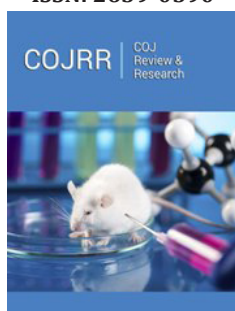# Academic Performance Prediction of Engineering Student

Ravindra Duche*, Rohan Tipnis and Srishti Singh

Department of Electronics and Telecommunication, India

Abstract

Early prediction of student's performance helps to take necessary remedies for better grades of students. To achieve the better education standard, several attempts have been made to predict the performance of the student, but the prediction accuracy is not acceptable. To accomplish the enhanced prediction, Neural Network (NN) based method is proposed. In this paper, an approach to predict student academic performance in college education based on artificial neural network is proposed. This paper also focuses on how the prediction algorithm can be used to identify the most important attributes in a student's data. We could actually improve students' achievement and success more effectively in an ancient way using educational data mining techniques. It could bring the benefits and impacts to students, educators and academic institutions.

Keywords: Academic performance; Prediction; Machine leaning; Neural networks; Regression

*Corresponding author: Ravindra Duche, Department of Electronics and Telecommunication, Mumbai, India

Introduction

Machine learning has been a booming field in the past few years, and it has been applied to almost every field. Predicting the academic performance of a student is a highly concerned topic, especially when it comes to higher education like degree courses as they determine the further path of a student's career. A student after graduation may choose to opt for higher studies or a job in a private sector or in the government sector. It is the CGPA of a student that often results in not qualifying for a particular job profile or a particular course in a post-Graduation college. The scope of the paper is to determine the outcome of a student so as to give warn on an appropriate time to students who are at risk and thereby will help the institution in determining to take remedial measures to improve the student's academic performance. The role of engineering students differs strongly from that of students in other fields. According to the graduate attributes defined by a number of engineering accreditation boards, an engineer must be able to: apply knowledge of mathematics, physics, and life sciences in order to understand, formulate, and solve engineering problems; design and conduct experiments; analyze and interpret data; develop designs that meet specified requirements; design solutions to new problems, possibly involving other disciplines; perform in multidisciplinary teams; understand engineers responsibilities, as well as the ethical social, economic, environmental, and political impact of the engineer profession. What differentiates engineering from other disciplines is thus its strong focus on mathematics and physics, combined with a range of domain specific abilities and knowledge.

A study conducted by Havan et al. [1] where they used Multivariate Linear Regression. The dataset initially had as few as 40 students and number of subjects were 17. The test set was of 10 students. The accuracy turned out to be 50percent for predicting the marks of a single subject. The dataset was further increased to 70 and the accuracy obtained then was 70 percent.

Behrouz M [2] have developed a prediction model that was based on multiple classifiers. The prediction model classified student on the basis of grade. The prediction performance was better compared to other classifier-based approaches since multiple classifiers are combined. Genetic algorithm was also applied to weighting feature vectors aiding in enhanced prediction. Accuracy in the prediction of about 94 percent was obtained, but complexity in a combination of multiple classifiers such as Quadratic Bayesian classifier, 1- nearest neighbor (1-NN), k-nearest neighbor (k-NN), Parzen-window, Multi-Layer Perceptron (MLP), and Decision Tree is not tolerable. The performance of multi classifier approach decreases with increase in attribute selected for prediction. Romero et al. [3] conducted a survey of educational data mining from 1995-2005. They found out that most projects are targeted towards improving student learning activities, instructor teaching methodologies, and institution structuring. They also introduced the use of data mining in predicting student performance in a course within the context of e-learning and intelligent tutoring systems. Nguyen et al. [4] used decision trees and Bayesian Network algorithms to predict a student's third year GPA using the students second year record. However, they have not identified factors that affect success or failure; hence, their techniques cannot be of further use in improving student performance. Azmi et al. [5] used similar techniques as Nghe to predict and classify students into groups according to their academic performance, based on the student's records. This classification though lacks the identification of the relevant predictors of success, and simple lumps a student's complete degree record in its analysis.

Methodology

Pervious actual academic performance of engineering students is collected as theory, term work, practical marks and attendance of each subject in each semester starting from second year to third year. We have used Feed Forward Neural Network (FFNN) to predict the performance of students of first semester of final year. Data collection is illustrated in the following (Figure 1). The student record simply shows the courses taken and grades obtained, as well as the GPA and program standing (i.e., good, probationary, failed). There might be other factors that affect a student's performance. Each semester has various sections in which the marks are divided such as theory, unit test, term work and practical exam. Thus, all of these marks were considered for every subject of every semester. First year marks were not considered because it does not have all subjects related to the field in which the student is studying. A training dataset of 30 students was taken from the previous years' batch and a testing set of ten students were taken from the current graduating batch [6].
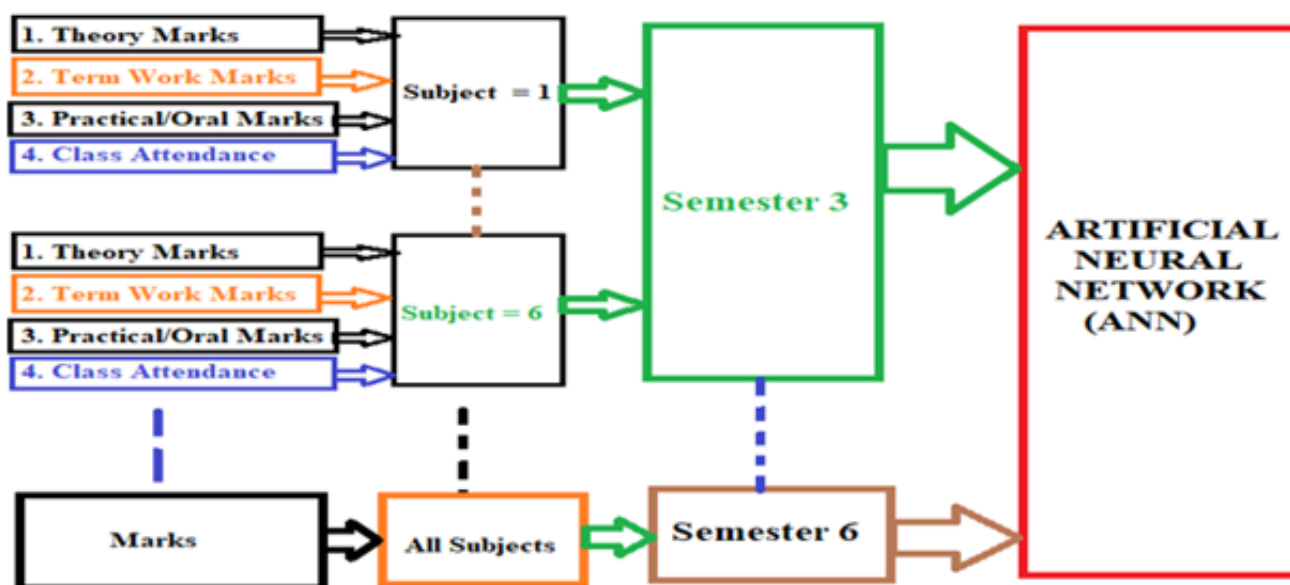

Figure 1: Data Set for academic performance predication.

Predication using Feed Forward Neural Network (FFNN)

Neural Networks (NN) are currently used prominently in voice recognition systems, image recognition systems, industrial robotics, medical imaging, data mining and aerospace application. NN are particularly effective for predicting events when networks have a large database of prior examples. So, we have decided to use FFNN for students' performance predication. FFNN with three layers is used. The 28 inputs such as theory, term work, oral, test marks and attendance of each subject of semester 3 to 6 are applied to first layer. Then Four neurons are used in intermediate layer with appropriate weights and bias. In the output layer Ten neurons are used for pointers respectively. Used FFNN structure is illustrated in below Figure 2.
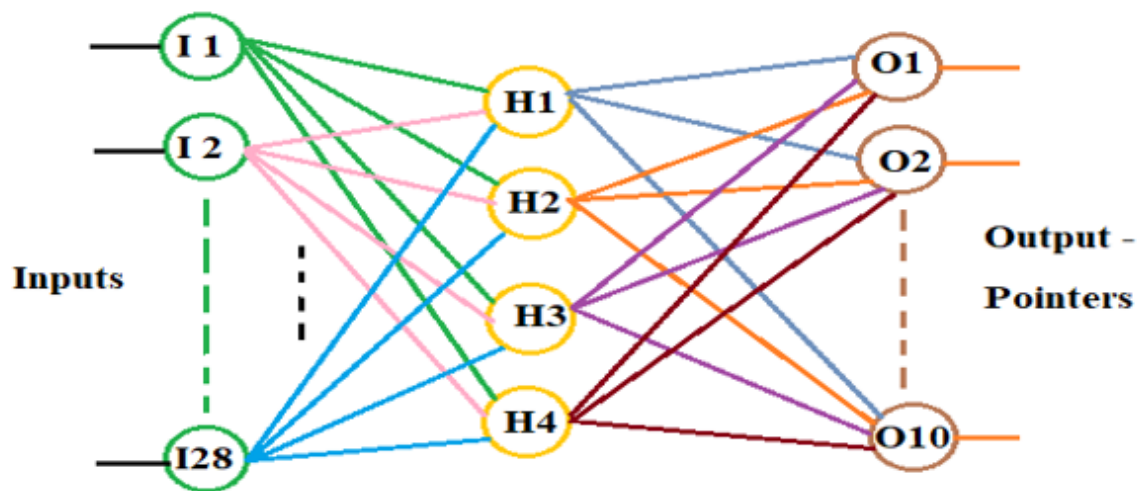
Figure 2: FFNN to predict performance in term of pointers.

Predication using linear regression

Linear regression requires the relation between the dependent variable and the independent variable to be linear. When the distribution of the data is more complex linear models cannot be used to fit non-linear data. To overcome under-fitting, we need to increase the complexity of the model. To generate a higher order equation, we can add powers of the original features as new features. Polynomial Regression is used for fitting the kind of complex data that is shown below Figure 3.
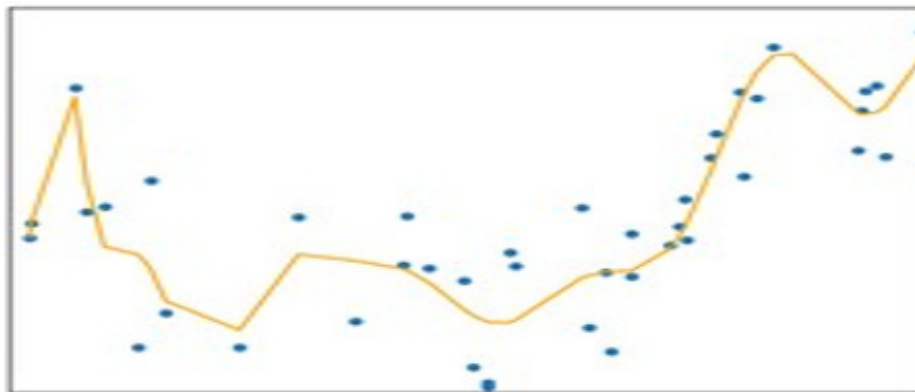

Figure 3: Data which is improperly fit in linear regression.

Result and Discussion

As we can see from the results since there exists a correlation between the subjects in consecutive Semester subject as well as the attendance in the respective semesters a prediction of the future exam pointer can be made from them. One of the important things in any neural network model is the size of the dataset. We know that the more the amount of data we have the better we can train the model. Since the size of our dataset was limited, we were facing problems of getting on to a local minimum rather than global minima. Due to this problem, we were getting a variable accuracy value. To tackle this problem, we tried hyper parameter tuning using the Grid Search CV. We got an average error margin of +1 to -1. We know that a regression type model is generally used for any type of prediction problems we tried to compare the results of our neural network model against a Polynomial Regression model.

From the comparison results shown in Figure 4, we can see that at some instances the regression model actually beats the neural network model, this happens mainly because the size of the training dataset is limited to only thirty samples. But at most of the cases the neural network model beats the regression model. Here we can make an inference that with sufficient amount of training data the neural network model will prove to be much better than the regression model. Once it was confirmed that the data conforms well to a machine learning algorithm, we conducted a comparative study of neural networks and Polynomial Regression, on the basis of varying training and test sets. The results were fairly surprising. In general, the neural networks tend to outperform Polynomial Regression. This is somewhat justified once one realizes that the input provided to the algorithm was on a continuous range, and Polynomial Regression traditionally requires discrete data.
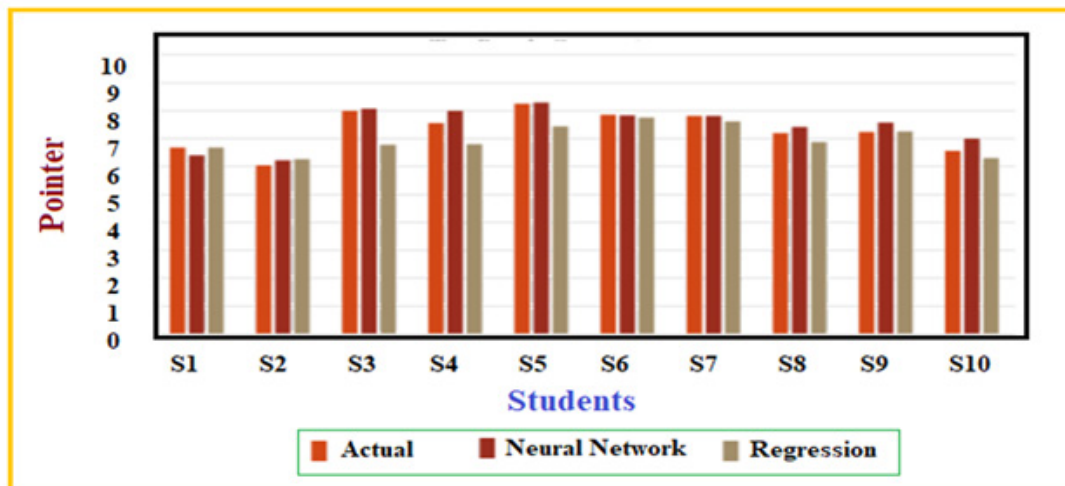
Figure 4: Comparison of actual, neural network and regression results of students' performance in term of pointer.

Conclusion

The methodology presented in this paper was an attempt to predict the semester performance of engineering students using learning capability of artificial neural network. The proposed prediction model was developed based on some input variables related to academic factors of students in which best factor for prediction. The selected features were given to an Artificial NN for prediction of semester marks. Adam Optimization algorithm was proposed for the training of neural network for optimal weight selection. The current study shows that the academic performance of a student is primarily dependent on his/her past performances. Our investigation confirms that past performances have indeed got a significant influence over students' performance. Further, we confirmed that the performance of neural networks increases with increase in dataset size. Machine learning has come far from its nascent stages and can prove to be a powerful tool in academia. In the future, applications similar to the one developed, as well as any improvements thereof may become an integrated part of every academic institution.

References

1. Havan A, Harshil M (2015) Student performance prediction using machine learning. International Journal of Engineering Research and Technology 4(3): 111-113.

2. Behrouz M, Punch WF (2003) Using genetic algorithms for data mining optimization in an educational web-based system. In proceedings of Genetic and Evolutionary Computation Conference, USA, pp. 2252-2263.

3. Romero C, Ventura S (2012) Educational data mining: A review of the state-of-the-art. IEEE Transactions on Systems 40(6): 601-618.

4. Nguyen TN, Janecek P, Haddawy P (2007) A comparative analysis of techniques for predicting academic performance. 37th ASEE/IEEE Frontiers in Education Conference, pp. T2G7- T2G12.

5. Azmi A, Ikmal P (2011) Academic performance predication based on voting technique. IEEE 3rd International Conference on Communication Software and Networks, China.

6. Amirah MS, Wahidah H, Nuraini AR (2015) A review on predicting student's performance using data mining techniques. Procedia Computer Science 72: 414-422.

For possible submissions Click below:

Submit Article