International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 13, 2017 ISSN: 2348-6600
Reference ID: IJCS-217 PAGE NO: 1354-1358
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

A REVIEW ON PREDICT THE STUDENT PERFORMANCE USING CLASSIFICATION METHODS IN DATA MINING

C.Francytheresa[1], Dr. G.Kesavaraj[2],
[1]M.Phil Full time Research Scholar, PG and Research Department of Computer Science,
[2]Head of the Department, PG and Research Department of Computer Science
Vivekanandha College of Arts and Sciences for Women (Autonomous),
Tiruchengode,Tamilnadu,Namakkal-637205.
francyjaya@gmail.com[1]
kesavaraj2020@gmail.com[2]

ABSTRACT- Nowadays, student database have large amount of datasets. Mining student database is important to predict the student who are at risk and then train the students to improve their performance. Predict student performance is a basic need of higher educational institutions. A criterion of the best educational institution is based on its excellent academic achievements. Predict the student performance using data mining techniques are classification, clustering. Compare classification and clustering, classification techniques are gave high accuracy. In this paper, I have discussed about various classification algorithms are highly used in predict the student performance.

Keyword: Data mining, Educational Data Mining, Classification methods.

I.INTRODUCTION

We are living in the information age" is a popular saying; however, we are actually living in the data age. Tera bytes or peta bytes of data pour into our computer networks, the world wide web (www), and various data storage devices everyday from business, society, science and engineering, medicine, and almost every other aspect of daily life [1]. Extract knowledge from the data is called "data mining". It's also called as KDD (Knowledge Discovery from Data). The data mining applied to Educational data base is called Educational Data Mining. The following are the useful tools for Educational Data Mining,

Table.1 Tools used for EDM

The main objective of this paper is to use data mining methodologies to study student's performance. In this review, we compare mainly used two classification methods namely decision tree, naive base used to predict the student performance

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 13, 2017 ISSN: 2348-6600
Reference ID: IJCS-217 PAGE NO: 1354-1358
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

Name of Tool and Developer	Function/ Features	Techniques /Tools	Environments
Intelligent Miner (IBM)	Provides tight integration with IBB's DB2 relational db system, Scalability of Mining Algorithm	Association Mining, Classification, Regression, Predictive Modelling, Deviation detection, Clustering, Sequential Pattern Analysis	Windows, Solaris, Linux
MSSQL Server 2005 (Microsoft)	Provides DM functions both in relational db system and Data Warehouse (DWH) system environment.	Integrates the algorithms developed by third party vendors and application users.	Windows, Linux
Oracle Data Mining (Oracle Corporation)	Provides an embedded DWH infrastructure for multidimensional data analysis	Association Mining, Classification, Prediction, Regression, Clustering, Sequence similarity search and analysis.	Windows, Mac, Linux
WEKA (University of Waikato, New Zealand)	Provides machine learning algorithms for data mining tasks. Well-suited for developing new machine learning schemes.	Data pre-processing, classification, regression, clustering, association rules, and visualization.	Windows, Linux

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 13, 2017 ISSN: 2348-6600
Reference ID: IJCS-217 PAGE NO: 1354-1358
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

I.LITERATURE SURVEY

Data Mining (sometimes called data or knowledge discovery) is the process of analysing data from different perspectives and summarizing it into the useful information – information that can be used to increase revenue, cuts costs, or both. Data Mining software is one of a number of analytical tools for analysing data. It allows users to analyse data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databaseFollowing are the survey papers being studied:

I.METHODOLOGY

Data Collection:

Collect the sample student database from the Higher educational institutions' or Universities which data you want to mine.

Data Pre-processing:

Data mining is also called as Knowledge discovery from data (KDD). The knowledge discovery process consists a sequence steps they are

Figure.1.

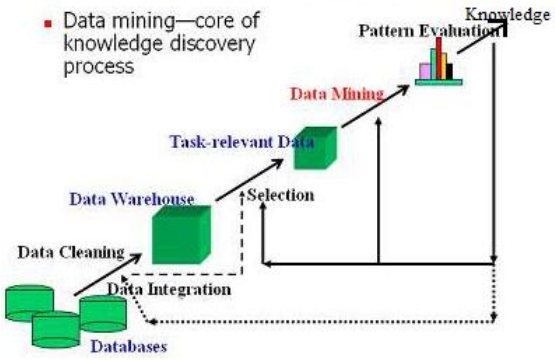

Figure 1 Knowledge discovery process

The first four steps in knowledge process is called the data pre-processing. In those steps data are cleaned and ready to apply data mining techniques. They are,

- Data cleaning: to remove noise and inconsistent data
- Data integration: when multiple data sources may be combined
- Data selection: where data relevant to the analysis task are retrieved from the database
- Data transformation: where data are transformed and consolidated into forms appropriate for mining by performing summary or aggregation operation

Data mining techniques:

An essential process where intelligent methods are applied to extract data patterns. In this study we see the two data mining techniques are decision tree and naive bayse classification algorithm.

Predict the student performance using decision tree:

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 13, 2017 ISSN: 2348-6600
Reference ID: IJCS-217 PAGE NO: 1354-1358
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

Decision Tree is one of a popular technique for prediction. Most of researchers have used this technique because of its simplicity and comprehensibility to uncover small or large data structure and predict the value [4,5,6]. Romeroet al. (2008) said that the decision tree models are easily understood because of their reasoning process and can be directly converted into set of IF-THEN rules.

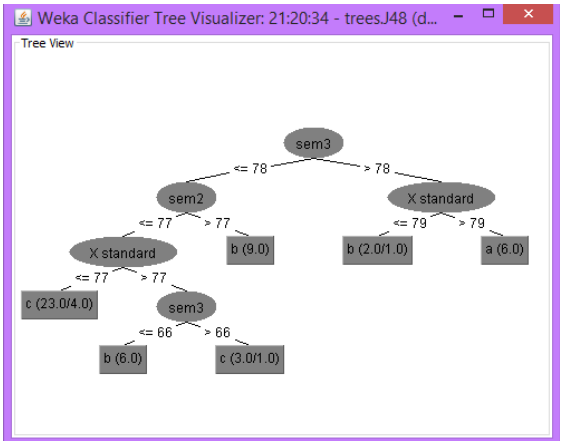

Figure 2 Decision tree Visualization

As in this case all the datasets were initially classified and then the results are passed through the decision trees algorithm in fig.3, and hence the results are tabulated through visualizing the tree. As in decision tree we can see that an accuracy of 87.751% is achieved.

Four attributes:

- Attendance at virtual lectures (nominal attribute)
- Take online quizzes (nominal attribute)
- Student gender (categorical attribute)
- Exam mark – final grade (categorical attribute)

The Naïve Bayes classification algorithm will calculate the probability of the final exam mark (final grade = A or B or C or D, etc.) in each class (i.e. from the number of times which the student attended an online lecture, or took an online quiz). For instance: X = A, B, C, D, or F and

C = 1, 2, 3, 4, 5, 6

X is a set of events C is a set of classes

To compute the probability that a student record will experience event X = A (student gets an A), IF s/he attends a lecture just once (i.e. C = 1), we use the Bayes theorem.

The data mining process in RapidMiner conducted in this study has many steps figure.4.

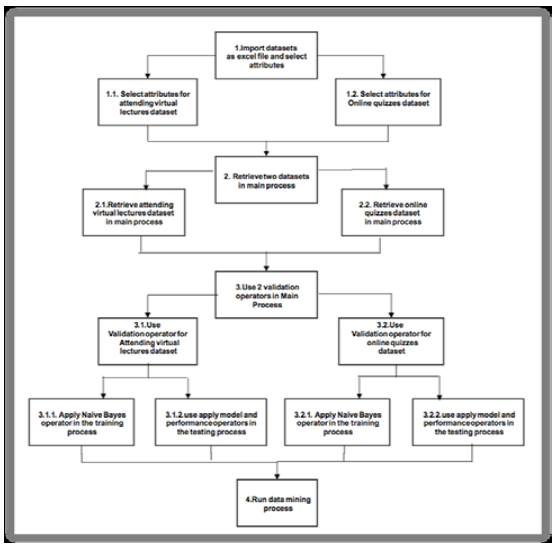

Figure 3 Data Mining Process using Naive Bayse

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 13, 2017 ISSN: 2348-6600
Reference ID: IJCS-217 PAGE NO: 1354-1358
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

I.RESULTS AND DISCUSSION

This section will discuss the results analysis of the recent works in predicting the student's performance. This meta-analysis is based on the highest accuracy of prediction methods and also the main also the attributes that may influence the student performance. Comparison of decision tree and naive base algorithm, decision tree gave the highest accuracy as shown as above (87.751%) and the naive bayse algorithm gave the high accuracy of (76%) with nominal and categorical attribute.

CONCLUSION AND FUTURE WORK

Educational data mining (EDM) is an area full of exciting opportunities for researchers and practitioners. This field assists higher educational institutions with efficient and effective ways to improve institutional effectiveness and student learning.

REFERENCES:

[1]. Jiawei Han, Micheline kamber, Jian Pei, "Data mining : Concepts and techniques", 3rd edition, 2012.

[2]. M.of Education Malysia, National higher education strategic plan (2015).

[3]. Osman Gursoy, Mehmet Akif Yaman, Emine yaman, "Performance analysis on student's GPAs and Course averages using data mining techniques", Research gate, conference paper January 2011.