Advances and Applications in Mathematical Sciences
Volume 21, Issue 9, July 2022, Pages 5281-5290

PREDICTION OF ACADEMIC PERFORMANCE OF ENGINEERING STUDENTS BY ARTIFICIAL NEURAL NETWORK

RADHESHYAM H. GAJGHAT, CHANDRAHAS C. HANDA, RAKESH L. HIMTE, VIJAY N. KALBANDE and SHITALKUMAR A. RAWANDALE

Department of Mechanical Engineering
Christian College of Engineering
and Technology, Bhilai-490026, India
E-mail: radhegaj@gmail.com

Department of Mechanical Engineering
KDK College of Engineering
Nagpur-440008, India
E-mail: chandrahashanda@gmail.com

Department of Mechanical Engineering
Rungta College of Engineering
and Technology, Bhilai-490026, India
E-mail: rakeshhimte@gmail.com

Department of Mechanical Engineering
Nagpur Institute of Technology, Nagpur-441501, India
E-mail: drvijaykalbande@gmail.com

Department of Mechanical Engineering
Pimpri Chinchwad College of Engineering
Pune-411044, India
E-mail: s.rawandale@gmail.com

2020 Mathematics Subject Classification: 68T07.

Keywords: engineering students, academic performance, influencing factors, university examination, artificial neural network.

Received January 1, 2022; Accepted May 14, 2021

Abstract

In this study, an ANN model was developed to predict the academic performance of engineering students. It will help the students to enhance their performance in university examination by knowing their expected academic performance in advance. In this research, initially 22 significant influencing factors were identified and validated by stakeholders. These 22 factors were grouped into four major categories viz. personal, pre-admission, institutional and self-learning factors which were used as independent variables and pass/fail result in university examination was taken as dependent variable. The data sets of 840 engineering students were collected from the technical institutions affiliated to Chhattisgarh Swami Vivekanand Technical University. After data filtration data sets of 419 students were used to develop ANN model by using MATLAB 2015a software. The overall R2 value of ANN model was found to be 0.8766. It indicates the potential capability of the ANN model as a prediction tool. This ANN model was also tested satisfactorily by using the data sets which were collected separately from the different batch of students.

1. Introduction

India, the largest democracy of the world, has demonstrated a huge development of its techno-financial advancement, throughout the most recent 70 years of post freedom period. Gradually, India is becoming self dependence in various sectors and key areas including space technology, IT sector, and food security. Presently, Indian economy has become world's biggest 6th economy, above all making India proud by fulfilling the requirement of techno manpower for the advanced nations of the world. This became possible basically on account of combined efforts and continuous support of Indian government and country's private technical education sector (Chopra and Sharma, 2010). During the most recent decade various private and government technical institutions have been started to satisfy the requirement of techno manpower. During this period, especially in the last decade, the country saw the quantitative growth in the number of engineering students at diploma, degree and postgraduate level in general.

The huge number of engineering seats available in the country has opened up the door for students in 12th grade with poorer results, which have an impact on their performance. The poor result of the university exam also negatively affected placements, which resulted in the largest concern about the engineering education system among engineering students. Many efforts were taken by the top administration and academicians to improve the results; however these efforts were not focused because of absence of right

approach for recognition of real influencing factors. If these factors due to which students results are influenced are to be identified and their academic performances are to be known in advance, the remedial actions like taking additional classes for these students, utilizing advanced tools for teaching can be taken by the college administration and the teachers to improve their results. This improvement in results will definitely help the students to enhance their placements. Good placement is one among the most essential criteria that can help the engineering institutes to attract the students (Kabra and Bichkar, [8]).

The academic performance of the students is not only influenced by their individual characteristics but also other different factors which are involved (Daniyal et al., [3]). There are some assured factors such as family conditions, individual characteristics, college environment, academic background etc. which have significant impact on the student performance. The past studies confirmed that students' performances can be enhanced by knowing the influencing factors in advance. As of now the various models have been developed for predicting the student performance at the university level. The majority of the research focuses on the performance of students in foreign universities which, because of social, cultural and academic differences, may not be suitable for Indian universities (Harb and El-Shaarawi, [7]). So a more appropriate model for Indian university students is to be developed.

In the majority of the models, the semester/cumulative point average (SGPA/CGPA) and the pass/fail result at university were taken as dependent variables. In several researches pre-admission factors, family background, individual characteristics, learning habits, college environment etc. are considered as independent variables when it comes to predicting student performance at the university level (Gajghat et al., [5]).

In the study conducted by Gajghat et al., 22 important factors were initially identified as influencing factors and pass/fail result was identified as performance criteria to evaluate the performance of engineering students. For simplification of model formulation, 22 factors were grouped into 4 categories as per their similarities; and named as personal factors, pre-admission factors, institutional factors and self-learning factors (Gajghat et al., [6]).

By using the above four influencing factors as independent variables and one performing factor as dependent variable, an artificial neural network model was developed to correlate engineering student's performance in their university examination. This model can calculate the probability of passing of engineering students in their university examination.

This investigation may help engineering students to enhance their academic performance in university examination by knowing the expected performance in their university exam in advance. This investigation will be also helpful for the parents and faculties by knowing the expected performance of the students and accordingly provide them appropriate guidance and support. College administrators will have the option to outline better educating learning policy and build up students' friendly learning process (Alos et al., [1]).

2. Methodology

2.1. Formulation of ANN Model for Pass/Fail Result

Formulation of model is the method of converting a real life problem into a mathematical equation/model or other decision model. It establishes relationship between dependent and independent variables. The model is formulated using a variety of statistical methodologies and advanced techniques. Some of the most commonly used methods and techniques are multiple linear regression, logistic regression, polynomial regression, generalized linear model, artificial neural network etc.

In this research, the engineering students' performances were recognized in terms of pass/fail result in their university examination and hence pass/fail result (binary) model was developed. Students who passed were coded as 1 and those who failed were coded as 0. As a result, the final outcome is stated in binary terms. When a linear model is developed, it will produce an output in the form of a fraction ranging from 0 to 1, which also shows the possibility of passing.

The goal of the suggested study was to create an artificial neural network model that could predict engineering student's probability of passing their university exams. Personal (X_1), Pre-admission (X_2), Institutional (X_3),

and Self-learning (X_4) characteristics were used as independent variables, while the probability of passing (Y_1) (which can be converted into pass/fail result) was used as the dependent variable. This relation can be expressed as

$$\text{Probability of Passing, } (Y_1) = f(X_1, X_2, X_3, X_4)$$

Here, the data sets of 840 engineering students were collected from the technical institutions affiliated to Chhattisgarh Swami Vivekanand Technical University and after data filtration data sets of 419 students were used to develop ANN model by using MATLAB 2015a software. After selection of data as shown in figure 1, ANN randomly divides total samples into three sets viz. training, validation and testing. By default, the data is divided into 70%, 15%, 15% as training, validation and testing data respectively (MATLAB Help). This proportion can be changed. In this case, 70% of the total data i.e. 293 samples were used as the training set, 15% i.e. 63 samples as the testing set, and 15% i.e. 63 samples used for cross validation. After classifying the data, the neural network topology was constructed based on the Multilayer Perceptron with two hidden layers and twenty processing elements per layer (Demuth and Beale, [4]).

During training the network is provided with training data and in response to error the network is adjusted. Validation data is used to measure network generalisation and to stop training when generalisation ceases to improve. Testing data have no effect on training and that's why it provides independent measure for network performance during and after training. Figure 3 shows Train Network window and figure 4 shows ANN Model for Pass/Fail Result.

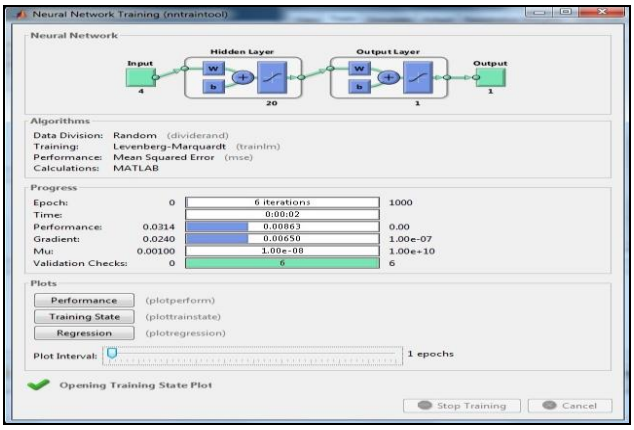
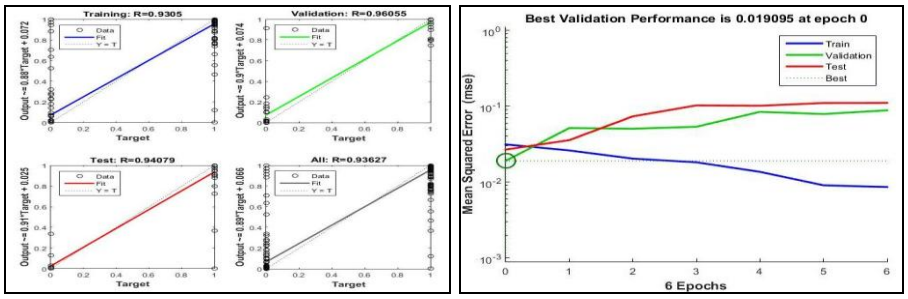


Figure 1. ANN Model for Pass/Fail Result.

3. Results and Discussions

3.1. Network Training and Validation Process

After selection of hidden neurons, the training has to be given to the network. For training neural network fitting tool uses Levenberg-Marquardt back propagation (trainlm) algorithm. Figure 5, 6 and 7 show the regression, performance and training state of pass/fail result model respectively. Multiple training results are generated because of various initial conditions. Mean squared error is the average squared difference between outputs and targets. It is better to have lower values of mean squared errors.



(a) (b)

Figure 2. ANN Model for Pass/Fail Result: (a) Regression (b) Performance.

Employing the number of runs set to three and the Epoch set to terminate at 1000, the network was trained with a supervised training

(Supervised and unsupervised learning). The performance of the training was assessed using the following performance measures:

The Mean Square Error (MSE):

$$MSE = \Sigma(dij - Yij)^2 / NP$$

Where,

P = Number of output of processing element

N = Number of exemplars in the data set

Y_{ij} = Network output for exemplars i at processing element j

d_{ij} = Desired output for exemplars i at processing element j

The number of neurons in this study varies from five to twenty and the performance of network was checked after repetitive training done on data. It is observed that the ANN network performance is best at overall R i.e. 0.93627 and Mean Square Error is 0.019095 at 20 neurons as shown in figure 5 and 6 respectively. The details of final ANN model for pass/fail result by using MATLAB R2015a software is shown in table 1.

Table 1. ANN Model for Pass/Fail Result.

Overall R_2	0.8766
Training R_2	0.8658
Validation R_2	0.9206
Testing R_2	0.8851
Mean Square Error	0.019095
No. of Hidden Layers	2
No. of Neurons	20
Input Variables	$4(X_1, X_2, X_3, X_4)$
Output Variables	$1(Y_1)$

3.2. Testing of ANN Model for Pass/Fail Result by using Separate Data Set Samples

The network was tested after training and cross-validation with separately collected data sets of different batch of students, which were not used for the development of this model. The following tabulated results were obtained for the tested dataset.

These separate test datasets have been utilised as the input variable data for the network without the output variable results. The network output data was then compared with the actual output data. This comparison of results is summarized in table 2.

Table 2. Testing of ANN Model by using Separately Collected Data Sets.

Sr. No.	X_1	X_2	X_3	X_4	Y	$Y - Y_b$	$Sq(Y - Y_b)$ (SST)	Y_p	$Y - Y_p$	$Sq(Y - Y_p)$ (SSE)
1	3.54	2.00	3.80	3.50	0	-0.43478	0.189036	0.00000	0.00000	0.00000
2	3.54	2.75	4.20	3.90	1	0.565217	0.319471	0.02707	0.97293	0.94659
-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-
45	4.13	2.50	4.00	3.90	1	0.565217	0.319471	0.98350	0.01650	0.00027
46	3.87	2.00	3.53	3.10	0	-0.43478	0.189036	0.76119	-0.76119	0.57941
Sum					20		11.30435			3.32306

The formula to calculate the coefficient of determination (R2) manually is given by

$$R^2 = 1 - \frac{\Sigma(Y - Y_p)^2}{\Sigma(Y - \overline{Y})^2}$$

Where $Y =$ Observed value of dependent variable (i.e. Pass, 1 or Fail, 0)

$Y_p =$ Predicted value of dependent variable (i.e. Probability of Passing)

$Y_q =$ Mean of $Y = 20/46 = 0.43478$

$R^2 =$ Coefficient of determination

$$R^2 = 1 - \frac{SSE}{SST}$$

Total Sum of Squares (SST) = 11.30435

Sum of Square Errors (SSE) = 3.32306

$$R^2 = 1 - \frac{3.32306}{11.30435} = 0.7060$$

The coefficient of determination (R^2) of the developed ANN Model for predicting the performance of engineering students is 0.8766 as shown in table 1. During the testing of this model by using separately collected data sets of different batch, the value of R^2 is found as 0.7060 which indicate the satisfactory performance of this ANN model.

4. Conclusions

This study shows the capability of the Artificial Neural Network for predicting and measuring the engineering student's performance in terms of probability of passing in university exam which will help them to enhance their performance in university examination by knowing their expected performance in advance. The coefficient of determination (R^2) is the measure of how well sample data set fits the design of the model. In this study, the maximum value of overall R^2 for the developed ANN model was found as 0.8766, implying that this model is capable of accurately predict engineering student's performance in university examinations. This ANN model was also tested by using separately collected data sets of different batch of students for which calculated value of R^2 is found as 0.7060, showing the potential efficiency of ANN model as a prediction tool. This approach can be used to predict engineering student's performance in university examinations at Indian technical colleges.

References

[1] S. B. Alos, L. C. Caranto and J. J. T. David, Factors affecting the academic performance of the student nurses of BSU, International Journal of Nursing Science 5(2) (2015), 60-65.

[2] K. L. Chopra and P. B. Sharma, Higher technical education in India-Profile of growth and future perspectives, Asian Development Bank Regional Technical Assistance Project, (Retrieved 10 April 2013, from ttp://www.namstct.org/ADB_RETA_Report/Prof_KL_Chopra_Prof_PB_S harma.pdf.), (2010).

[3] M. Daniyal, T. Nawaz, M. Aleem and A. Hassan, The Factors Affecting the Students' Performance: A Case Study of Islamia University of Bahawalpur, Pakistan, African Journal of Education and Technology 1(2) (2011), 45-51.

[4] H. Demuth and M. Beale, Neural Network Toolbox User's Guide, The Math Works, (2002).

[5] R. H. Gajghat, C. C. Handa and R. L. Himte, Factors Influencing Academic Performance of the Students at University Level Exam: A Literature Review, International Journal of Research in Engineering and Technology 6(5) (2017a), 102-110.

[6] R. H. Gajghat, C. C. Handa and R. L. Himte, Identification of important factors influencing the performance of engineering students in university examination: a systematic approach, Research Journal of Engineering and Technology 8(4) (2017b), 447-452.

[7] N. Harb and A. El-Shaarawi, Factors Affecting Students' Performance, Munich Personal Archive (MPRA), (2006).

[8] R. R. Kabra and R. S. Bichkar, Performance prediction of engineering students using decision trees, International Journal of Computer Applications 36(11) (2011), 8-12.

[9] MATLAB Help available through MATALB R2015a software.

[10] Supervised and unsupervised learning, http://www.nnwj.de/supervised-unsupervised.html, (Browsing date: 20th July 2019).