iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** IV **Month of publication:** April 2017

DOI: http://doi.org/10.22214/ijraset.2017.4163

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Educational Data Mining for Classification of Students based on their Performance

Amruta R. Joshi[1], Prof. A. Thomas[2]
[1]Research Scholar (M. Tech.), [2]Head of the Department (HOD), Computer Science and Engineering Computer Science & Engineering, G. H. Raisoni College of Engineering, Nagpur, India

Abstract: ***Performance of any institution is largely dependent on the success of its students. These students performance may be dependent on various parameters such as student's knowledge, educational background, teacher's teaching methods, financial background and their psychology. Due to large educational databases, predicting student's performance becomes a challenging task. There are ample student's present in any Institution so focusing on each and every student becomes a difficult task for a teacher. Also sometimes student don't understand where they are lacking and hence results in Low Grades. To monitor the student's progress and performance is at most important task. The existing system will lead to decrease the gap between student's performance level as well as the Teacher's Assessment. The main objective of this paper is to predict Students final grades and classify them according to their performance level. This paper also focuses on how the prediction and classification algorithm can be used to identify the student's performance level with the help of past data.***
With the help of this system, teachers will be able to categorize student according to their performance and can use that knowledge to help them to enhance their performance. It could bring the benefits and impacts to students, educators and academic institutions.

I. INTRODUCTION

Student past performance monitoring is an essential part in predicting future performance and hence classify them accordingly. This will help teachers to understand and know which student is good and which student is weak in studies. By this, the teacher can more focus on the weaker students so that the students can improve their performance in finals. EDM helps to understand how an individual student learns, and the way of their learning [1]. It enables analyzing student's data and makes decision for improving the current educational system and learning material. In this students' data such as semester grade, background qualification is explored to extract useful knowledge that can be used to improve students overall skills and provide with the best possible learning outcome and hence enhance the productivity of educational systems as well.

The objective is to gather the past performance information of students and to predict their final grades and classify them in different classes to know their performance level. So, the flowchart in Figure 1.1 gives an overview of the whole system.

The rest of the paper is arranged as follows. The Related Work of the proposed system is described in Section 2.Section 3 describes the Methodology about the essential aspects of Prediction and Classification of the students. Implementation is described in section 4.The enrollment of database for training and testing is created and is presented in Section 5 which is the result and discussion part. In Section 6, Conclusion of paper is described.

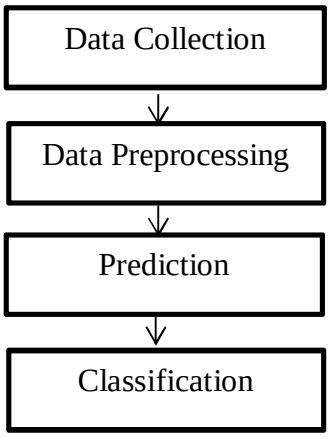


Figure 1.1: Flowchart of system

©IJRASET: All Rights are Reserved

II. RELATED WORK

Educational Data Mining is used for describing the research disciplines which uses data from educational settings. In [2], author have discussed about various methods and techniques for exploring the data collected from different educational sources and he has categorized the data in two forms which are offline data which can be collected from the traditional databases of the educational institutions and online data which can be collected from the online media. In this proposed system, the offline data of the students is collected. In [4], Author has presented a framework for data preparation to transform the data in understandable format for clustering, classification and association and for this purpose Data understanding and Data Preparation is required. Data understanding requires a detailed study of data sources, to become familiar with data meanings and its formats. In [5], discussed about data preparation includes activities that needed to be performing to generate final data set from the previous inconsistent data. In [6], Author have performed the analysis on the learning behavior of the students by using different data mining models, namely classification, clustering, decision tree, sequential pattern mining and text mining. They used open source tools such as KNIME, RAPIDMINER, WEKA, CARROT, ORANGE, RProgramming, and iDA. These tools have different compatibilities and it provided an insight into the prediction and evaluvation. In [7], Author have discussed about the advantages of RapidMiner tool which is used for Preprocessing of data as it have the facility of Descriptor Selection, Parameter optimization of machine learning/ statistical methods and Model validation using cross- validation and/ or independent validation set. In [8], discussed about the Neuro-Fuzzy classification method which is used for predicting and classifying students based on their past academic performance details and then to predict their future academic grades. Their performance is then categorized into several bands ranging from weak students to excellent students. In this proposed system, neural network toolbox is used prediction as it provides algorithms, pre-trained models, and apps to create, train, visualize, and simulate both shallow and deep neural networks to perform prediction and classification. In [9], Author has discussed about the classification method used for prediction with Decision tree classifiers and methods to solve the class imbalance problem. Decision tree classifier is used for classifying students according to their grades [9]. In [10], Author has stated about the best model for classification of students. They applied the combination of neural network and fuzzy IF-THEN rules on the data set containing four parameters motivation, talent, interest and intelligence.

III. METHODOLOGY

The techniques which are used for building up the final system are described in Figure 1.1. It gives a detail flowchart of the system. The following are the steps involved in proposed system.

A. Data Collection.
B. Data Preprocessing.
C. Prediction.
D. Classification.

A. Data Collection

In this step, student's past and current information is gathered which includes student's SSC grades; HSSC grades, and first to sixth semester's grades.

B. Data Preprocessing

The data received in the first step is unstructured. In this step the raw data is processed and is converted into structured data. For this preprocessing, RapidMiner tool is used. In RapidMiner, There are five types of views are available for performing the process - repository, process, design, result, operators. Firstly, data need to add by repository view. There are various operators which can be used for performing operations on data, we just need to drag and drop that operator to use it. And after performing the operations, we can get the result by clicking on run button. For pre-processing the data file is dragged from the repository view then replace missing value operator is used for eliminating missing values and that operator and data file is connected. The result file is generated after this connection process.

C. Prediction

In this step Student's future performance is predicted depending on their past and current data. Students' performance is one of the most explored issues in educational data mining. To predict if students will achieve the outcomes of the subject based on the previous results enables teachers to adapt the learning design of the subject to the teaching-learning process. However, this

©IJRASET: All Rights are Reserved

adaptation is even more relevant if we could predict the fulfillment of the educational objectives of a subject, since teachers should focus the adaptation on the learning resources and activities related to those educational objectives.

D. Classification

In this step Classification is done on the basis of performance, which is categorized as – Excellent, Good, Average and Poor. It is a two way technique which maps data into a predefined class. Supervised learning is used to separate data into different data sets. The students having grades between following range come under respective categories-

9 to 10 CGPA - Excellent
8 to 9 CGPA - Good
7 to 8 CGPA - Average
6 to 7 CGPA - Poor

The corrective Measures for improving the Student's learning material and enhancement of educational system can be taken. This way performance of weaker students can be improved.

IV. IMPLEMENTATION DETAILS

In this proposed system, 200 student's samples of three batches of Computer science department from one local college situated in Nagpur was considered. The grades they scored in SSC, HSSC, first semester to sixth semester grades were taken into consideration.

RapidMiner tool is used for preprocessing the student's past and current data. Pre-processing helps in removing data redundancy, missing values and to transform raw data into understandable form.

To perform prediction and classification neural network toolbox is used. In this, Feed-Forward back propagation algorithm is used as we got the best results using this algorithm. Only one layer is used for training in this proposed system. And the numbers of neurons used are 50. There are different training functions which are available but TRAINGDM training function is used in this proposed system. There are two types of Adaption learning function in neural network toolbox which are Learngdm and Learngd. Learngdm is used in this proposed system. And the best suitable performance function for these Traingdm and Learngdm is MSE. Depending on above selected training parameters the training is done and output is generated.

V. RESULT AND DISCUSSION

Name	Feed-forward back Propagation network	NARX	Feed-Forward distributed time delay	Layer Recurrent
Time taken For training the network	11 seconds	24 seconds	41 seconds	49 seconds

Table 5.1: Comparative analysis of different algorithms.

Above table shows the time duration of different algorithms took for training. As Feed-forward back propagation takes less time than other algorithms and it is a non-parametric algorithm, this algorithm is used in this proposed system.

No. of Students	No. of students with accurate predicted grade points	No. of students with different predicted grade points	Accuracy percentage
100	82	18	82%

Table 5.2: Predicted grades using NARX.

©IJRASET: All Rights are Reserved

Volume 5 Issue IV, April 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

No. of Students	No. of students with accurate predicted grade points	No. of students with different predicted grade points	Accuracy percentage
100	74	26	74%

Table 5.3: Predicted grades using Feed-Forward time delay.

No. of Students	No. of students with accurate predicted grade points	No. of students with different predicted grade points	Accuracy percentage
100	79	21	79%

Table 5.4: Predicted grades using Feed-Forward time delay.

No. of Students	No. of students with accurate predicted grade points	No. of students with different predicted grade points	Accuracy percentage
100	86	14	86%

Table 5.5: Predicted grades using Feed-Forward back propagation algorithm.

Above tables shows the analysis of predicted grade points using different algorithms. In this proposed system, Feed-forward back propagation algorithm is used as it is taking less time and better results. For training purpose, data of 100 students was taken who has already passed out. It was observed that from these 100 students, the predicted grade points of 86 students were correct and in remaining 14 student's grades was having dissimilarity in their predicted and actual grade points. So, the accuracy rate of this proposed system for prediction is 86%.

No. of Students	Excellent	Good	Average	Poor
100	23	38	26	13

Table 5.6: Classification of students in different categories.

After prediction, these 100 students were categorized in four groups according to their predicted grade points. From these 100 students, 23 students were categorized as Excellent, 38 were categorized as Good, 26 categorized as Average and 13 students were classified as Poor students.

Now, because of this categorization, it will be easy for teachers to identify the weaker students of the class and can more focus on them to improve their performance.

VI. CONCLUSION

The paper introduces method for predicting the future performance of the students according to their past and current data and classifying the students into four different groups like – Excellent, Good, Average and Poor depending on their predicted future performance. This in turn will help the teachers to identify and focus more on the weaker students which will help the students to improve their performance in final semester and will lead to their bright and successful future.

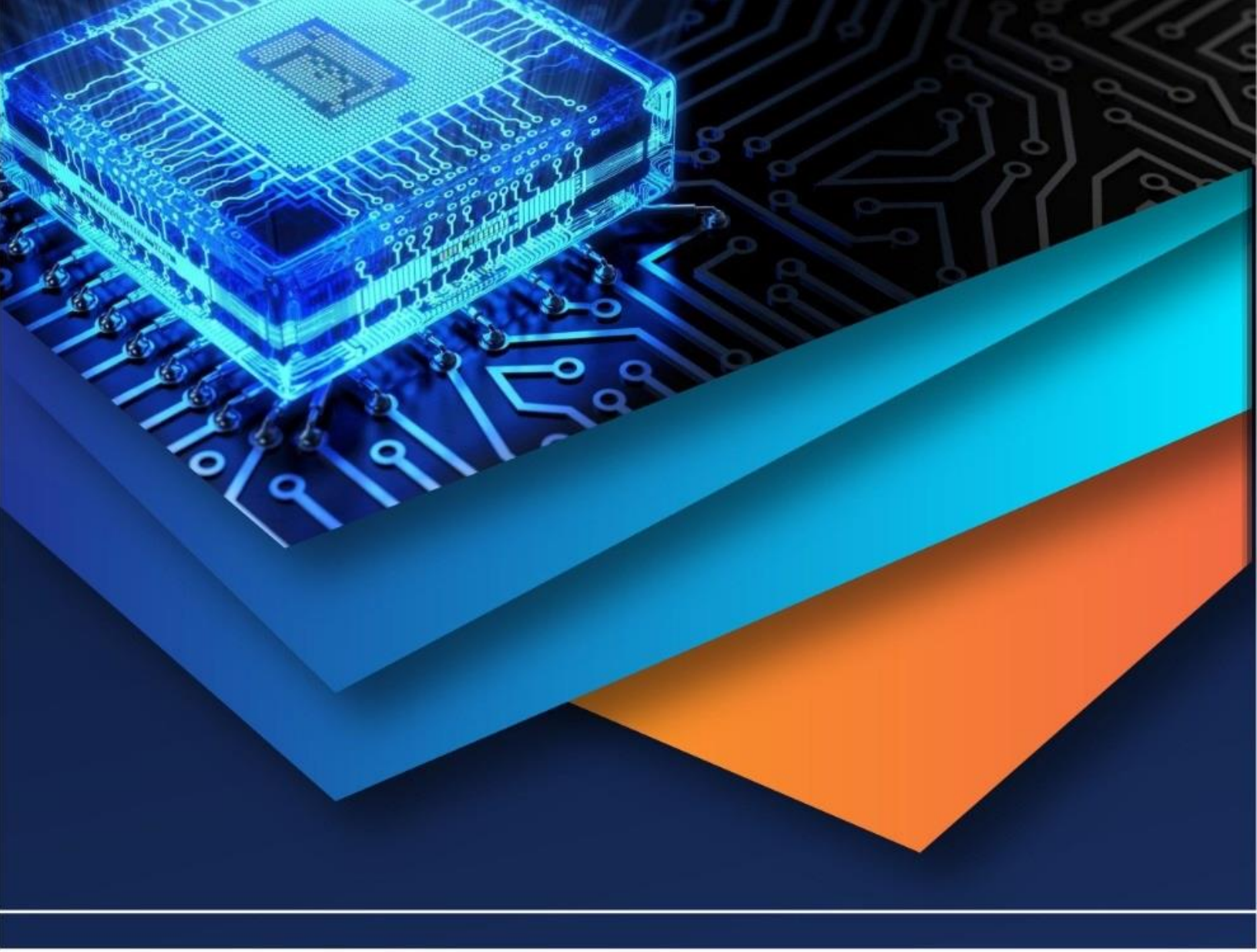
REFERENCES

[1] Norlida Buniyamin, Usamah bin Mat, Pauziah Mohd Arshad, "Educational Data Mining for Prediction and Classification of Engineering Students Achievement," 2015 IEEE 7th International Conference on Engineering Education (ICEED).

©IJRASET: All Rights are Reserved

[2] Karan Sukhija Dr. Manish Jindal Dr. Naveen Aggarwal, "The Recent State of Educational Data Mining: A Survey and Future Visions." 2015 IEEE 3rd International Conference on MOOCs, Innovation and Technology in Education (MITE).
[3] Debraj DeWen, Zhan Song, "DCTP: Data Collecting Based on Trajectory Prediction in Smart Environment" IEEE 2014.
[4] Elaf Abu Amrieh, Thair Hamtini, Ibrahim Aljarah, "Preprocessing and Analyzing Educational Data Set Using X-API for Improving Student's Performance" 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT).
[5] Mirela Danubianu, "A data preprocessing framework for students' outcome prediction by data mining techniques" IEEE 2015.
[6] Nikitaben Shelke and Shriniwas Gadage, "A survey of data mining approaches in performance analysis and evaluation", International Journal of Advanced Research in Computer Science and Software Engineering , vol 5, iss 4, 2015.
[7] Shraddha Dwivedi, Paridhi Kasliwal, Prof Suryakant Soni, "Comprehensive Study of Data Analytics Tools (RapidMiner, Weka, R tool, Knime)," Symposium on Colossal Data Analysis and Networking (CDAN) ©2016 IEEE.
[8] Bo Guo, Rui Zhang, Guang Xu, Chuangming Shi and Li Yang "Predicting Students Performance in Educational Data Mining." 2015 IEEE International Symposium on Educational Technology .
[9] Alana M. de Morais and Joseana M. F. R. Araújo Evandro B. Costa "Monitoring Student Performance Using Data Clustering and Predictive Modelling." IEEE , 2014.
[10] Shaleena K.P, Shaiju Paul "Data Mining Techniques for Predicting Student Performance." 2015 IEEE International Conference on Engineering and Technology (ICETECH), 20th March 2015, Coimbatore, TN, India.
[11] Krina Parmar Prof. Dineshkumar Vaghela Dr Priyanka Sharma"Performance prediction of students using distributed Data Mining." 2015 IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems (ICIIECS)
[12] Hanife GÖKER, Halil Ibrahim BÜLBÜL Improving an Early Warning System to Prediction of Student Examination Achievement", 2014 13th International Conference on Machine Learning and Applications.
[13] Pensri Amornsinlaphachai, "Efficiency of data mining models to predict academic performance and a cooperative learning model," 2016 IEEE Conference.
[14] Adam Mukharil Bachtiar, Hira Laksmiwati." Development of Requirement Elicitation Model for Prediction of Student Achievement in University", 2014 2nd International Conference on Information and Communication Technology (ICoICT).
[15] S Manuel Fern´ andez-Delgado, Manuel Mucientes, Borja V´azquez-Barreiros and Manuel Lama," Learning Analytics for the Prediction of the Educational Objectives Achievement", 2014 IEEE Conference.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)