Bulletin of Education and Research
December 2018, Vol. 40, No. 3 pp. 157-164

Prediction of Students' Academic Performance using Artificial Neural Network

Zahoor Ahmad[*] and Erum Shahzadi[**]

Abstract

Universities play a remarkable role in the development of a country by producing skilled graduates for the country. Graduation rate is low as compared to the enrollment rate in the higher education institutions. Academic failure is main reason for non-degree completion. Students' retention and high academic performance are significant for students, academic and administrative staff of universities. In this paper, our objective is to Predict the chance of students being at risk (AR) or not 'Not at risk' (NAR) with respect to their degree. Population of study consisted of all students of social sciences studying in 4th semester and they enrolled in 2007 session of BS and MA/MSc program at University of Gujrat Hafiz Hayat Campus. By using stratified sampling with proportional allocation method, a sample of 300 students was selected. We have used Multilayer Perception Neural Network Model to predict the chance of students being at risk (AR) or not 'Not at risk' (NAR) with respect to their degree on the basis of CGPA at the end of 2nd semester, Study time. Previous degree marks, Home environment, Study habits Learning skills, Hardworking and Academic interaction. In classifying the students at risk/not at risk, we could achieve a rate of correct classification of over 95% in training sample and over 85% in holdout sample. The estimated models can be used to predict the students being at risk or not with respect to their degree.

Keywords: Student's Retention, at risk, not at risk, Multilayer Perception, Neural Network.

[*]Department of Statistics, University of Gujrat. Email: zahoor.ahmed@uog.edu.pk
[**]Department of Statistics, University of Gujrat. Email: erum.shahzadi@uog.edu.com

Introduction

Universities have came under increasing performance assessment because they played a critical role in national development efforts. Enrollment in higher education institutions and the proportion graduates entering colleges and universities have been gradually increasing, graduation rates have not kept rapidity and have remained remarkably flat. Academic failure at the undergraduate level is one of the main reasons for non-degree completion and consequently needs to be addressed to decrease the drop-out percentages (Cheesman, Simpson, & Wint, 2006).

Several studies have been conducted on the prediction of academic performance of university students. Naik and Ragothaman (2004) carried out a study to predict MBA student success. This study demonstrated that the neural network model is useful tool for predicting MBA student performance. Many studies highlighted a series of factors that lead to the prediction of academic performance of university students. Vandamme, Meskens, and Superby, (2005) reported the factors that influence student performance and explored prediction methods. The main aim of this study was to classify students into three groups: 'low risk' students, 'medium risk' students, and 'high risk' students. Neural Network, Decision tree and Discriminant analysis was used to predict the achievement of freshmen. In this study discriminant analysis performed better than neural network.

Different researches have focused on comparing the performance of different methods for prediction of academic performance of university students. Ibrahim and Rusli (2007) conducted a study for predicting students' academic performance. Three predictive models had been developed namely Artificial Neural Network, Decision Tree and Linear regression. The result depicted that more than 80% accuracy was achieved by all of three models. This study concluded that performance of Artificial Neural Network was better than others.

Oladokun, Adebanjo, and Owaba-Chales, (2008) presented a study on predicting student academic performance. In this study an Artificial Neural Network (ANN) model was used for predicting the student performance as being the university student. Multilayer Perceptron Neural Network model was used and trained using data spanning five generations of graduates from engineering department of university of Ibadan. Results showed that the prediction accuracy of ANN model is 70%. Erimafa, Iduseri, and Edokpa, (2009) carried out a study to predict the class of degree obtained in university system by using discriminant analysis. By using stepwise approach we found two variables made significant independent and combined contribution. The linear discriminant function correctly predicted 87.5% of graduating student's class of degree. Through this function we get hit ratio of 88.2% when generalized, as a valid tool to

classify fresh students of unknown group membership. Prediction of academic performance is essential for students, academic and administrative staff of higher education institutions. Bhardwaj and Pal (2011) conducted a study to predict the academic performance of students by classification methods of data mining. Results concluded that mother qualification, students other habit, family annual income, students' family status are potential variables for prediction of academic performance.

Research Objective

1. To identify students who might be termed 'at risk' (AR) and 'Not at risk' (NAR) with respect to their degree.

The first group of students are those who are in danger of failing in a class, not getting the degree from the university. The second group of students are those that will get by using the factors that contribute significantly to their academic performance during the session. This paper employed Neural Network approach for data analysis purpose.

Research Methodology

In this section we present the methodology that we utilized to collect and analyze the data. Population of study consisted of all students of social sciences studying in 4^{th} semester and they are enrolled in 2007 session of BS and MA/MSc program of University of Gujrat Hafiz Hayat Campus. Our population consisted of 708 students. Sample of 300 students is selected by using stratified random sampling with proportional allocation. The sample size was determined by Yamane (1967) formula n = $N(1 + Ne^2)^{-1}$ using 5% margin of error.

The students of social sciences (Statistics, Sociology, CSIT, Business Administration, and English) are not homogeneous with respect to academic performance across disciplines and programs (BS and MA/MSc). We have used stratified random sampling with proportional allocation.

Instrument

Questionnaire is used for data collection. Questionnaire developed itself by the researchers. First part of the questionnaire is designed to obtain information on the demographic characteristics of university students, like gender, age, region, family system, profession of father of respondent. Next part designed to obtain information on some quantitative variables related to student performance. Then there are 39 items that consist of a combination of two categories nominal items, and 37 items using a 5-point Likert-Scale. Items are designed to assess six dimensions associated with student

academic performance. (Previous Achievements, Home Environment, Study Habits, Learning Skills, Hardworking, Academic Interaction).

Data Analysis

The Multilayer Perceptron Neural Network Model:

The following diagram illustrates a Perceptron network with three layers:

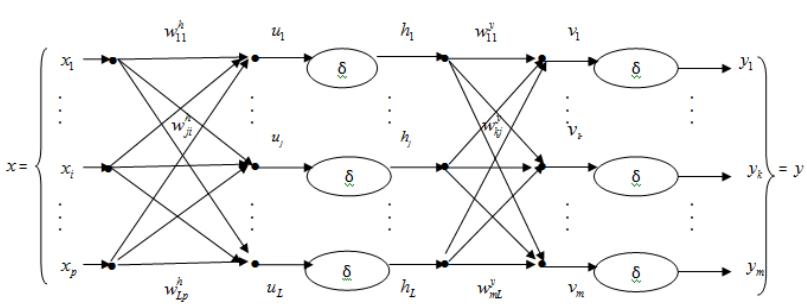

Figure 1: *Multilayer Perceptron Neural Network Architecture*

This network has an input layer (on the left) with three neurons, one hidden layer (in the middle) with three neurons and an output layer (on the right) with three neurons. There is one neuron in the input layer for each predictor variable ($x_1 \ldots x_p$).

Results

For achievement of first objective i.e. to predict the students being at risk or not with respect to their degree Firstly, we predict the chance of students being at risk or not at risk. The Multilayer Perceptron Neural Network Model is used and the network architecture is given in Figure 1 of Appendix. In which our dependent variable has two categories (at risk and not at risk) and independent variables are Study time, CGPA at the end of second semester, Previous degree marks Home environment, Study Habits, Learning skills, Hardworking and Academic interaction. In analysis 180 cases were assigned to training sample and 79 cases to the hold out sample.

Figure 1 depicts that eight, six and two number of nodes in the input layer, hidden layer and output layer respectively. Eight independent variables that are entered in the analysis all are covariates. Table 2 shows that for training and holdout the percentage of correct classification is high. It means the estimated model will accurately distinguish between being a student at risk and not at risk.

Table 1
Classification of Students being at risk or not at risk for Training and Holdout Sample

Sample	Observed	Predicted		
		At Risk	Not at Risk	Percent Correct
Training	At Risk	56	0	100.0%
	Not at Risk	0	130	100.0%
	Overall Percent	30.1%	69.9%	100.0%
Holdout	At Risk	16	2	88.9%
	Not at Risk	3	52	94.5%
	Overall Percent	26.0%	74.0%	93.2%

Dependent Variable: Risk

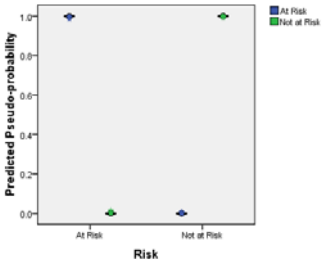

Figure 2. *Predicted Pseudo Probability*

All the graphs (Predicted Pseudo Probability Chart, ROC Curve, Lift Chart and gain chart) are calculated by using the data of training sample. Classification accuracy is 100% in training sample, so all the results calculated by training sample are accurate. The predicted –by-observed shown in Figure 2.The leftmost box plot shows, for cases that have observed category At Risk, the predicted pseudo-probability of category At Risk. In this case, all portion of box plot is above 0.5 so our predicted pseudo probabilities are very high for at risk category.

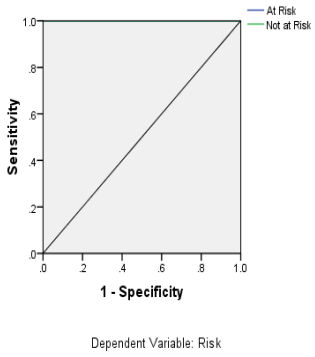

Figure 3. *ROC Curve*

The next box plot to the right shows, for cases that have observed category At Risk, the predicted pseudo probability of category Not at Risk.

Table 3

Area under the Curve

		Area
Risk	At Risk	1.000
	Not at Risk	1.000

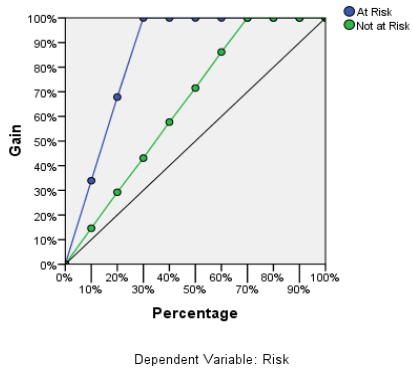

Figure 4.*Gain Chart*

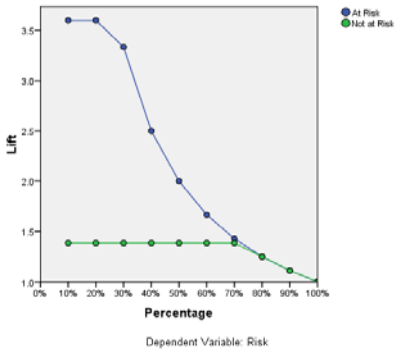

Figure 5. *Lift Chart*

The chart shown in Figure 3 two curves, one for the category At Risk and one for the category Not at Risk. Both curves are close to upper left corner; which shows that our fitted model is best. Table 2depicts the area under curve. Our model has an area of 1.0 so

our model is perfectly accurate. In figure 4, cumulative gains chart showed that the first point on the curve for the At Risk category is at (10%, 37%), meaning that if you score a data set with the network and sort all of the cases by predicted pseudo-probability of At risk category. You would expect the top 10% to contain approximately 37% of all of actually take the category At Risk. Figure also demonstrates that in first (10-30) %at risk move sharply upwards. Not at risk move upwards with almost with same rate on all the percentage levels.

Discussion

In previous researches conclude that CGPA (at the end of second semester), previous degree marks and academic interaction were most important predictors respectively for academic performance in universities. In this study CGPA (at the end of second semester) affect significantly to identify students at risk or not at risk. By using Neural Network we predicted approximately all the cases correctly. Neural Network is more flexible classifier as compared to others.

Conclusion

Our findings demonstrate that the most significant predictor of academic performance of students of social sciences at University of Gujrat (Hafiz Hayat Campus) was CGPA (at the end of second semester). Results of this paper underlined the importance of evaluating factors that are related to academic performance. In classifying the students at risk/not at risk, we could have achieved a rate of correct classification of over 95% in training sample and over 85% in holdout sample. The classification accuracy of CGPA of final degree is over 95% in training sample and over 75% in holdout sample.

References

Cheesman, J., Simpson, N., & Wint, A. G. (2006). *Determinants of student performance at university: Reflections from the Caribbean.* Unpublished Manuscript.

Bhardwaj, B. K., & Pal, S. (2011). Data mining: A prediction for performance improvement using classification, *International Journal of Computer Science and Information Security, 9*(4).

Erimafa, J. T., Iduseri, A., & Edokpa I. W. (2009). Application of Discriminant Analysis to predict the class of degree for graduating students in a university system. *International Journal of Physical Sciences, 4*(1),16-21.

Ibrahim, Z., & Rusli, D. (2007). *Predicting students' academic performance: comparing artificial neural network, decision tree and linear regression.* 21st Annual SAS Malaysia Forum, 5th September, Shangri-La Hotel, Kuala Lumpur.

Mohd, R. N., Zaidah, I., & Mohd, J. R. (2008). *Predicting students' academic achievement: Comparison between logistic regression, artificial neural network, and neuro-fuzzy.* International Symposium on information Technology, IT Sim 2008. Kuala Lumpur, Malaysia.

Naik, B., & Ragothaman, S. (2004). Using neural network to predict MBA student success. *College Student Journal, 38*(1), 143-149.

Nghe, N. T., Janecek, P., & Haddawy, P. (2007). *A comparative analysis of techniques for predicting academic performance.* 37th ASEE/IEEE Frontiers in Education Conference October 10-13.

Oladokun, V. O., Adebanjo, A. T., & Owaba-Chales, O. E. (2008). Predicting students academic performance using artificial neural networks: A case study of engineering course. *The Pacific Journal of Science and Technology, 9*(1), 72-79.

Vandamme, J. P., Meskens, N., & Superby, J. F. (2005). *Fresher's achievement: Prediction methods and influencing factors.* Higher Education, Multijuridictionality and Golbalisation Conference, Ateliers des Fucam, Mons, Belgium. Dec 16-17.