Predicting Effect of Past Qualification in Successful Honoring Master of Computer Applications Degree – A Neural Network Approach

Shanti Verma

L.J.Institute of Computer Applications, Ahmedabad, India

Abstract

This in last decade number of private university in India grown exponentially. For technical courses it is very important that what student's previous educational background was? As in Master of Computer Applications (MCA) -3 year course of Gujarat Technological University (GTU) either student have commerce or science background in Higher Secondary qualification (HSC) and Bachelor of Science (B.Sc.), Bachelor of commerce (B.Com.) or Bachelor of Computer Applications (BCA) as Graduate qualification, they take admission in MCA program. So it is very important to see which previous qualification has more impact in awarding the MCA degree in 3

years. The objective of study is to develop a predictive model using Neural Network- Multilayer perceptron approach to validate how much the previous (HSC, Graduate) qualification of student associated with MCA degree awarded in specified time span of 3 years? The present study involves a group of 126 students of GTU who take admissions in MCA program in academic year 2011-2014. The analysis clearly shows in classification table that 73.2% training data was classified correctly corresponding to 26.8% incorrect predictions in model summary table. The analysis also shows that graduate qualification is more important (59.5%) as compared to HSC qualification (40.5%).

Keywords: Neural Network; Educational qualification,; Model Validation,; Classification Table; ROC Curve

Introduction

All India Council for Technical Education (AICTE) report of academic year 2013-2014 tells that number of private university in India grown exponentially [1].For Master of Computer Applications (MCA) 3 year program in Gujarat Technological University (GTU)criteria for admission is any graduate and any Higher secondary qualification (HSC)[2]. Since MCA program is technical program authors try to find out which HSC and Graduate qualification have more impact on getting MCA degree in specified time -3 years. Knowledge Discovery in Database (KDD) uses collection of analytical techniques to discover unidentified knowledge from large Database [3]. Various data mining techniques are applied in different areas like Medical, Production, Manufacturing and education [8]. Many researchers perform various data mining techniques in educational database to discover useful knowledge for improvement of quality and results. In this paper author has applied Neural Network- Multilayer Perceptron technique to discover how much the pervious(HSC, Graduate) qualification associated with MCA degree awarded in 3 years.

Neural Networks are the preferred tool for many data mining applications because of their power, flexibility and ease of use. Neural Networks are useful in Forecasting, Predicting the probability and detecting fraudulent transactions in database [4].

This paper is organized as Introduction is provided in section I, Objective of study is defined in section II, Literature review of Neural Network and Education Mining discussed in section III, section IV discuss data pre processing, section V discuss the findings of experiment using SPSS tool and conclusion is provided in section VI.

OBJECTIVES OF STUDY

For improving the educational results it is very important to apply data mining techniques in educational database to discover unidentified results. As students are the future of country it is important that they get proper guidance to choose higher studies program. So this study aims to establish a model between HSC, Graduate qualification with post Graduate MCA degree awarded in 3 years. The objectives of study are

1) To validate either HSC or Graduate have more impact on MCA degree awarded in 3 years.
2) To check what extent holdout data was classified correctly for defined model.

LITERATURE REVIEW

Related Work- Neural Network

Neural Network technique is used by may researchers for forecasting, predicting probability and detecting fraudulent transactions in various databases.

i. Xuhui Wang and Jian Xu proposed the teaching quality evaluation model using back propagation neural network and also give various applications of neural network in education mining [5].

ii. Timothy Wang and Antonija Mitrovic used neural network to predict student performance. They develop an intelligent problem selection agent based model to predict number of errors a student will make [6].

iii. Changjun Zhu, Zhenchun Hao used neural network model for comprehensive evaluation of teaching levels in fluid mechanics. They prepare a one kind of black box model which contains types of neural network, data pre-processing, training sample, input pattern, network topology, parameter estimation and model examination [7].

iv. Neural Network technology was also used by Yoa Gu, Lan Guo, Qingyu Sun for performance evaluation of e- commerce enterprises in china [9].

v. Liu Zhao-cheng, LIU Xi-yu, ZEHNG Zi-ran, WANG Gong-xi were used neural network model to forecast scales of Higher Education in China. They observed that GRNN efficiently managed the uncertainties presented in the raw historical data and could use the few samples [10].

vi. In assessment of student academic achievement considering categorized individual difference at engineering education Mustafa, et.al used Neural network approach .The simulation results concerned with students attitudes either introversion or extroversion [11].

vii. Ruba Alkhasawneh, Rosalyn Hobson used qualitative and traditional statistical techniques to identify the factors that affect student retention. They propose two models. The first model is applied to predict incoming retention identify pre college correlation factor. The second model is applied to classify students in groups [12].

Related Work- Educational Data Mining

In current scenario data mining can be used in various fields. Education is one of the most upcoming fields where data mining is used, which is also called as "Education Data Mining". Various data mining approaches are used by different researchers in different areas such as:

i. Naive Bayes classification data mining technique used by Pandey and pal to predict new comers students will be performer or not? [13].

ii. Linear regression model was applied by Hijazi and Naqvi to discover a functional relationship between mother's education, family income and student academic performance. They also try to find the degree of association between these factors [14].

iii. Association rule, lift technique and chi-square test were used by Jigesh to predict failures in related subjects [15].

iv. As per Ayesha, Mustafa, Sattar and Khan, researchers use k-means clustering algorithm to predict student learning activities [16].

DATA PRE-PROCESSING- DATA SELECTION, COLLECTION AND TRANSFORMATION

The continuous explosion of data has promoted the development of the process of data mining DM [13] [15], or, Knowledge Discovery [14]. DM is defined as an interactive, iterative, nontrivial process of deriving valid, interesting, accurate, potentially useful, and unpredictable patterns from data. The data mining process is usually divided into many subtasks, in Figure 2 the DM process applied to M-business data [11].

In this paper author has selected data sets of 3 institutes of one of largest university of India, "Gujarat Technological University". The data sample is taken from Post Graduate Course "Master of Computer Applications (MCA)" of 126 students of academic year 2011-2014.

Data selection
      i. Data of 3 institutes out of 40
      ii. 126 data collected out of 320

Data Collection
      i. Prepare a excel sheet for data collection in which three variables are defined.
          a. HSC Stream (Science, commerce)
          b. Graduate Degree name (BCA, BCOM, BSC)
          c. MCA degree awarded status (Yes, No) within 3 years
      ii. 126 respondents fill the excel sheet as a sample.

Data Transformation
      i.   Convert HSC stream as quantative data

Table 1: HSC Stream coding

HSC Stream	Science	Commerce
Coded_HSC_stream	1	2

      ii.   Convert Graduate Degree Name as quantative data

Table 2: Graduate Degree coding

Graduate Degree Name	BCA	BSC	BCOM
Coded_graduate_degree	1	2	3

      iii.   Convert MCA degree awarded status as quantative data

Table 3: MCA degree status coding

MCA Degree awarded Status	Yes	No
Coded_MCA_Status	1	0

Tool used for Analysis
      i. Data mining tool SPSS
      ii. Technique- Neural Network

FINDINGS AND DISCUSSION

Author applied the Multilayer perceptron- Neural Network technique to predict the probability of importance of independent variables for dependent variable. Here independent variables are HSC stream and Graduate degree and dependent variable is MCA degree awarded. The following results are get after perform the Multilayer perceptron- Neural Network technique to sample data of 126 students.

Classification Table

Table 4 : Classification Table for Model

Sample	Observed	Predicted		
		NO	YES	Percent Correct
Training	NO	5	9	35.7%
	YES	6	10	62.5%
	Overall Percent	36.7%	63.3%	50.0%
Testing	NO	7	19	26.9%
	YES	8	21	72.4%
	Overall Percent	27.3%	72.7%	50.9%
Holdout	NO	7	8	46.7%
	YES	3	23	88.5%
	Overall Percent	24.4%	75.6%	73.2%

The classification table shows the practical results using the network. For each case, the predicted response is yes if that cases predicted probability is greater than 0.5. The holdout sample help to validate the model; For the above case author used to create a model, where 73.2% training cases are classified correctly, corresponding to 26.8% incorrect cases. This suggest that, overall, model is in fact correct about 3 out of four times approximately.

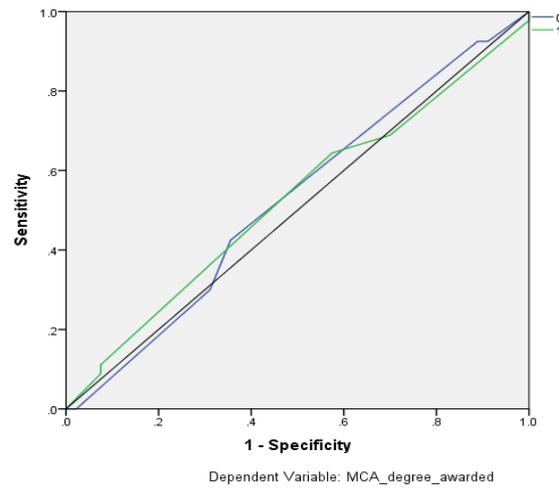## ROC Curve

Dependent Variable: MCA_degree_awarded

Fig1: ROC curve for dependent variable MCA_degree_awarded

The ROC curve gives a visual display of sensitivity and specificity for all possible cut offs in a single plot. The chart shown here display two curves for dependent variable MCA_Degree_Awarded one for the value yes (1) and another for the value No (0).

Independent Variable Importance

Table 5: Independent Variable Importance

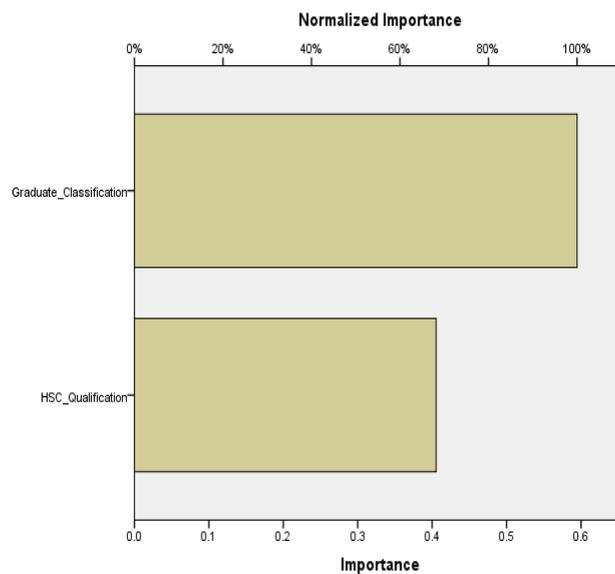|	Importance	Normalized Importance
Graduate_Classification	.595	100.0%
HSC_Qualification	.405	68.2%

Fig 2: Normalized importance chart

The importance of the independent variable (HSC and Graduate qualification) is a measure of how much the predicted value of network changes the different values of independent variable. The importance of graduate qualification is visible in the normalized importance chart. It means that for the awardance of MCA degree graduate qualification play major role as compared to HSC qualification.

CONCLUSION

The objective of validate independence variable importance from findings authors say that graduate qualification have more impact (59.5%) as compared to HSC qualification (40.5%). The hold out data was classified 73.2% correctly means that model is good enough to predict dependent variable value. It is expected that these findings are useful for university to design the course curriculum for MCA program so that the university results get increased.

References

[1] All India Council for technical Education (AICTE) approval process handbook (2013-14), pp. 4.-7.
[2] http://www.gtu.ac.in
[3] Richard M. Felder, G. N. Felder, E.J. Dietz: "The Effects Of Personality Type On Engineering Student Performance And Attitudes", http://www4.ncsu.edu /unity/lockers/users/f/felder/public/Papers: Journal of Engineering Education, 91(1), 3–17 (2002).
[4] IBM SPSS neural network 20
[5] Xuhui Wang and Jian Xu, "The Model of Teaching Quality Evaluation Based on BP Neural Networks and Its Application", IEEE First International Workshop on Education Technology and Computer Science,pp. 916-919, 2009
[6] Timothy Wang and Antonija Mitrovic, "Using Neural Networks to Predict Student's Performance", Proceedings of the International Conference on Computers in Education (ICCE'02), ISSN:0-7695-1509-6/02
[7] Changjun Zhu, Zhenchun Hao, "Application of Artificial Neural network in Fluid Mechanics Teaching Evaluation System", IEEE First International Workshop on Education Technology and Computer Science, ISSN: 978-0-7695-3557-9/09, pp. 12-14, 2009
[8] G K Gupta, Introduction to data mining with case studies, Second Edition, PHI Learning pp.24-29, 2010
[9] Yoa Gu, Lan Guo, Qingyu Sun, "The application of Balanced Scorecard and Neural Network on the Performance Evaluation for Enterprises" International Symposium on Intelligent Ubiquitous Computing and Education, ISSN: 978-0-7695-3619-4/09, pp. 105-107, 2009
[10] Liu Zhao-cheng, LIU Xi-yu, ZEHNG Zi-ran, WANG Gong-xi,," General Regression Neural Networks in Forecasting the Scales of Higher Education", IEEE First International Workshop on Education Technology and Computer Science, ISSN: 978-1-4244-3930-0/09, pp. 1257-1261, 2009
[11] Mustafa, et.al, "On Assessment of Students' Academic Achievement Considering Categorized Individual Differences at Engineering Education (Neural Networks Approach)", ISSN: 978-1-4799-0086-2/13
[12] Ruba Alkhasawneh, Rosalyn Hobson , "2011 IEEE Global Engineering Education Conference (EDUCON) – "Learning Environments and Ecosystems in Engineering Education", ISSN: 978-1-61284 -643-9/11, pp. 660-663
[13] U . K. Pandey, and S. Pal, "Data Mining: A prediction of performer or underperformer using classification", (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), pp.686-690, ISSN:0975-9646, 2011
[14] S. T. Hijazi, and R. S. M. M. Naqvi, "Factors affecting student's performance: A Case of Private Colleges", Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006

[15] JigneshDoshi , Result Mining: Analysis Of Data Mining Techniques In Education, International Journal of Technical Research and Applications e-ISSN: 2320-8163, Volume 2, Issue 3 (May-June 2014), PP. 25-28
[16] Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, "Data mining model for higher education system", Europen Journal of Scientific Research, Vol.43, No.1,