Survey on Different Classification Approach for Analysis of Students' Performance

Varsha Choudhary[1] Prof. Mohit Jain[2]

[1,2]Department of Computer Science
[1,2]BM College, Indore, India

Abstract— The assessment in outcome based learning is very vital and significant approach Toward measuring the student's performance. There are many traditional methods existing in this context. The data mining is one of the intelligent computing methods which are having widely accepted features that enable the idea of its usage in Assessment. Much work has been done to measure the student performance by using Different methodologies and modern technologies. In this work, we have gone through the current datasets of students of the university and different classification methods of data mining are used to measure the accuracy of student performance. Based on the Analysis of the result, it has been concluded that accuracy and the other measures of SVM is more than the other classification methods.

Key words: Student Performance

I. INTRODUCTION

Higher education has gained importance manifolds in the past few decades. The higher educational institutes are forced to revise its scope and objects because of the private participation. The controller of regulatory body has put some guidelines with regard to infrastructure, faculty and other resources. New technologies are being developed in the field of data management and analysis due to large supply of data being present in several companies, including both private and public. The main aim of the techniques of data mining is to discover hidden and insignificant links within the information having diverse characteristics. Various techniques of data mining are being used in different fields including the educational environment. A very encouraging area to attain this objective is the usage of Data Mining (DM) [1]. In fact, classification is one of the most helpful DM work in e-learning. Data mining has been executed well in the business applications, but its use in higher education and higher learning institutions is still relatively new.In the sector of education, educational data mining proves to be an emerging practice which is very recent and its practice is preconceived to identify and extract new and valuable knowledge from the data .

The aim is to resolve problems of research areas of education and improve the whole educational process using various statistical techniques, machine learning programming (MLP) and data mining algorithms. Educational data Mining (EDM) is a prospering practice that can be used for analytics and visualization of data, prediction of student performance, student modeling, grouping of students etc. Educational Data Mining is focused on developing methods to explore the unique and increasingly large dataset which arrives from educational sources and further employing those methods to understand the students and the environment in which they learn in a better way. Educational Data Mining (EDM) is the process to convert raw data from education systems to beneficial information which can be further bemused by parents, teachers, educational developers, other educational researchers and students.

II. RELATED WORK

Dhahiri et al.[10] have also provided an over view on several techniques of datamining that were applied to predict and analyze performance of students, concentrating on the identification of most valuable attributes in a student's data by employing the prediction algorithm. They provide a systematic literature review to improve the student's achievements by using the techniques of data mining. The various analytical methods used cumulative grade point average (CGPA) as their data sets, thus helping the system of education to monitor the performance in a very systematic way.

Osmanbegovic and Suljic [11] applied three supervised data mining algorithms to assess the data of first year students to predict favorable outcome in a course and evaluating the performance based on certain factors like convenience, accuracy and approach of learning. A very high emphasis is given on some socio-demographic factors, high school results obtained, attitude towards study and marks in entrance examinations. The whole data was collected from University of Tuzla, academic year2010-2011. The authors believe that exams play a very important role to determine the future of the students, in addition to the internal assessments. They used WEKAfor their study and implemented it in java and also conducted four tests to assess the input variables: Info Gain test, Chi-square test, Gain Ratio test and One R-test. A model has also been developed based on some selected input attributes assembled through questionnaire method

Ramesh et al. [12].conducted a survey cu experimental methodology to generate database for the students for predicting the performance. The three main objectives were to identify the essential predictive variables on higher secondary students, know the best classification algorithm and to predict the grade at higher examinations. The study shows that parent's occupation plays a major role and not the type of school in predicting the grades. The data for the study was collected from schools and internet and the authors found out that multilayer perceptron algorithm is the best one for grade prediction. This algorithm is more efficient showing the accuracy of 72%.

Goga et al. [13] designed a tool by using .NET framework to predict student's grade by providing various parameters as input. Models based on the student's enrollment records were developed by using ten classifications trees (OneR, Random forest, ZeroR, random tree, Decision stump, REPTree, JRip, , J48, PART, and Decision table) and a multilayer perceptron (Artificial Neural

Network) learning algorithms by operating on WEKA(Waikato Environment for Knowledge Analysis). A framework is designed for intelligent recommender system which recommends suitable actions for improvement. The work is based on the background factors that predicts tertiary first year academic performance of the students. The data for the study is taken from Babcock University, Nigeria. The background factors for the students were collected through in-depth interview. The various demographic factors are father occupation, mother occupation, family income, place of birth, family size, academic qualification of parents, parent's marital status. The benchmarks used in the comparison of the generated models include confusion matrix, accuracy and speed. Random tree outperformed the other algorithms in terms of benchmarks. Therefore, random tree is adopted as the best algorithm in the domain of this study to serve as a building block for designing a generic system.

R.Campagni et al. [14] Presents the methodology to determine the future career of university graduate students. The main aim is to identify the strategy to improve the performance and scheduling of the exams by using various approaches of datamining. "Ideal career" is introduced which is basically the career of an ideal students who has given the exam just after the end of studying the particular course i.e. giving exams without any delay. The methodology is applied to a real case study and it has been observed that the performance, in terms of final grade and graduation time, increases manifolds if the student follows the order as given by the ideal career. Bubble sort distance is used to calculate the career distance between the normal student and the student following the ideal career. The practical implementation of the obtained results has been used for the students enrolled in the Computer Science and engineering department at the University of Florence, helping them to improve.

III. EXISTING WORK

A model is proposed along with an algorithm to predict the performance of students. Data pertaining to the study were collected from the same institution for which the performance chance prediction and percentage student analysis performance need to be found from 2006to 2015. Data collected is divided into historic data from 2016to1014 and test data.DM classification techniques to predict the performance of students. We applied three classifiers (Naive Bayes, Decision Tree and K-NN) and found that DT classifier gives the best results when used with students' data (social an academic attributes).

IV. PROBLEM STATEMENT

Examination has an important role in the life of students. The result decides the future of the students according to the marks obtained in examination. Therefore, prediction of student's result, pass or fail, in any examination becomes very vital. More efforts can be taken to improve the studies and the performance if the student is expected to not perform well. This will help the students to pass the examination and improve performance [40]. Although many studies are being carried out on the prediction of student performance, but very few studies focus on investigating how the performance of students evolves during their course of study. Most of the approaches have used only the factors like demographic factors, academic marks as their basis of prediction. Very little work is done which take the grades of each semester into the account and work solely on the grades obtained by the students.

V. OBJECTIVES

The main purpose of any institution is to improve the quality of the education and to impart managerial decision which would be beneficial to the society. Good measurement and prediction of students is one way to reach the highest quality leveling higher education system. The main objective of this research work is to find the methods which can best measure the academic performance of the students in university. We have employed data mining techniques for the following purpose:
1) To predict the final grade of the students by using the obtained grades in four semesters.
2) Improving or discovering the main attributes which can be useful for Measuring the performance.
3) Studying about the various algorithms which uses the attribute's values at it's Best and thus performs better in terms of accuracy.
4) Learning the classification algorithms in detail so that the best model is Discovered and attained to measure the performance of the students.

VI. PROPOSED ALGORITHM

A. Classification

A data mining approach that allows items in a collection to target categories or classes is known as classification. The chief goal of classification is to predict the target class for each case in the data accurately. Eg: a classification model could be useful form ensuring the performance of the students to be high, medium or low. It is considered to be the "best-understood" technique among all data mining approaches. Classification task initiates building data for which the target values or class Assignments are known.
1) Naive Bayes Adaptive
2) Support Vector
3) Bayes Network

B. Regression

Statistical Regression is one of the predictive data mining model that analyzes the dependency within attribute values, that are dependent on the values of other Attributes. This is the main difference between regression and classification. In other Way, target attribute containing continuous (or floating-point) values requires a Regression technique. The most commonly used regression type is linear regression. In

This, the line that minimizes the average distance among all the points from the line
I.e. that best fits the data is calculated.

C. Time-Series Analysis

Time-series database includes sequences of values or events acquired within repeated measurements of time. It is a sequence database having values that are typically measure at equal time interval such as weekly, hourly, daily. Time series analysis is the process of analyzing time series data for extracting meaningful statistics and other characteristics of the data. This time series obtains the represents sequential measurements by collecting the values. It can be helpful for observing natural phenomena like wind, earthquake, treatments, atmosphere and temperature.

D. Descriptive Data Mining Model

This approach for mining employs techniques of clustering, association rules meninges. to find patterns that are covered in large data set and further aid in intelligent decision-making. As the name implies, these models "describe" or summarize raw data. They are the models that describe the past and are interpretable by humans. Descriptive data mining are useful because they allow learning from past behaviors and understand how they might influence future outcomes. It can also be used to find relevant subgroups in the bulky data. The various descriptive data mining approaches are as described below:

1) Clustering

Clustering is one of the most important descriptive data mining models. Clustering is the process of discovering natural groups (or clusters) in a database. The data items in the set in a cluster have similar characteristics. The goal of clustering is to find clusters of high quality so that the intra-cluster similarity is high and inter-cluster similarity is low. Clustering models divide the data into groups that were not defined before. Clustering is useful for exploring data, for anomaly detection, to find natural groupings and do not use a target.

2) Summarization

The other names of summarization are abstraction or generalization of data. The aim of this technique is to map the data into subsets having simple descriptions. The data after summarization gives overview of the data with aggregated information. As the name suggests, it is that concept of data mining which involves the concept of finding compact description of dataset. Summarization can be viewed from different angles and can be scaled up to different levels of abstraction. The main approaches of summarization are standard deviation, variance, mean, median, tabulation and mode. The applications are usually involved in data visualization, data analysis and automated report generation.

3) Association

To find relationships between attributes and items, associations or link analysis technique is used. Association rule is one of the important techniques for market basket analysis. This is so because all possible combinations of product outings can be explored. Therefore, it can be easily used to establish statistical relationships among various interdependent variables of a model. By using if/then statements, association rules helps to uncover relationships among unrelated data in various databases such as transaction oriented database, information repository or other relational database. Association rules can be used to analyze and predict the behavior of consumers, catalog design, basket data analysis, store layout and product Clustering.

VII. CONCLUSION

This work would be done on factual and real data. From the above analysis, The proposed methodology can be adopted to help the teachers as well as the students to enhance the quality of learning and student's performance by taking significance Decision at right time.

VIII. REFERENCES

[1] C. Romero, S. Ventura, "Educational data mining: a survey from 1995 to 2005", Expert Systems with Applications, vol. 33, no.1, pp.135-146, 2007.

[2] D. P. Nithya, B. Umamaheswari, A. Umadevi, "A survey on educational datamining in field of education, "International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol. 5, no. 1, pp. 69–78, Jan.2016.

[3] E. Osmanbegovic, M. Suljic, "Data mining approach for predicting student performance", Economic Review – Journal of Economics and Business, vol. 1,no. 1, pp. 3-12, 2012.

[4] R. Srivastava, M. Gendy, M. Narayana, Y. Arun, J. Singh, University of the future — "A thousand year old industry on the cusp of profound change", Melbourne, Australia: Ernst & Young (Retrieved fromhttp://www.ey.com/Publication/vwLUAssets/Univ ersity of the future/$FILE/University of the future 2.12.pdf), 2012.

[5] Y. Ma, B. Liu, C. Wong, P. Yu, S. Lee, "Targeting the right students using datamining", in: Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data mining, pp. 457–464, 2000.

[6] D. Suthers, K. Verbert, E. Duval, X. Ochoa, "Clow, MOOCs and the funnel of participation", In: International Conference on Learning Analytics and Knowledge, pp. 185–189. ACM New York, 2013.

[7] Suthers, K. Verbert, E D. Silva, M. Vieira, "Using data warehouse and datamining resources for ongoing assessment in distance learning", In: IEEE International Conference on Advanced Learning Technologies, pp. 40–45, 2002.

[8] J. Anderson, A. Corbett, K. Koedinger, "Cognitive tutors", J. Learn. Sci., vol.4, no.2, pp.67–207, 1995.

[9] P. Brusilovsky, C. Peylo, "Adaptive and intelligent web-based educational systems", Int. J. Artif. Intell. Educ. 13, vol.2, no.4, pp. 159–172, 2003.

[10] A. M. Shahiri, W. Husain, N. A. Rashid, "A review on predicting student's performance using data mining techniques, "Procedia Computer Science, vol.72, pp. 414–422, 2015.

[11] E. Osmanbegovic, M. Suljic, "Data mining approach for predicting student performance", Economic Review – Journal of Economics and Business, vol., no.1, pp. 3-12, 2012.

[12] V. Ramesh, P. Parkav, K. Rama, "Predicting student performance: A statistical and data mining",

International Journal of Computer Applications, vol. 63, no.8, pp. 35-39, 2013.

[13] M. Goga, S. Kuyoro, N. Goga, "A Recommender for improving the student academic performance" ,Procedia - Social and Behavioral Sciences, vol. 180,pp. 1481–1488, May 2015.

[14] R. Campagni, D. Merlini, R. Sprugnoli, and M. C. Verri, "Data mining models for student careers, "Expert Systems with Applications, vol. 42, no. 13, pp.5508–5521, Aug. 2015.