the University of Lagos, Nigeria, in 1990, a PGD and M.Sc. degree in Information Technology from Queen Mary and Westfield College, University of London, in 1991, and a Ph.D. degree in Mathematical Modelling and Computational Mathematics from University of Nigeria, Nsukka, in 2018. From 2023 to 2024, he is a post-doctoral fellow at the Institute for Intelligent Systems, University of Johannesburg, South Africa. From 1999 to 2023, he assumed the following lectureship positions at the University of Nigeria, Nsukka (Lecturer II, Lecturer I, Senior Lecturer and Associate Professor). He is the author of three books and more than twenty articles. His research interest includes Artificial Intelligence, Machine Learning, Natural Language Processing. Dr. Osondu Oguike is a member of Nigeria Computer Society, Science Association of Nigeria, A recipient of the following awards: Best graduating student, Department of Mathematics, University of Lagos, Nigeria, 1988/89; ODA Shared Scholarship Scheme, Queen Mary and Westfield College, University of London, 1990/1991; Post doctoral fellowship award, Institute for Intelligent Systems, University of Johannesburg, South Africa, 2023-2024.

Emmanuel C. Ukekwe, has a B.Sc degree in Computer/Statistics, a Masters(Computer Programming) as well as a Ph.D degree in Computer Science all from the University of Nigeria, Nsukka. His Ph.D focused on application of Artificial Intelligence (Expert system) in improving Human Capititał Investment quality for productive labour market. Since then, he has developed a keen interest in the application of machine learning, Data science, expert system and general artificial intelligence in solving societal problems. He has also published his findings in some reputable journals and has contributed to research generally in the field. His current research aims at improving service delivery of network operators in Nigeria based on the locality subscription.

Elufidodo A. Gabriel, earned his Bachelor of Science degree in 2015 from Obafemi Awolowo University, located in Ile-Ife, Osun State, Nigeria. He is currently pursuing a Master of Science degree at the University of Nigeria. He serves as a Research Technologist in the Department of Computer Science at the University of Nigeria. His research interests encompass Data Science and Machine Learning, Deep Learning and Computer Vision, Data Mining, Natural Language Processing, and Predictive Modeling and Analytics. He is committed to contributing to these fields through both his academic and professional pursuits. He has published several papers in reputable journals and presented his research at various conferences, demonstrating his active engagement and contribution to the scientific community. In addition to his work in academia, he is an active member of the Data Scientist Network, formerly known as Data Science Nigeria. Through this membership, he collaborates with fellow professionals to advance the field of data science.

Retrieval Number: 100.1/ijsce.D36401404092
DOI: 10.35940/ijsce.D3640.14040924
Journal Website: www.ijsce.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)