International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 12, ISSN No. 2455-2143, Pages 99-102
Published Online October-November 2016 in IJEAST (http://www.ijeast.com)

STUDY ON STUDENT PERFORMANCE PREDICTION

Pankit Arora
Department of IT
Amity University, Noida, India

Akshath Dhar
Department of IT
Amity University, Noida, India

Abstract— **Data mining is the way towards mining the definite information from a database . It can be said that data mining is mining the knowledge from the data. A database is accumulation of interrelated information. The performance of the student depends on the training and preparation provided by the teacher. J48 characterizes the straightforward C4.5 choice tree for characterization. The binary tree is a decision tree approach is most useful in classification problem. The present framework utilizes C4.5, Naïve Bayes and ID3 calculation for dissecting the execution of the understudies is utilized for characterizing the understudies into low and high class. This paper gives relative investigation of foreseeing understudy level execution utilizing arrangement calculation, for example, Naïve Bayes, C4.5 and ID3.**

Keywords— **ID3, C4.5, Naïve Bayes, Data Mining**

I. INTRODUCTION

Through the process of Data mining, the required data is explored from large amounts of data- business or market related- also known as the "Big Data". By applying the detected patterns to new subsets of data, the user search for reliable patterns or efficient relationships between variables.

There are many classification algorithms used for classifying the data. The classification algorithms can be classified into four categories namely, Basic learning/mining tasks deals with the mining of data from the database based on query of search and Inferring rudimentary rules provides for mechanism that generates rules by concentrating on specific class at time. Decision tree learning model maps the observation about the item and provide conclusion about the target value and the Covering algorithm Remove positive examples covered by this rule. This paper used for classifying the student dataset into high and low performance students. This paper uses decision tree algorithm for classifying the student based on their performance.

II. RELATED WORK

A large portion of the prior specialists have analysed the performance of the students taking into account their tenth , twelfth marks, and on living area, medium of instructing, family yearly wage and student's family status, and so on.

With Big Data Using C4.5 and Bayesian Classifier compares C4.5 and Bayesian Classifier algorithm using the performance of the students. Comparison of these two algorithms and classified the data set into different classes and in different phases.

Naive Bayes algorithm provides best accuracy level 81% compared to C4.5 algorithm

We use ID3 algorithm for classifying the dataset. This paper used dataset obtained from a website. We analyze the performance of the students based on their previous year marks, seminar performance, assignment and end semester marks. From this paper, the students with low performance can be easily identified and high concentration can be provided in order to improve the performance level of such students

Classification Model of Prediction for Placement of Student:
Apply various classification algorithm with Naive Bayes, MLB, and J48 for analysis the student's academic performance for Training and placement. This model determines the relations between academic achievement of the students and their placement in campus selection. J48 algorithm gives the best accuracy level of 86.15% then other classification algorithms.

We have used decision tree algorithm for prediction of students academic performance in higher education. These learnings will be helpful to identify the students. Different decision tree algorithm are used J48, NB tree and simple card algorithm. J48 method identify two factors ―weak and success students. It gives the best accuracy level of 80.15%

A prediction for performance improvement using classification:
Raw data was pre-processed in terms of filling up lost ethics, transform ethics in one shape into one added and relevant attribute variable selection.

By utilizing the Bayesian classification technique on the student database, we can predict the student division on the premise of earlier year database. With the help of this study, both the student and the teacher can enhance the division of

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 12, ISSN No. 2455-2143, Pages 99-102
Published Online October-November 2016 in IJEAST (http://www.ijeast.com)

the student. The decision tree algorithms such as ID3,C4.5 and CART are applied on the student's data to foresee their performance in the final examination. This paper generate the result of the decision tree predicted the number of students and the C4.5 algorithm gives the highest level of accuracy 67.77% compared to other classification.

Foreseeing students' performance with Id3 and c4.5 using classification algorithms :
Break down the data set containing data concerning student, for example, sexual classification, the grade scored in their board examinations of higher classes, the position achieved by them in their placement test and the grades in the first year of the past batch of student. Compared to other classification algorithms ID3 (Iterative Dichotomies 3) produces better accuracy level is 67% and C4.5 categorization algorithms

Decision tree algorithm and Naive Bayesian Classifier algorithm are applied on pre-processed student data to reveal classification accuracy between 93.33 % and 71.67 %

The highest accuracy is accomplished for the Decision Tree model (93.33%).Brawny class is predicted with the higher precision using the decision Tree model, while the scrawny class is predicted with the help of other three models . The information ascribe associated with the students' personal information and under graduation information are among the components impacting most the classification process.
Using the Data Mining Techniques ,study on Student Data Analysis: explores the socio-demographic variables times, masculinity, name, lesser class rating, upper class blot,
Degree expertise ,etc. This paper used classification algorithm to categorize the level of students. To discover the data source distribution of information, as well as other data mining algorithm as a pre processing rung, a separate tool known as the clustering analysis is used. This groups the students according to their grade and proficiency.

Using the student's qualified data, an analysis on performance of decision tree algorithms: This paper compares the ID3, C4.5 and CART algorithm. The performance based on Parent qualification, Living Location and Economic Status, Friend and Relative Support, Attendance Result. CART shows the best classification accuracy when compared other classification. It produces the highest accuracy level of 55.83%.

III. PROPOSED METHODOLOGY

Decision tree learning utilizes a decision tree as a prescient model which maps perceptions around an item to decisions about the item's target value. Tree shape where the aim variable can take a limited arrangement of qualities are called classification trees.

In these tree structures, leaves represent class labels and branches represent conjunctions of elements that lead to those class names. Decision trees where the target variable can take consistent qualities are called regression trees. The proposed system uses random forest algorithm which is among decision tree to classify the student dataset into high and low category. The uploaded dataset of the student is classified based on the performance of the student.
An issue with Naive-Bayes is that it has no occurrences of a class label and a certain attribute value together then the frequency-based probability estimate will be nothing. Agreed Naive-Bayes' provisional independence supposition, when all the probability are multiply it will get zero and this will affect the posterior probability estimate.

The major nuisances of the C4.5 algorithm are as follows:
Values generated using this algorithm neither contribute to generate rules nor help to construct any class for classification task. It crafts the tree bigger and more complex. Branches reduce the usability of decision trees.
Some of the issues ID3 algorithm is time complexity is high when compared to other algorithms. It cannot be provide with exact accuracy of classified students. This problem happens when samples are being drawn from a population and the drawn vectors are not fully representative of the inhabitants. For forecast, another example is pushed down the tree. It is dole out the mark of the planning test in the terminal node it ends up in. This process is iterated in abundance of all foliage in the band, and the standard vote of all trees is report as random forest prediction.

Decision tree Algorithm Algorithm
Decision Tree Algorithm are an outfit learning methodology for course of action, backslide and distinctive errands, that limit by building countless choice trees at preparing ready time and yielding the class that is the strategy for the classes (classification) or mean desire (regression) of the individual trees. Random forest correct for decision trees propensity for over fitting to their preparation set. It is a standout amongst the most precise learning calculations accessible for some information set as it delivers a very exact classifier. Random forest classifier used for number of decision trees in order to improve the classification rate. This method combines the Breiman's in ―bagging idea and the random selection features.

It is a group of classifier that consists of many decision trees and outputs the class that is mode of the class output by individual trees.
Decision trees are individual learners that are combined for one of the popular learning methods commonly used for data exploration. To improve the performance, ensembles are used that is a divide and conquer approach.

International Journal of Engineering Applied Sciences and Technology, 2016
Vol. 1, Issue 12, ISSN No. 2455-2143, Pages 99-102
Published Online October-November 2016 in IJEAST (http://www.ijeast.com)

All classifiers taken together are a strong learner while taken independently is a weak learner. Random forest techniques examines a large ensemble data set .It is first generating a random sample of the original data with replacement and a user defined number of variable selected at random from all the of the variables to determine node splitting.
On entering a new input into the system, it is run down all the trees. The result may either be an average or weighted average of all the terminal nodes that are reached. Each of the trees is grown to the largest extent possible .

Decision Tree Prediction

$$s= \frac{1}{k} \sum_{k-1}^{k} k^{th}$$

The average of the predictions of the tree is taken to be the prediction of the random forest. Where the index K run over the individual trees in the forest.

This algorithm run time is fast efficiently on large data base they are able to deal with unbalanced and missing data.Without deleting any variable thousands of input variable can be handled. For estimating missing data and maintaining accuracy, this method is effective.

Data Selection

This paper uses data sets that consists of 1000 students. The performances of these students are analyzed based on their assignment mark, seminar mark, internal marks and their external marks. The random forest algorithm is applied to the dataset collected from the above college.
The variables, description and possible values of the variables are listed below in the table.

Table. I. Variables and Description

Vairable	Discription	Possible Value
AM	Assignment marks	<5
SM	Seminar marks	<10
IM	Internal marks	<50
SEM	Semester marks	<100
CFAC	College Facility Internet Facility	Brilliant , superior, reasonable

If the low performance students are high the reason for low performance is taken from the Test and CFAC. This helps to improve the performance of the students in future.

IV. CONCLUSION

The accuracy level of accessing algorithms such as ID3, Naïve Bayes (NB) and C4.5 are compared. The ID3 algorithm provides 67% of accuracy when comparing the student's concert. The Naïve Bayes algorithm provides 81% of accuracy level in and C4.5 provides 75.145%.

The accuracy level for the accessing algorithm is provided in the figure:

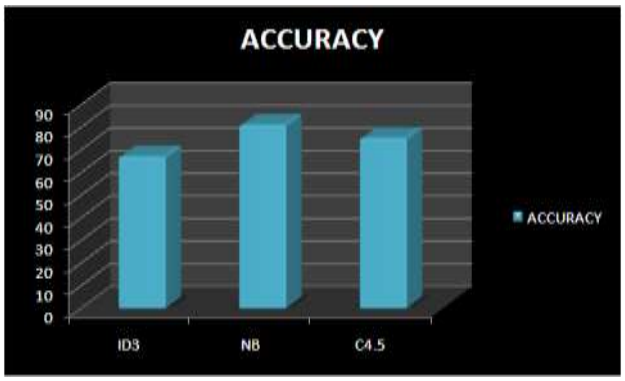

Figure.1. Performance accuracy

The accessing algorithms such as ID3 (Iterative Dichotomies 3), Naïve Bayes, C4.5 algorithm provided classification accuracy based on the provided dataset. This paper compares the Naïve Bayes, ID3, C4.5 algorithm and aims at displaying that random forest algorithm performs better than the other classification algorithm with more than 81% of accuracy.

V. REFERENCE

[1] Bharti Thakur, ―Data Mining With Big Data Using C4.5 and Bayesian Classifier‖, , International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 8, August 2014.

[2] Brijesh Kumar Baradwaj, Saurabh Pal ―Mining Educational Data to Analyze Students Performance‖. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011.

[3]Ajay Kumar Pal, Saurabh Pal, I.J.Modern ―Classification and Computer Science, 2013, 11, 49-56, Published Online November 2013 in MECS.

[4] Mrinal Pandey and Vivek Kumar Sharma in ―Data Mining: A Decision Tree Algorithm Pertaining to the Student Performance Analysis and Prediction‖. International Journal of Computer Applications (0975 – 8887) Volume 61– No.13, January 2013

[5] Brijesh Kumar Bhardwaj, ―Data Mining: A prediction for performance improvement using classification**",** International Journal of Computer Science and Information Security, Vol. 9, No. 4, April 2011**.**

[6] Surjeet Kumar Yadav ―Data Mining: A Prediction for

Performance improvement of Engineering Students using Classification‖, World of Computer Science and Information Technology Journal (WCSIT) ISSN: 22210741 Vol.2,No.2,51-56,2012**.**

[7]Kalpesh Adhatrao, ―Predicting students' performance using Id3 and c4.5 classification algorithms‖, International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.3, No.5, September 2013.

[8] Dr. A. Padmapriya ―Prediction of Higher Education Admissibility using Classification Algorithms‖,International Journal of Advanced Research in Computer Science and Software Engineering.

[9] Umamaheswari. K, S. Niraimath in ―A Study on Student Data Analysis Using Data Mining Techniques‖, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 8, August 2013.

[10] T.Miranda Lakshmi, A. Martin, R.Mumtaj Begum and Dr.V.Prasanna Venkatesh ―An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data‖.I.J.Modern Education and Computer Science, 2013, 5, 18-27 Published Online June 2013 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijmecs.2013.05.03