

Predicting Students Academic Performance using Ridge, LASSO and Elastic Net with Efficient Model Selection: A Case Study of Federal University Gusau Nigeria

Kabir Bello, Olayemi Joshua Ibidoja and Yusuf Abubakar
 Department of Mathematics, Federal University Gusau, Gusau, Nigeria

Corresponding Author's Email: ojibidoja@fugusau.edu.ng

Received on: April, 2024

Revised and Accepted on: May, 2024

Published on: July, 2024

ABSTRACT

Academic establishments function in a very demanding and competitive environment. Most of the schools' face challenges in providing the students with a high-quality education, creating methods for assessing student achievement, analysing performance, and anticipating the needs of their students in the future. This research proposes a sparse regression model that solves the problem of multicollinearity among the six variables. The significant variables that determined the academic performance at graduation were identified, using ridge, LASSO and elastic net. Evaluation metrics and seven selection criteria were used to select the best model. The results show that the age of the student has no impact on the CGPA at graduation. For the training data, the elastic net using SGMASQ significantly suggests better performance in comparison to other models. For the testing data, the elastic net using SGMASQ significantly shows better results in comparison to other models including the training data.

Keywords: Machine learning, Sparse regression, Selection criteria, Test data, Modelling.

1.0 INTRODUCTION

The factors that determine the academic performance of students at the end of an academic program are crucial (Issah *et al.*, 2023a; Roy, 2023). Several variables, such as: educational background, gender, family history, state of origin, age, marital status, teaching staff, type of accommodation, etc., affect the academic performance of students (Banda *et al.*, 2023; Montero *et al.*, 2023; Vadivel *et al.*, 2023). In education, there is a growing amount of interest in predicting undergraduate academic success (Balogun *et al.*, 2020; Banda *et al.*, 2023; Ibidoja *et al.*, 2016; Roslan and Chen *et al.*, 2023). From the standpoint of the student, maintaining and predicting high academic performance boosts the chances of finding a job, because companies consider excellent academic achievement (Alsariera *et al.*, 2022).

Regression analysis has been used to predict the academic performance of students (Ibrahim & Rusli, 2007; Rajalaxmi *et al.*, 2019; Ying & Ma, 2024). A multiple regression model can be in this matrix form in equation (1):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1)$$

where \mathbf{y} represents an $n \times 1$ vector of response variables, the design matrix is \mathbf{X} of order $n \times p$, $\boldsymbol{\beta}$ means a $p \times 1$ vector of parameters and $\boldsymbol{\varepsilon}$ is a $n \times 1$ vector of the matrix which is identically and independently distributed errors. Ordinary Least Squares (OLS) is crucial to find the unknown parameters in the regression model. According to (Gujarati and Porter 2004; Ibidoja *et al.*, 2023; Obadina *et al.*, 2021), the (OLS) estimator of $\boldsymbol{\beta}$ in equation (2) is found as:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (2)$$

One of the problems that can happen is when the explanatory variables exceed the observations (Ibidoja *et al.*, 2023). Sparse regression is useful for regulating the coefficients of the regression and to overcome the problem of multicollinearity. This will lower the sampling error, reduce variance, and solve multicollinearity. Multicollinearity leads to overfitting, to solve this problem, ridge regression that shrinks the coefficients to non-zero values to prevent overfit, but keeps the variables., LASSO regression shrinks regression coefficients, with some shrunk to zero, this it also helps for feature selection. Elastic net is a mixture of ridge and LASSO. It has been used in many scientific disciplines (Ahrens *et al.*, 2020; Bottmer *et al.*, 2022; Friedman, 2012; Hirose *et al.*, 2013; Vincent & Hansen, 2014; Yu & Liu, 2016). Sparse regression is used to boost predictive accuracy in a high-dimensional circumstances (Belloni & Chernozhukov, 2011; Ibidoja *et al.*, 2024; Ma *et al.*, 2016; Pun & Wong, 2019; Vincent & Hansen, 2014; Wang & Yin, 2008; Zhang *et al.*, 2021).

In this research, a sparse regression model is proposed to model and predict the academic performance of students at Federal University Gusau, Nigeria, using seven efficient model selection criteria. Many studies have used features selection (Hussain and Khan 2023; Ibrahim and Rusli 2007; Issah *et al.*, 2023b; Lim *et al.*, 2024; Rajalaxmi *et al.*, 2019; Ravi *et al.*, 2024; Ying and Ma 2024; Zaffar *et al.*, 2018). But no study has been conducted to study the academic performance at Federal University Gusau, Nigeria. It is rare to find a study that incorporates the effect of age on the academic performance of students and the seven efficient models (Issah *et al.*, 2023c).

2.0 MATERIALS AND METHODS

Supervised machine learning algorithms such as: Ridge, LASSO and Elastic net will be used to determine the variables for predicting the academic performance of students. Evaluation metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Coefficient of Determination (RSQUARED), Sum of Squared Error (SSE), with seven efficient model selection criteria will be used for validation and model selection.

For the data partition, a set seed was used to obtain a reproducible result, and two independent samples with sampling with replacement. The data were divided into 70% and 30% for the train and test respectively. After the partitioning, the train data has 877 and the test 435 observations. *TrainControl* function within the caret package was used to create the custom control parameters for the cross validation in the RStudio. For consistency and reproducible results, the same set seed

was used for developing all the models. The cross validation used 10 folds, it was repeated 5 times, the training data was divided into 10 parts, each time it used 9 parts to create the model and 10 parts was used for testing the error and repeated 5 parts. Hence, the RMSE was based on the data that was not used for creating the model, for example, in the summary of the sample sizes of 789, out of 877, the remaining data points were used for calculation the RMSE.

2.1 Data Description

The data were collected from 23rd May, 2024 to 30th April, 2024, between the hours of 8:00 am to 5:00 pm at Federal University Gusau, Nigeria. The data contains the age and Cumulative Grade Point Average (CGPA) of the graduating students in 2021/2022 academic session. It comprises a total of 6 variables of 1312 observation, with 5 independent variables and one dependent variable. Table 1 shows the 6 variables with the relevant information.

Table 1: Representation of Variables

Symbol	Variable	Meaning
x_1	Dependent Variable	Age at entry to the University
x_2	Dependent Variable	CGPA at the end of first year
x_3	Dependent Variable	CGPA at the end of second year
x_4	Dependent Variable	CGPA at the end of third year
x_5	Dependent Variable	CGPA at the end of fourth year
y	Independent Variable	CGPA at graduation

3.0 RESULTS AND DISCUSSION

In this research, the assumptions of linear regression are checked. Figure 1 displays the scatter plot of every possible combination of the independent variables with the correlation coefficients, when the variables are extremely correlated, it causes multicollinearity problems, hence, the prediction using the regression model may not be precise since the estimates are unstable because of the multicollinearity problems.

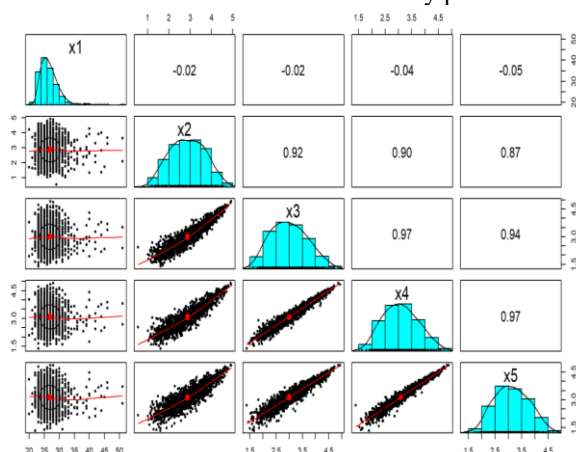


Figure 1: Scatter plot of all the independent variables

Figure 2 reveals the coefficients vs log lambda plot for ridge, on the x-axis, there is log lambda, and the coefficients on the y-axis, as the log lambda is about 6.3 or more, all the coefficients are zero. As the lambda is relaxed, the coefficients grow, and the sum of the squares of the coefficients becomes larger and larger. At the top, it means that at every time, there are 5 independent variables in the model, so increasing the lambda will not make the coefficient grow, so increasing the lambda helps to reduce the size of the coefficients, it does not make the variables that are not contributing to be zero. Figure 3 is the plot of variable importance with ridge, it shows that x_4 is the most important variable to predict academic performance at graduation, followed by x_5 , x_3 and x_2 . The variable x_1 is not important and has no contribution to predicting the academic performance of the students at graduation.

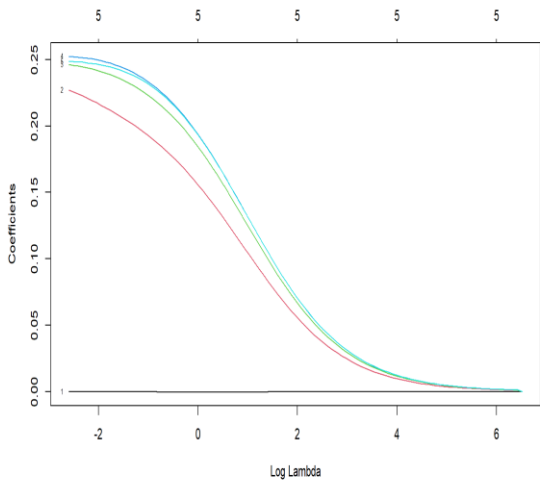


Figure 2: Coefficients vs log lambda plot for ridge

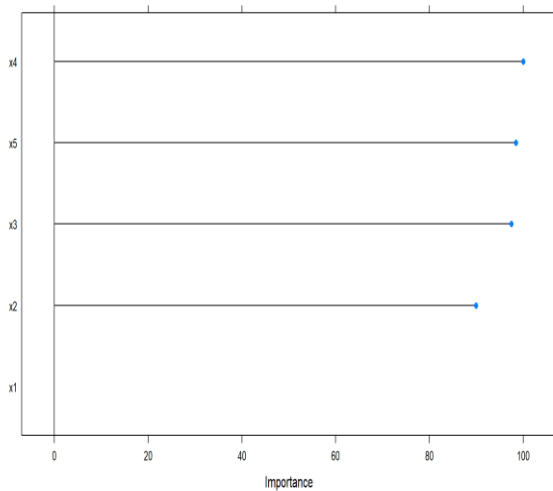


Figure 3: Plot of variable importance with ridge

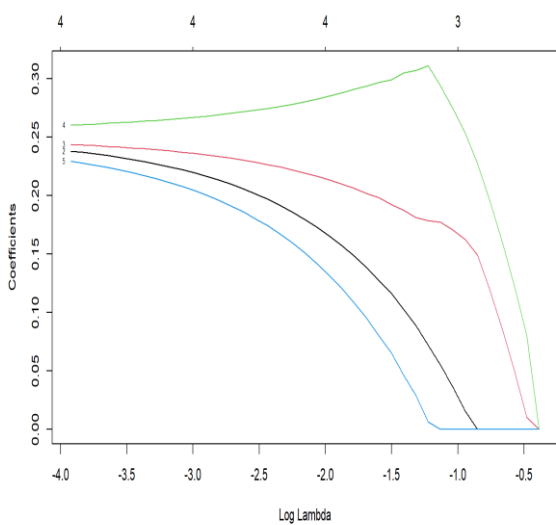


Figure 4: Coefficients vs log lambda plot for LASSO

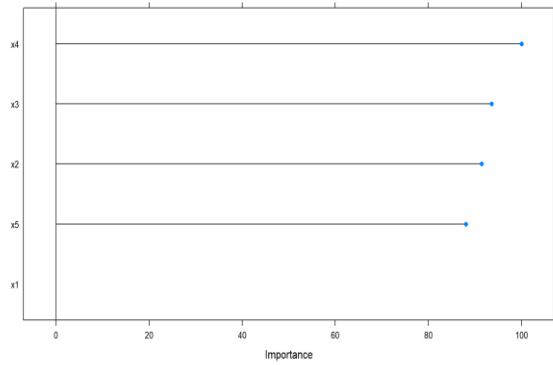


Figure 5: Plot of Variable Importance with LASSO

Figure 4 shows the coefficients vs log lambda plot for LASSO, on the x-axis, there is log lambda, and the coefficients on the y-axis, as the log lambda is -0.5, all the coefficients are zero. As the lambda is relaxed, the coefficients to grow and the sum of squares of the coefficient becomes larger and larger, at the top. Figure 5 is the plot of variable importance with LASSO, it reveals that x_4 is the most important variable to predict the academic performance of the students at graduation, followed by x_3 , x_2 and x_5 . The variable x_1 is not important and has no contribution to forecasting the academic performance of the students at graduation.

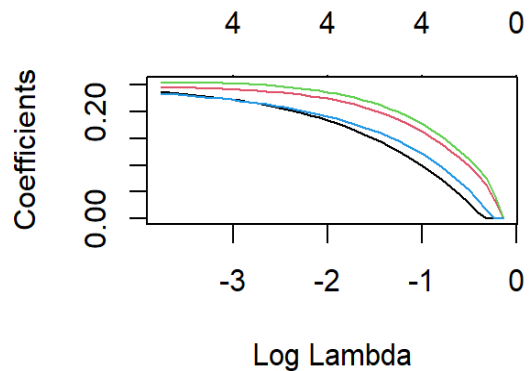


Figure 6: Coefficients vs log lambda plot for elastic net

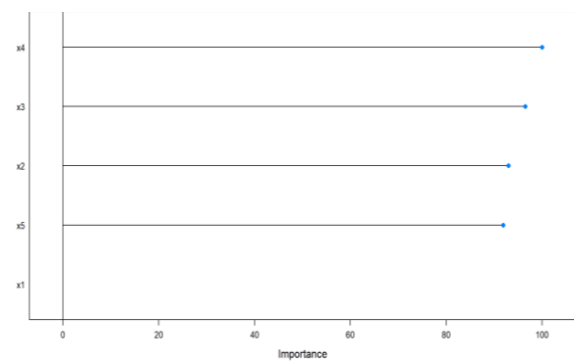


Figure 7: Plot of variable importance with elastic net

Figure 6 shows the coefficients vs log lambda plot for elastic net, on the x-axis, there is log lambda, and the coefficients on the y-axis, as the log lambda is -0.3, all the coefficients are zero, as the lambda is relaxed, the coefficients to grow and the sum of squares of the coefficient becomes larger and larger, at the top. Figure 7 is the plot of variable importance with elastic net, it shows that x_4 is the most important variable to predict the academic performance at graduation, follow by x_3, x_2 and x_5 . The variable x_1 is not important and has no contribution for forecasting the academic performance of the undergraduates at graduation.

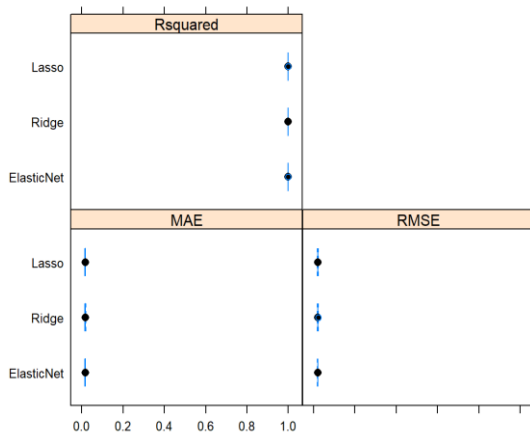


Figure 8: Box Plot of the results of the models

Figure 8 shows the box plots of ridge, LASSO and elastic net using the RSQUARED, Mean Absolute Error (MAE) and RMSE. The differences in the models are small, it is not easy to know the ones that are substantially superior from each other. Figure 9 shows the scatter plot between ridge and LASSO, the points above the line mean that the models achieve better with LASSO, the points below the line mean that the models achieve better with ridge. Figure 10 shows the scatter plot between ridge and elastic net, the points above the line means that the models achieve better with elastic net, the points below the line means that the models achieve better with ridge. Figure 11 shows the scatter plot between LASSO and elastic net, the points above the line mean that the models perform better with elastic net, there is no point below the line, it means that the models did not perform with LASSO.

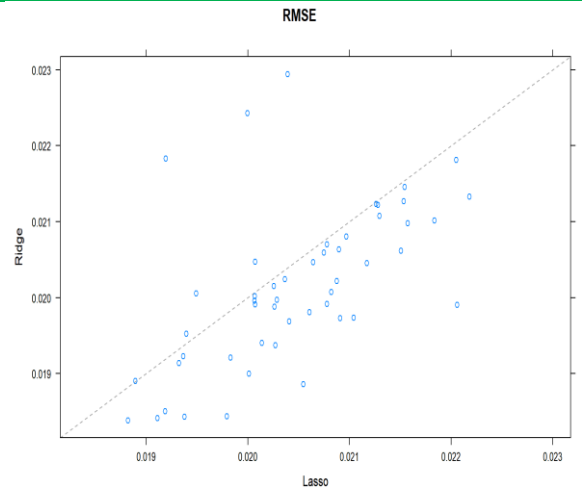


Figure 9: Scatter plot between ridge and LASSO

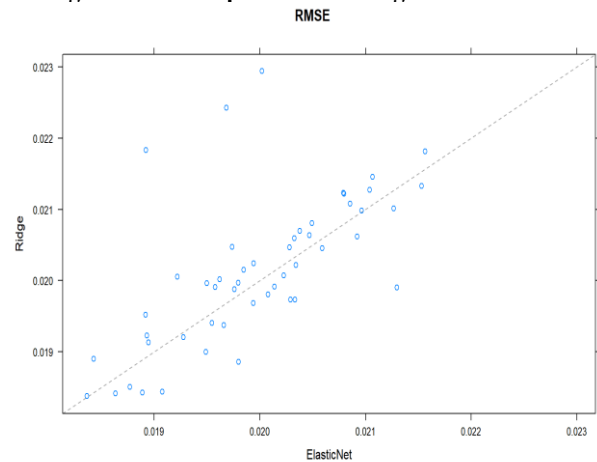


Figure 10: Scatter plot between ridge and elastic net

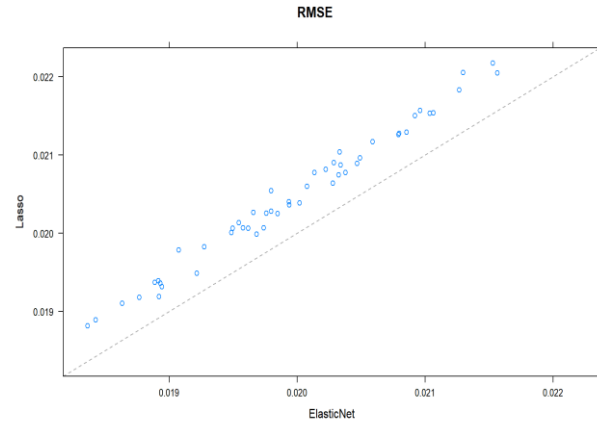


Figure 11: Scatter plot between LASSO and elastic net

Table 2: Results for the evaluation metrics for the models

	MSE	RMSE	RSQUARED	SSE
Training Phase (70 % of the data, n= 877)				
Ridge	0.0004024048	0.02006003	0.9991548	0.352909
LASSO	0.0004191825	0.02047395	0.9991196	0.3676231
Elastic Net	0.0003990938	0.01997733	0.9991618	0.3500052
Testing Phase (30 % of the data, n=435)				
Ridge	0.0003774662	0.01942849	0.9991432	0.1641978
LASSO	0.0003993539	0.01998384	0.9990935	0.173719
Elastic Net	0.0003785228	0.01945566	0.9991408	0.1646574

Table 2 presents the assessment metrics such as: MSE, RMSE, RSQUARED, and SSE for the training and testing phases. For the training data, the MSE for ridge, LASSO and elastic net are 0.0004024048, 0.0004191825 and 0.0003990938 respectively. The lowest MSE of the elastic net with 0.0003990938 represents the most efficient model. This is because the lower the MSE, the better the prediction accuracy. For the RSQUARED, the ridge, LASSO and elastic net have 0.9991548, 0.9991196 and 0.9991618 respectively. The higher the RSQUARED value, the better the model. In addition, assessment metrics with MSE (0.0003990938), RMSE (0.01997733), RQUARED (0.9991618) and SSE (0.3500052), indicate that significantly improved results are found by elastic net for the training data for forecasting the academic performance of the students. This also

confirms the results of (Baradaran *et al.*, 2023; Ibidoja *et al.*, 2023b; Ogutu *et al.*, 2012; Zou & Hastie, 2005) where the elastic net absolutely performed better than the other methods. If the error assessment metrics are small, it is high prediction accuracy (Ibidoja *et al.*, 2023; Sumari *et al.*, 2021). The value of RSQUARED (0.9991618) implies that the selected variables can explain 99.92% of the dependent variable (CGPA at graduation).

For the testing data, the MSE for ridge, LASSO and elastic net are 0.0003774662, 0.0003993539, and 0.0003785228 respectively. The lowest MSE of the ridge with 0.0003774662 represents the most efficient model. The value R-square (0.9991432) implies that the selected variables can explain 99.91% of the dependent variable (CGPA at graduation).

Table 3: Results for 7 Selection Criteria for the Models

	AIC	RICE	GCV	SCHWARZ	SGMASQ	SHIBATA	FPE
Training Phase (70 % of the data)							
Ridge	0.00040701	0.00040704	0.00040703	0.00041825	0.00040471	0.00040699	0.00040701
LASO	0.00042398	0.00042401	0.00042400	0.00043569	0.00042158	0.00042396	0.00042398
Elastic Net	0.00040367	0.00040369	0.00040368	0.00041481	0.00040138	0.00040364	0.00040367
Testing Phase (30% of the data)							
Ridge	0.00083015	0.00083037	0.00083026	0.00086996	0.00082071	0.00082993	0.00083015
LASO	0.00040864	0.00040875	0.00040869	0.00042823	0.00040399	0.00040853	0.00040864
Elastic Net	0.00038732	0.00038742	0.00038737	0.00040590	0.00038292	0.00038722	0.00038732

AIC= Akaike Information Criterion, RICE = Rice Information Criteria, GCV= Generalized Cross-Validation, SGMASQ= Sigma Square, FPE= Final Prediction Error

Table 3 presents the seven selection criteria (7SC) such as AIC, RICE, GCV, SCHWARZ, SGMASQ, SHIBATA, and FPE. For the training data, elastic net has the lowest 7SC. For the training data, elastic net has the lowest 7SC when compared to ridge and LASSO. The lowest 7SC represents the most efficient model for modelling and predicting the academic performance of the students at graduation. Hence, elastic net using SGMASQ significantly indicates better results in comparison to others. For the testing data, elastic net also has the lowest 7SC. The lowest 7SC represents the most efficient model for modelling and predicting the academic performance of the students at graduation. The elastic net using SGMASQ significantly indicates better results in comparison to others.

4.0 CONCLUSION AND FUTURE WORK

The significant variables that determined the academic performance at graduation were identified, using ridge, LASSO and elastic net. These models are valuable for determining the important variables to overcome the problem of multicollinearity. The age of the student has no impact on the CGPA at graduation. For the training data, the elastic net using SGMASQ significantly indicates better performance in comparison to other

models and selection criteria. For the testing data, the elastic net using SGMASQ significantly indicates better results in comparison to others.

For further studies, the effects of interaction and higher order interaction among the variables can be investigated. The variables such as gender, educational background, state of origin and ethnicity can be investigated in conjunction with robust estimators.

ACKNOWLEDGEMENTS

This research was sponsored by TETFUND under the grant allocation number ‘‘TETF/DR&D/CE/UNI/GUSAU/IBR/2024/VOL.1’’ for their financial support. The authors are immensely grateful to Federal University Gusau Nigeria for providing the environment for this research.

AUTHOR CONTRIBUTIONS

KB: He contributed to the writing, logic and editing. OJI: He conducted the analysis, designed the experiment, logic, writing and manuscript development. YA: He contributed to the manuscript writing. All the authors approved the final manuscript.

REFERENCES

- Ahrens, A., Hansen, C. B., & Schaffer, M. E. (2020). lassopack: Model selection and prediction with regularized regression in Stata. *Stata Journal*, 20(1), 176–235. <https://doi.org/10.1177/1536867X20909697>
- Alsariera, Y. A., Baashar, Y., Alkaws, G., Mustafa, A., Alkahtani, A. A., & Ali, N. (2022). Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance. In *Computational Intelligence and Neuroscience* (Vol. 2022). Hindawi Limited. <https://doi.org/10.1155/2022/4151487>.
- Balogun, O. S., Olaleye, A. S., & Ibidoja, O. J. (2020). Academic Performance: An Approach of Tau Statistic. *International Business Information Management Association (IBIMA)*, 6169–6181. <https://digitalcommons.unl.edu/libphilprac>.
- Banda, L. O. L., Liu, J., Banda, J. T., & Zhou, W. (2023). Impact of ethnic identity and geographical home location on student academic performance. *Heliyon*, 9(6). <https://doi.org/10.1016/j.heliyon.2023.e16767>
- Baradaran Rezaei, H., Amjadian, A., Sebt, M. V., Askari, R., & Gharaei, A. (2023). An ensemble method of the machine learning to prognosticate the gastric cancer. *Annals of Operations Research*, 328(1), 151–192.
- Belloni, A., & Chernozhukov, V. (2011). High dimensional sparse econometric models: An Introduction Springer Berlin Heidelberg.
- Bottmer, L., Croux, C., & Wilms, I. (2022). Sparse regression for large data sets with outliers. *European Journal of Operational Research*, 297(2), 782–794. <https://doi.org/10.1016/j.ejor.2021.05.049>
- Friedman, J. H. (2012). Fast sparse regression and classification. *International Journal of Forecasting*, 28(3), 722–738. <https://doi.org/10.1016/j.ijforecast.2012.05.001>
- Gujarati, D. N., & Porter, D. N. (2004). *Basic econometrics* (4th ed.). The McGraw-Hill Companies.
- Hirose, K., Tateishi, S., & Konishi, S. (2013). Tuning parameter selection in sparse regression modeling. *Computational Statistics and Data Analysis*, 59(1), 28–40. <https://doi.org/10.1016/j.csda.2012.10.005>
- Hussain, S., & Khan, M. Q. (2023). Student-Performer: Predicting Students' Academic Performance at Secondary and Intermediate Level Using Machine Learning. *Annals of Data Science*, 10(3), 637–655. <https://doi.org/10.1007/s40745-021-00341-0>
- Ibidoja, O. J., Shan, F. P., Sulaiman, J., & Ali, M. K. M. (2023a). Robust M-Estimators and Machine Learning Algorithms for Improving the Predictive Accuracy of Seaweed Contaminated Big Data. *J. Nig. Soc. Phys. Sci*, 5, 1137. <https://doi.org/10.46481/jnsps.2022.1137>
- Ibidoja, O. J., Ajare, E. O., & Jolayemi, E. T. (2016). Reliability Measures of Academic Performance. *IJS GS*, 2(4), 59–64.
- Ibidoja, O. J., Shan, F. P., Suheri, M. E., Sulaiman, J., & Ali, M. K. M. (2023). Intelligence System via Machine Learning Algorithms in Detecting the Moisture Content Removal Parameters of Seaweed Big Data. *Pertanika Journal of Science and Technology*, 31(6), 2783–2803. <https://doi.org/10.47836/pjst.31.6.09>
- Ibidoja, O. J., Shan, F. P., & Ali, M. K. M. (2024). Modified sparse regression to solve heterogeneity and hybrid models for increasing the prediction accuracy of seaweed big data with outliers. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-60612-7>
- Ibidoja, O. J., Shan, F. P., Sulaiman, J., & Ali, M. K. M. (2023b). Detecting heterogeneity parameters and hybrid models for precision farming. *Journal of Big Data*, 10(1). <https://doi.org/10.1186/s40537-023-00810-8>
- Ibrahim, Z., & Rusli, D. (2007). Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree and Linear Regression. 21st Annual SAS Malaysia Forum, 1–6. ht
- Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023a). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. In *Decision Analytics Journal* (Vol. 7). Elsevier Inc. <https://doi.org/10.1016/j.dajour.2023.100204>
- Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023b). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. In *Decision Analytics Journal* (Vol. 7). Elsevier Inc. <https://doi.org/10.1016/j.dajour.2023.100204>
- Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023c). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. In *Decision Analytics Journal* (Vol. 7). Elsevier Inc. <https://doi.org/10.1016/j.dajour.2023.100204>
- Lim, L. M., Sathasivam, S., Ismail, M. T., Mohamed, A. S. A., Ibidoja, O. J., & Ali, M. K. M. (2024a). Comparison Using Intelligent Systems for Data Prediction and Near Miss Detection Techniques. *Pertanika Journal of Science and Technology*, 32(1), 365–394. <https://doi.org/10.47836/pjst.32.1.20>
- Lim, L. M., Sathasivam, S., Ismail, M. T., Mohamed, A. S. A., Ibidoja, O. J., & Ali, M. K. M. (2024b). Comparison Using Intelligent Systems for Data Prediction and Near Miss Detection Techniques. *Pertanika Journal of Science and Technology*, 32(1), 365–394. <https://doi.org/10.47836/pjst.32.1.20>
- Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra- and inter-

- category promotional information. *European Journal of Operational Research*, 249(1), 245–257. <https://doi.org/10.1016/j.ejor.2015.08.029>
- Montero, L., Mejia-Dorantes, L., & Barceló, J. (2023). The role of life course and gender in mobility patterns: a spatiotemporal sequence analysis in Barcelona. *European Transport Research Review*, 15(1), 44. <https://doi.org/10.1186/s12544-023-00621-1>
- Obadina, O. G., Adedotun, A. F., & Odusanya, O. A. (2021). Ridge Estimation's Effectiveness for Multiple Linear Regression with Multicollinearity: An Investigation Using Monte-Carlo Simulations. *Journal of the Nigerian Society of Physical Sciences*, 3(4), 278–281. <https://doi.org/10.46481/jnsps.2021.304>
- Ogutu, J. O., Schulz-Streeck, T., & Piepho, H. P. (2012). Genomic selection using regularized linear regression models: Ridge regression, lasso, elastic net and their extensions. *BMC Proceedings*, 6(SUPPL. 2). <https://doi.org/10.1186/1753-6561-6-S2-S10>
- Pun, C. S., & Wong, H. Y. (2019). A linear programming model for selection of sparse high-dimensional multiperiod portfolios. *European Journal of Operational Research*, 273(2), 754–771. <https://doi.org/10.1016/j.ejor.2018.08.025>
- Parkavi, R., Karthikeyam, P., & Sheikh, A. A. (2024). Predicting academic performance of learners with the three domains of learning data using neuro-fuzzy model and machine learning algorithms. *Journal of Engineering Research (Kuwait)*. <https://doi.org/10.1016/j.jer.2023.09.006>
- Rajalaxmi, R. R., Natesan, P., Krishnamoorthy, N., & Ponni, S. (2019). Regression Model for Predicting Engineering Students Academic Performance. *International Journal of Recent Technology and Engineering (IJRTE)*, 7(6S3), 71–75. <https://www.researchgate.net/publication/333102911>
- Kumar, P. R., Ali, M. K. M., & Ibiyoja, O. J. (2024). Identifying heterogeneity for increasing the prediction accuracy of machine learning models. *Journal of the Nigerian Society of Physical Sciences*, 2058. <https://doi.org/10.46481/jnsps.2024.2058>
- Roslan, M. H. Bin, & Chen, C. J. (2023). Predicting students' performance in English and Mathematics using data mining techniques. *Education and Information Technologies*, 28(2), 1427–1453. <https://doi.org/10.1007/s10639-022-11259-2>
- Roy, S. K. (2023). YouTube's influential factors for academic achievement: A two-stage approach. *Telematics and Informatics Reports*, 10. <https://doi.org/10.1016/j.teler.2023.100060>
- Sumari, A. D. W., Charlinawati, D. S., & Ariyanto, Y. (2021). A Simple Approach using Statistical-based Machine Learning to Predict the Weapon System Operational Readiness. *The 1st International Conference on Data Science and Official Statistics*, 343–351.
- Vadivel, B., Alam, S., Nikpoo, I., & Ajanil, B. (2023). The Impact of Low Socioeconomic Background on a Child's Educational Achievements. *Education Research International*, 2023. <https://doi.org/10.1155/2023/6565088>
- Vincent, M., & Hansen, N. R. (2014). Sparse group lasso and high dimensional multinomial classification. *Computational Statistics and Data Analysis*, 71, 771–786. <https://doi.org/10.1016/j.csda.2013.06.004>
- Wang, Q., & Yin, X. (2008). A nonlinear multi-dimensional variable selection method for high dimensional data: Sparse MAVE. *Computational Statistics and Data Analysis*, 52(9), 4512–4520. <https://doi.org/10.1016/j.csda.2008.03.003>
- Ying, D., & Ma, J. (2024). Student Performance Prediction with Regression Approach and Data Generation. *Applied Sciences (Switzerland)*, 14(3). <https://doi.org/10.3390/app14031148>
- Yu, G., & Liu, Y. (2016). Sparse Regression Incorporating Graphical Structure Among Predictors. *Journal of the American Statistical Association*, 111(514), 707–720. <https://doi.org/10.1080/01621459.2015.1034319>
- Zaffar, M., Hashmani, M. A., Savita, K. S., & Rizvi, S. S. H. (2018). A study of feature selection algorithms for predicting students academic performance. *International Journal of Advanced Computer Science and Applications*, 9(5), 541–549. <https://doi.org/10.14569/IJACSA.2018.090569>
- Zhang, Y., Zhu, R., Chen, Z., Gao, J., & Xia, D. (2021). Evaluating and selecting features via information theoretic lower bounds of feature inner correlations for high-dimensional data. *European Journal of Operational Research*, 290(1), 235–247. <https://doi.org/10.1016/j.ejor.2020.09.028>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Statist. Soc. B*, 67(2), 301–320.