Predicting Student Academic Performance Using Classification Algorithms: A Comparative Analysis

ISSN (e) 2520-7393
ISSN (p) 2521-5027
Received on 3rd Dec, 2018
Revised on 26th Dec, 2018
www.estirj.com

Urooj Makhdoom[1], Dr. Prof Tariq Jamil Saifullah khanzada[2]

[1]*Department of computer system, Mehran university of engineering and technology*
[2]*Department of computer system, Mehran university of engineering and technology*

Abstract: Data mining techniques are utilized to work on huge volumes of data to find hidden patterns and relationships between data that is helpful in decision making. It provides powerful techniques for various fields including education. With the help of some valuable tools data mining techniques discover a pattern. This paper proposes model that predicts student's academic performance from large available dataset by using classification techniques like decision tree and naive bayes. The experiment result shows that naive bayes is best technique among other by revising high accuracy 77%. The information extracted from prediction model will be used by instructors to identify student category that either student belong to which category good, average or poor with the help of our research, we could have provided educational institutions with a way of reducing number of failures.

Keywords: *Data mining techniques, Decision tree, Naive bayes, Educational field, Students academic performance (SAP.*

1. Introduction

The Data mining is one of the process which is used to mine the valuable information or knowledge from huge amount of data, data are mostly represent and stored in various forms example text data, voice data, video data etc. data mining is one the powerful technique through which an unstructured data converted in to meaningful structured data mining also used for prediction.

Educational field is rapidly increasing due to large amount of student's data are involved in any educational institute this can be very difficult to test knowledge about student performance, so it is necessary to develop a prediction model by using classification algorithm's this prediction model used to analyze student performance therefore to develop a prediction model is very essential for any educational academy.

In this project a lot of studies have been applied on data mining classification methods to predict SAP (student academic performance). The popular methods of classification are decision tree and naive Bayes. First ambition of using data mining techniques is to develop a prediction model the extracted model is then used to predict student's educational performance the overall student performance is predict from their previous record. The students' performance is used as predictor parameter in this way extracted model will help the instructors to identify the students' category in order to enhance the student performance in academics.

2. Related Work

In this [3] the neural network model that can be developed by using data mining techniques and MATLAB tool to build an expert system to predict dengue fever cases in early stages.

Another research [4] in which they survey most relevant study carried out in EDM they focused on some data mining techniques which will proved to be very useful to analysis the results to find out poor academic record of students and also find out solution to improve it.

This paper has been carried out to make a performance comparison of classification algorithm in the context of financial institute to find different conditions of employs this can also calculate the Classification accuracy and cost analysis by using WEKA tool [5].

In this paper, they compare decision tree algorithm with clustering techniques so as to predict the performance of these algorithms by using WEKA tool in terms students' academic performance and it proves that j48 is more accurate then clustering techniques [6].

This present study develops and compares four predictive mathematical models to predict students' academic performing in engineering dynamics these models are linear regression, multilayer perception model, radial basis network model and SVM [7].

Corresponding author Email address*: uroojshahmuhammad@yahoo.com*

In this research a training neural networks is developed by using MATLAB in order to predict student academic performance by compare two algorithms cuckoo search and gravitational search algorithms [8].

2.1 Classification Techniques

There are many classification techniques the most valuable two techniques I used in this project, the first one is decision tree and second is naïve Bayes. Decision Tree and naïve Bayes are supervised learning approach that totally worked on labeled data set and applied different models to get refined outputs which are beneficial in decision making and forecasting. Decision tree is one of the useful techniques that is applied in educational data mining. Decision tree provides a natural and human responsive description for decision makers to making further decisions [1] whereas naïve Bayes uses the Bayes probability theory which assumes the effect of parameter value of a given class is independent of the values of other parameters it represent the approach to make prediction on values of the data using known results found from dissimilar data [2].

3. Problem Statement

In order to discover hidden knowledge and information from the student's data there are some elements such as parameters, methods, techniques, tools and algorithms are need to be identified and considered in order to produce the best prediction of student's academic performance.

Mostly in all kinds of institute all the information about students is stored in a database at student's administration department so it is very difficult for lecturer to know and get knowledge about students in order to analyze their performance Moreover the work proposed earlier is applied on very small corpus.

Beside that there also must be find out the relationship between the dependent parameters and the independent parameters, the efficiency of the systems proposed also need to be improved so by making focus on these all problems a valuable prediction model can be developed that can be used by instructors to make a prediction on students' academic performance

4. Methodology

This part of the paper will present the proposed framework in which we will discuss about how a prediction model can be developed by using selected classification techniques and one of the best tool python, methodology shows some steps involved in developing model to predict SAP

4.1 collection of dataset

During the first step of collecting student's academic record so one of the suitable educational dataset has been obtained from Kaggle website [7]. The dataset contains 16 attributes with three class labels: High, Middle and Low. The dataset contains 500 cases record and was separated in to two section the first is about 70% is used to train the model. The second section is about 30% is engaged for testing the model. The training dataset used for building the model while the other was used to validate the model.

4.2 Data preprocessing

After that this data would be analyze by data selection in data selection only 5 parameters were selected for mining process then data cleaning process in this process, we rename the column name and encode our categorical column in order to remove redundancy from data.

4.3 Extract prediction model

In this stage the python software tool is used to conduct the experiment. In this experiment firstly, classification techniques one by one applied on training dataset to build the model after that developed models applied on test data set to validate it, the two classification techniques for this model are decision tree and naïve Bayes.

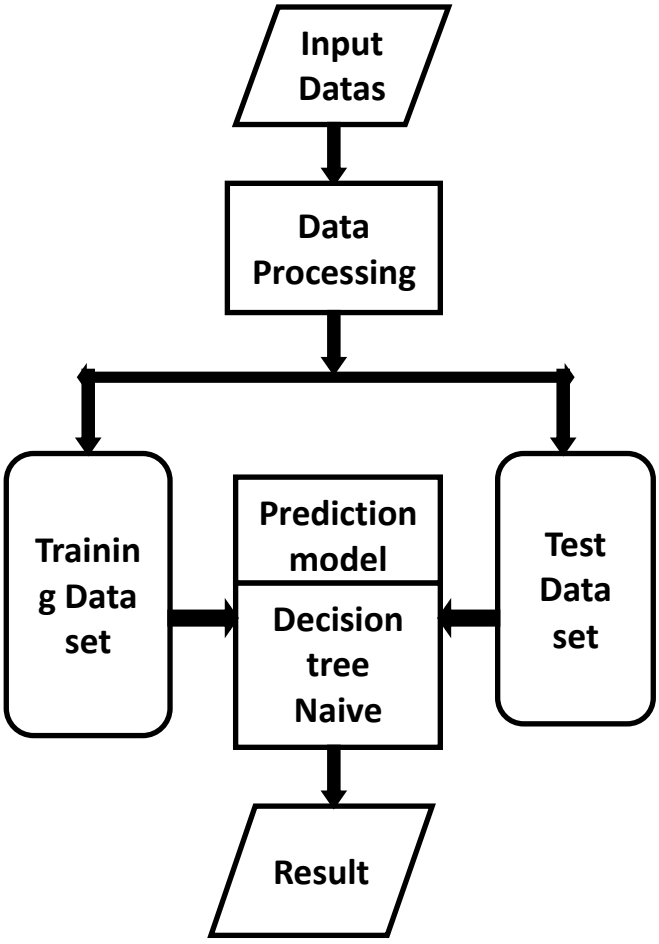

Figure.1. the Experimental system for developing prediction

This prediction model can be used by instructors to predict that either student belongs to good, average or poor category in order to improve their performance.

5. Results and Discussion

In this section the results after this whole process is represented. As in this project we make a comparison between two algorithms so once the prediction model developed from these two algorithms then results will be appear by using confusion matrix technique because confusion matrix is one of the techniques that is used to summarize the performance of classification techniques and it shows the results in some understanding format. Confusion matrix is based on some rates these are precision, recall, F1 score, support and accuracy so in the results, I have calculated all these rates with their accuracy for high, middle and low category students shown in figure:2

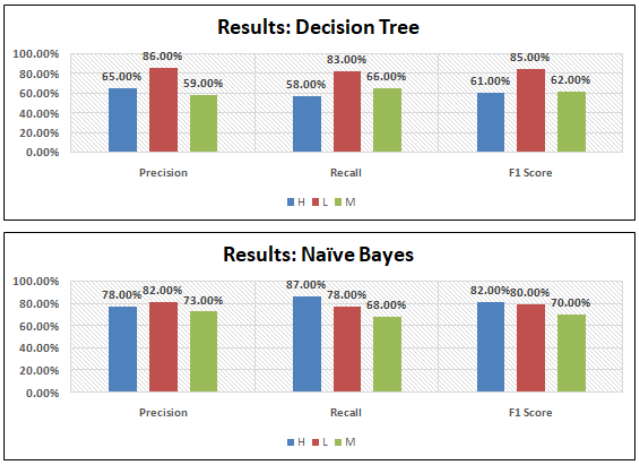

Figure.2. compare the results of two algorithms

After that we take the average accuracy of both of these algorithms in terms of high, middle and low category students. In decision tree average accuracy for high category students is 68% for low 68% and for middle 68% and in naïve bayes the average accuracy for high category students is 77% for low 77% and for middle 77%. Shown in figure:3

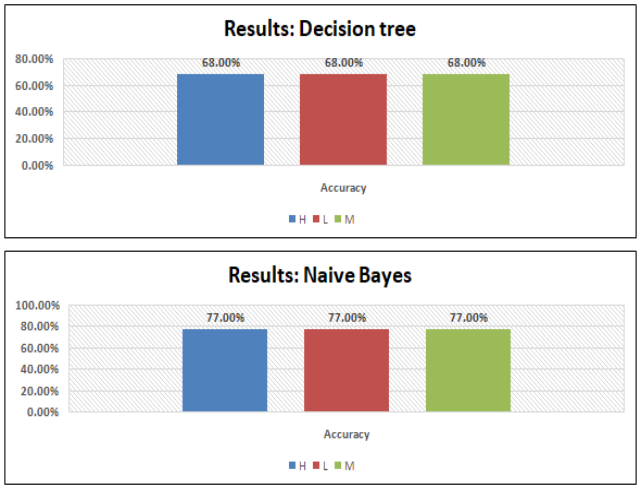

Figure 3: decision tree and naïve bayes average accuracy

Now we get the final accuracy of both of these algorithms decision tree is 68% and naive bayes is 77% so the experiment results show that naïve bayes is more accurate to develop prediction model that predict SAP

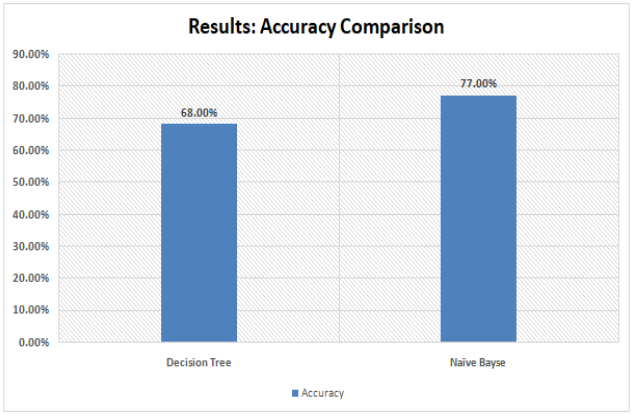

Figure 4: accuracy comparison

6. Conclusion

In educational institute lecturer face difficulty to analyze students' performance therefore in this research a prediction model can be developed by using two classification techniques these are decision tree and naïve Bayes in order to predict students' academic performance this study make a comparative analysis between these two classification techniques by using python software tool in order to find that which techniques give more accuracy result so the experiment results shows that naïve Bayes give more accuracy value then decision tree. The extracted model can predict the new student record and that is used by instructors to assist good, average or poor category students to improve their academic performance.

7. Acknowledgement

The authors pay special thanks to the Computer System Department, Mehran University of Engineering and Technology, Jamshoro, Pakistan, where this research was done.

8.References

[1] D. T. Larose, Discovering Knowledge in Data: An Introduction to Data Mining, Hoboken, NJ, USA: Wiley, 2005. http://dx.doi.org/10.1002/0471687545

[2] U. Kumar, & P. S. Pal, Data Mining: A prediction of performer or underperformer using classification, International Journal of Computer Science and Information Technologies (IJCSIT), 2 (2011), no. 2, 686–690.

[3] M.V.Jagannatha, Reddy1and B.Kavitha CSE Dept." Expert System to Predict the Type of Fever Using Data Mining Techniques on Medical Databases"

International Journal of Computer Sciences and Engineering Research Paper .Volume-03, Issue-09

[4] Ms. Priti S. Patel1, Dr. S.G.Desai "Various Data Mining Techniques used to Study Student's Academic Performance" International Journal of Computer Science and Mobile ApplicationsVol.3 Issue. 6, June-2015, pg. 55-58.

[5] Anshul Goyal and Rajni Mehta, Assistant Professor,CSE "Performance Comparison of Naïve Bayes and J48 Classification Algorithms" International Journal of Applied Engineering Research, ISSN 0973-4562 Vol.7 No.11 (2012)

[6] Ritika Saxena "Educational Data Mining Performance Evaluation of Decision Tree and Clustering Techniques using WEKA Platform" HCTL Open International Journal of Technology Innovations and Research (IJTIR) Volume 14, April 2015

[7] Shaobo Huang, Ning Fang "Predicting student academic performance in an engineering dynamic course a comparison of four types of predictive mathematical models" department of engineering education, Utah State University, Logan, UT 84322, USA vol.61 (2013) 133–145

[8] Jeng fung chen and quang hung "Training neural network to predict student academic performance a comparison of cuckoo search and gravitational search algorithm" International Journal of Computational Intelligence and Applications Vol. 13, No. 1 (2014) 1450005 (18 pages)

[9] Sevilla Bronson "Autonomy support environment and autonomous motivation on nursing student academic performance: An exploratory analysis" School of Nursing, Florida A&M University Nurse Education Today vol. 44 (2016) page 103–108

[10]Kaggle Data set of student performance prediction, https://www.kaggle.com/rmalshe/d/aljarah/xAPI-Edu-Data/student-performance-prediction/notebook [Accessed: 11 FEB, 2017]

About Authors

Urooj Makhdoom student of master engineering in computer information engineering from Department of computer system, Mehran university of engineering and technology

Dr. Prof Tariq Jamil Saifullah khanzada from Department of computer system, Mehran university of engineering and technology.