On The Comparison of Artificial Neural Network (ANN) and Multinomial Logistic Regression (MLR)

O. C. Asogwa[1] and A. V. Oladugba[2]

[1]Department of Mathematics, Computer Science, Statistics and Informatics,
Federal University Ndufu-Alike Ikwo, Ebonyi State..qackasoo@yahoo.com
[2]Department of Statistics, University of Nigeria, Nsukka. [2]bolvic2003@yahoo.com

Abstract

This work presented the results of an experimental comparison of two models: Multinomial Logistic Regression (MLR) and Artificial Neural Network (ANN) for classifying students based on their academic performance. The predictive accuracy for each model was measured by their average Classification Correct Rate (CCR). The Average Classification Correct Rate for the Artificial Neural Network was higher than Multinomial Logistic Regression.

Keywords: Multinomial Logistic Regression, Artificial Neural Network, Correct classification rate.

Introduction

Prediction of student academic performance is necessary in educational institutions, especially when admission officers are considering applications, accurate predictions help them to distinguish between suitable and unsuitable candidates for an academic program and identify candidates who would likely do well in the university. The failure to perform an accurate admission decision may result in an unsuitable student being admitted to the university. The quality of an educational institution is mainly reflected in its research and training. Hence, the quality of admitted candidates affects the quality level of institutions.

In addition, the results obtained from the prediction of academic performance may be also used for classifying students, which enables educational instructors to offer them additional support such as customized assistance and tutoring resources. The results of prediction can also be used by instructors to specify the most suitable teaching actions for each group of students and provide them with further assistance tailored to their needs. Accurate prediction of student achievement is one way to enhance the quality level and provide better educational services.

However, neural network model is originated from the mathematical neurobiology and the commonest practical models have been used in a wide variety of applications where statistical methods are traditionally used. Many literatures have pointed out the potentials of neural networks for prediction and classification problems. This can be due to the fact that these networks have attempts to model the capabilities of human brain. Neural network and traditional methods have been applied as a comparative model performance measure in solving different real life problems ranging from management sciences, business schools, and others [10], [12], [14], [17].

Moreover, this study aims at comparisons of the model performance of neural network and statistical technique (Multinomial Logistic Regression) in view of other objectives, using secondary data from the department of pharmaceutical science, University of Ibadan, to predict the academic performance of students at their penultimate year. Penultmate year is considered in this work because it acts as a key element in selecting the appropriate possible graduates. Over the past several years, there has been an increase in the number of students that either repeat class or be demoted at their

penultimate year and hence it is important to adequately and carefully study those possible factors that might likely contribute to the increased rate of the failure. These factors are enumerated in the methodology.

In view of this, many authors have worked on the academic performance prediction in educational institutions, especially that of the higher institutions of learning. [1] Used bivariate regression models to examine relationships between undergraduate GPA and a number of factors and found out that undergraduate GPA and PUTME were significant variables in predicting academic success. [2] showed that academic gain and learning performance in the University is affected by numerous factors: gender, age, etc. [16] compared the results of Artificial Neural Network model to that of a logistic regression analysis for modeling student enrollment decision making. It was shown that the improvement gained by using artificial neural network is significantly higher than that of the other model.

In this study, a student is classified in one category if he or she is promoted to final year and to the other groups if he or she is not promoted or demoted. The criterion for the promotion is the student's penultimate year GPA. Promoted category has GPA of 3.5 and above in 5-point scale while repeated class has penultimate year GPA between 2.5 to 3.4 and the demoted class have GPA below 2.5.

Artificial Neural Network (ANN):

Artificial neural network (ANN) commonly called neural network is a system composed of many single processing elements operating in parallel, whose function is determined by network structure, connection strength and the processing performed at computing elements or nodes [6]. Artificial Neural Network is considered nonlinear statistical data modeling tools where the complex relationships between inputs and outputs are modeled. It has three layers that are interconnected. The first layer which consist of the input layer. The intermediate layer called the hidden units (layer) or summing junction and the third layer which is the output layer. The hidden neuron receives data from the input neuron, works on it with the help of the transfer function and sends it to the output layer as the response.

The model proposed by [3] was used. The model is given as:

$$y = f(X,w) = \alpha X + \sum_{h=1}^{H} \beta_h \, g\left(\sum_{i=0}^{I} \gamma_{hi} x_i\right) \qquad (1)$$

where the transfer function, $g = (1 + e^{-x})^{-1}$

$$\mathbf{X}=(x_0 = 0,\ x_1, \ldots, x_I)\ ; \quad w = (\alpha, \beta, \gamma)$$

y = the output variable; x = the input variables;
α=the weight of the input unit(s); β= the weight of the hidden unit(s),
γ= the weight of the output unit(s)

Multinomial Logistic Regression:

Multinomial Logistic regression is a classification method that generalizes logistic regression to multi-class problems. (i.e with more than two possible discrete outcomes), [8]. Moreover, multinomial logistic regression is also a linear regression analysis to conduct when the dependent variables are categorical with more than two levels. Thus, it is an extension of logistic regression, which analyses dichotomous (binary) dependents. However, Multinomial Logistic Regression does not make any serious assumption about the distribution of the predictor variables. Considering a collection of k independent variables which will be denoted by a row vector $x = (x_1, x_2, \ldots, x_k)$. Then the log of the Multinomial Logistic Regression model is given in equation 1 below in form of linear function.

$$log(p_{ij} / p_{i1}) = \beta_{0j} + \beta_{1j} x_{i1} + \beta_{2j} x_{i2} + \cdots + \beta_{pj} x_{ip} \qquad (2)$$

where i is the i^{th} individual and j is the j^{th} category of the dependent variables. It is necessary to make one of the categories the baseline category (i=1). From equation 1, β_{0j}= the intercept.

β_{ij}'s, are the coefficients associated with the explanatory variables $x_{i1}, x_{i2}, \ldots, x_{ip}$

Materials and Methods

Through extensive review of literatures, a number of pre-admission and socio-economic factors which influence student's academic

performance were identified. Some of these factors can be seen in [11], [7], [4] and others.

The data used in this work are secondary data and were collected from faculty of pharmaceutical Sciences, University of Ibadan, Oyo state. The data were the extraction of the student's pre-admission and some socio-economic factors contained in the student's file in their respective departments. Also, the student's penultimate year GPA was collected from the departmental examination officers. A total sample of 150 student's files were drawn using simple random sampling, after which 120 were considered in this study. Files with incomplete information were ignored. Those factors which were extracted from the student's file in the departments are shown in the appendix.

The output variable is the student's performances, while those considered pre-admission and socio-economic factors are considered the input variables. The output variable will be multi-categorized. Students with penultimate year GPA above 3.5 in 5-point scale were promoted to final year (possible graduants); those students with GPA below 2.5 were demoted to immediate lower class, while those students between 2.5 to 3.4 repeated a class. This restriction is to enable the students graduate well. The penultimate year is a year to the final year.

A multi-layer perception algorithm which was trained with back propagation with a logistic transfer function was programmed. This programmed model was compared with Multinomial Logistic Regression model based on some of the existing similarities between the models. A one unit neuron was used in the hidden layer in other to maintain the similarities. Mean Correct Classification Rate (CCR) was used as a measure of comparison.

Artificial Neural Network Model Architecture:

In this section the Neural Network technique is applied to the data. A three-layer feed forward network was considered in this analysis. The input layer contained eighteen (18) neurons. Based upon recommendations from [5] and [9] that one-hidden-layer network is sufficient to model any complex system, the designed network model will have only one hidden layer. As to the issue of determining the optimal number of hidden nodes it is a crucial yet complicated one [13]. The best way to determine the number of middle layers and their sizes is trial-and-error. This can also be helpful in revealing the underlining relationships between variables. The rule of thumb is to start with the smallest size possible for a given problem to allow for generalization, then to increase the size of the middle layer(s), until the optimal results are achieved [18]. In this work, one hidden neuron in the hidden layer was used, as the network performance was best at this number and the output layer contained three (3) neurons, namely the classes of student, the promoted students, the repeated students and the demoted students. The back-propagation learning algorithm was used for training the network. The logistic activation function was used at the hidden layer and the identity activation function was used at the output layer. 60% of the data were used for training, 30% for testing and 10% for validation in this research. This analysis resulted in a Mean Correct Classification Rate for the test data. Each network was run for 1000 epochs.

Statistical Approach:

MATLAB was used as a tool to assist in the analysis by importing a variety of factors specific to the independent variables. These factors were; Sex, PUTME score, Penultimate year GPA, etc as seen in the appendix. The analysis included all independent variables into the equation by using stepwise regression analysis. The conducted stepwise analysis is a way to know the order of contribution of those factors to the model.

Results

Result of Multinomial Logistic Regression (MLR)

Independent variables or predictors were introduced into the model by a stepwise regression approach to determine the significant ordering of the variables to the model fit. It was observed that the order of the contribution is as follow: GPA, father's education, mother's education, O' level English, type of secondary school attended, parental marital status, location of secondary school attended, etc. These arrangements of the variable order of

importance to the model were evaluated using variable's individual normalized important value. However, considering the significance of the variables to the model, it can be concluded that the Student's rate of Performance or promotion outcomes in the aforementioned faculty is mainly determined by Student's GPA and others as follow: Father's Education, Mother's Education, O' Level English, etc.

Table 1: Classification Table for Multinomial Logistic Regression

Performances	**Promoted**	**Repeated**	**Demoted**
Promoted	60 (82.19%)	11 (15.07%)	2 (2.74%)
Repeated	5(22.72%)	15(68.18%)	2(9.09%)
Demoted	3(12.00%)	0(0.00%)	22(88.00%)
Mean Correct Classification Rate (CCR) = 80.83%			

Table1 above shows the percentage of accuracy in prediction and classification. The twelve independent variables in the predicted equation predicted classification membership correctly with an average of (80.83%). It can equally be depicted from the above Table that the model was able to predict correctly 60 students out of 73 students who were promoted. 15 students out of 22 students who were repeated were correctly predicted and also predicted correctly 22 students out of 25 students who were demoted. The diagonal entries show the correctly classification and the Mean Correct Classification Rate was calculated by taking the ratio of percentage of the correctly classifies to the total entries. It was represented by CCR and was estimated at 80.83%.

Result of Artificial Neural Network (ANN)

The appropriate model was developed with a back propagation algorithm, having three layers with a feed forward structure. A logistic transfer function was used in the hidden layer while an identity function was used in the output layer. The input layer had eighteen (18) neurons. The hidden layer had only one (1) unit and ten (10) neurons while the output layer had three (3) neurons. The network was partitioned using 60% for training, 30% for testing and 10% for validation. One thousand (1000) epochs was equally used in the model.

Table2: Classification Table for Artificial Neural Network.

Performances	**Promoted**	**Repeated**	**Demoted**
Promoted	71 (97.26%)	1 (1.37%)	1 (1.37%)
Repeated	1 (4.55%)	21(95.45%)	0(0.00%)
Demoted	1 (4.00%)	1(4.00%)	23(92.00%)
Mean Correct Classification Rate (CCR) = 95.83%			

Table2 above shows the percentage of accuracy in prediction by Artificial Neural Network. The prediction Table above predicted Classification group membership correctly with an average of (95.83%). It can be vividly seen from the above Table 2 that the model was able to predict correctly 71 students out of 73 students who were promoted. Twemtu-pme (21) students were also predicted out of twenty-two (22) students who repeated class correctly and also predicted correctly twenty-three (23) students out of twenty-five (25) students who were demoted. The diagonal entries show the correct classification and the Mean Correct Classification Rate was calculated by taking the ratio of percentage of the correctly classified to the total entries. It was represented by CCR and was estimated at 95.83%.

Comparison of Multinomial Logistic Regression and Artificial Neural Network

To compare the performance of the models: Artificial Neural Network and Multinomial Logistic Regression, the Mean or Average Correct Classification Rate (CCR) was the criterion used. The predictive accuracies for Multinomial Logistic Regression and Artificial Neural Network were shown in Tables 1 and 2, respectively. Moreover, it can be clearly seen that Artificial Neural Network model predictive accuracy is higher than that of Multinomial Logistic Regression; hence Artificial Neural Network performed better than the other models in predicting students' academic performance classification, judging from their Mean Correct Classification Rate (%).

Table 3: Table of model comparison

Model Comparison	Mean Correct Classification Rate CCR (%)
Multinomial Logistic Regression (MLR)	CCR_{MLR} = **80.83%**
Artificial Neural Network (ANN)	CCR_{ANN} = **95.83%**

Table 3 above shows the classification results for the Multinomial Logistic Regression approach and the Artificial Neural Network model as used in this research work. The Mean Correct Classification Rate for the Artificial Neural Network was higher than that of the Multinomial Logistic Regression. The best efficiency was 95.83% and 80.83% respectively.

Conclusion

The purpose of this work was to compare the prediction classification efficiency of two models in predicting student's academic performance rates. The result of the two compared models showed that Artificial Neural Network outperformed Multinomial Logistic Regression in terms of the Mean Correct Classification Rate which was (95.83%) against (80.83%).

However, some of the variables which have the most significant impact or contribution to the model fit were equally identified in this study. The order of the significance of the variables to the model has been listed earlier and these ordering concur with the results of research by [11], [7] and [15].

Acknowledgement

Faculty of Pharmaceutical Science, University of Ibadan was strongly acknowledged for making the data for this research available.

References

[1] Ahmadi, M., Raiszadeh, F., & Helms, M. (1997) An examination of admission criteria for the MBA program: A case study. *Education,* **177**, 540-546.

[2] Ali, S., Zubair, H., Fahad, M., Hamid, K., & Awais, A. (2013) Factors contributing to the students' academic performance: A case study of Islamic University Sub-Campus. *American Journal of Educational Research* **1** (8), 283-289

[3] Anders, U. (1996) Model selection in neural networks, ZEW *Discussion Papers 96-21*. Retrieved from http://hdl.handle.net/10419/29449

[4] Arsad, P. M., Buniyamin, N., Manan, J. L. A., & Hamzah, N. (2011) Proposed Academic Students' Performance prediction model. A Malaysian case study. *3rd International Congress on Engineering Education* (ICEED).

[5] Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function. *Journal Mathematics of Control, Signals, and Systems*, Vol. **2**, no. 4, pp. 303-314

[6] Defense Advanced Research Projects Agency (DARPA) (1988) *Neural Network Study.* AFCEA International Press, p. 60

[7] Diaz, A. L. (2003) Personal, family, and academic factors affecting low achievement in secondary school. *Electronic Journal of Research in Educational Psychology and psychopadology,* Vol. **1**. No 1, pp. 43-66.

[8] Greene, W. H.,(1993) *Econometric Analysis,* fifth edition, Prentice hall, 1993:720–723.

[9] Hornik, K., Stinchcombe, M. & White, H. (1989) Multilayer feed forward networks are universal approximators, *Neural Networks,* Vol. 2, No. 5, pp. 359-366

[10] Kumar, A., Rao, V. R., & Soni, H. (1995) An empirical comparison of neural network and logistic regression models. *Marketing letters*, **69**(4), 251 – 263

[11] Omidi, M. & Hedjazi, Y., (2008) Factors affecting the Academic success of Agricultural Students at university of Tehran, Iran. *Journal of Agricultural Science and Technology*. Vol. **10**. No. 3. Pp. 205-214, April 2008.

[12] Schumacher, M., Robner, R., & Vach, W. (1996) Neural networks & logistic regression: Part 1. *Computational Statistics and Data Analysis,* **21**(6), 661 – 682.

[13] Leea, T., Chiub, C., Chouc, Y., & Lu, C. (2006) "Mining the customer credit using classification and regression tree and multivariate adaptive regression splines," *Computational Statistics & Data Analysis*, Vol. 50, No. 4, pp. 1113 – 1130

[14] Tam, K. Y., & Kiang, M. Y. (1992) Managerial application of neural networks: The case of bank failure predictions. *Management Science*, **38** (7), 926-947.

[15] Thipsuda, W. & Pusadee, S. (2010) A comparison of classical Discriminant Analysis and Artificial Neural Network in predicting student graduation outcomes. *Proceedings of the Second International Conference of knowledge and Smart technologies*: 24-25.

[16] Walczak, S., & Sincich, T. (1999) A comparative analysis of regression and neural works for university admissions. *Information Sciences*, **119** (1-2), 1-20.

[17] Warner, B. & Misra, M. (1996) Understanding neural networks as statistical tools. *The American Statistician*, **50**(4), 284 – 293.

[18] Zhang, E.Y., & Trimble, P.E (1996) Forecasting Water Availability by Applying Neural networks with Global and Solar Indices. *Proceedings of the Sixteenth Annual American Geophysical Union*

APPENDIX A

Those factors that were considered in this research work are listed in the table below;

Appendix: Table containing both response and predictor variables considered in this work and the appropriate code

No.	Variables descriptions	Data type	Location	Code
1	Gender	Categorical	Input	1=male, 0=female
2	Parents Marital Status	Categorical	Input	0= Married and living together 1= Father die 2= Mother die 3= Mother & father die 4= Married but living apart because of occupation 5= Married but living apart because other reasons 6 = Divorced
3	Father's Education	Categorical	Input	0 = Uneducated 1= Primary school 2 = Secondary school 3 = Degree 4 = Vocational 5 = Diploma 6 = Masters degree 7= Doctors degree
4	Mother's Education	Categorical	Input	0 = Uneducated 1 = Primary school 2 = Secondary school 3 = Degree 4 = Vocational 5 = Diploma 6 = Masters degree 7 = Doctors degree
5	Fathers Occupation	Categorical	Input	0 = Government service(govt. officer) 1 = Government service(temporary) 2 = State enterprise employee 3 = Company employee 4 = Ownership 5 = Farmer, fisher

				6 = Freelance 7 = Unemployed 8 = Others
6	Mother's Occupation	Categorical	Input	0 = Government service(govt. officer) 1 = Government service(temporary) 2 = State enterprise employee 3 = Company employee 4 = Ownership 5 = Farmer, fisher 6 = Freelance 7 = Unemployed 8 = Others
7	O' Level Results	Categorical	Input	English: 0 = A1-A2 1 = A3-C4 2 = C5-C6
				Mathematics: 0 = A1-A2 1 = A3-C4 2 = C5-C6
				Chemistry: 0 = A1-A2 1 = A3-C4 2 = C5-C6
				Physics: 0 = A1-A2 1 = A3-C4 2 = C5-C6
				Biology: 0 = A1-A2 1 = A3-C4 2 = C5-C6
8	Age at Entry	Categorical	Input	0 = Below 23 years 1 = 23 years and above
9	Time delay before admission	Categorical	Input	0=1year 1=2years 2=3years and above
10	Type Of Secondary School Attended	Categorical	Input	0= Private 1= State 2= Federal
11	Location of Secondary School attended	Categorical	Input	0=Urban 1= Rural
12	Post UTME	Numerical	Input	True value
13	Physically impaired	Categorical	Input	0 = Disable 2 = Not disable
14	Performance outcome	Categorical	output	1 = Promoted students 2 = Repeating students 3 = Demoted students

.