International Journal of Scientific Research and Management Studies (IJSRMS)
ISSN: 2349-3771 **Volume 1 Issue 5, pg: 156-161**

AN APPROACH ON EARLY PREDICTION OF STUDENTS' PERFORMANCE IN UNIVERSITY EXAMINATION OF ENGINEERING STUDENTS USING DATA MINING

Dineshkumar B. Vaghela[1], Priyanka Sharma[2]
[1]PhD Scholar, [2]PhD Guide,
Gujarat Technological University, Chandkheda, Gujarat, India

ABSTRACT

Data mining, also popularly known as Knowledge Discovery in Database, refers to extracting or "mining" knowledge from large amounts of data. Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. Many application areas such as medical, research, stock market, weather forecasting, business strategies…etc data mining is very much helpful to gain the hidden and useful information. Nowadays the universities also have been started to use the data mining in-order to achieve highest quality in teaching. One way to achieve highest level of quality in higher education system is by discovering knowledge for prediction regarding enrolment of students in a particular course, alienation of traditional classroom teaching model, detection of unfair means used in online examination, detection of abnormal values in the result sheets of the students, prediction about students' performance and so on. The present paper focuses on the prediction of students upcoming university examination with the help of certain key attributes related with the students' performance. In this paper the decision tree of classification method is used which helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising/counseling.

KEYWORDS*-Educational Data Mining (EDM); Classification; Knowledge Discovery in Database (KDD); ID3 Algorithm.*

I. INTRODUCTION

The extensive usage of information technology in different areas leads the generation and collection of large volumes of data storage in different formats like records, files, documents, images, sound, videos, scientific data and many new data formats. For better decision making, the data collected from large repositories of different applications require proper method of extracting knowledge. The Data mining of KDD process aims at the discovery of useful information from large collections of data [1]. The main functions of data mining are applying various methods and algorithms in order to discover and extract patterns of stored data [2]. Data mining and knowledge discovery applications have got a rich focus due to its significance in decision making and it has become an essential component in various organizations. Data mining techniques have been introduced into new fields of Statistics, Databases, Machine Learning, Pattern Reorganization, Artificial Intelligence and Computation capabilities etc.

There are increasing research interests in using data mining in education. This new emerging field, called Educational Data Mining, concerns with developing methods that discover knowledge from data originating from educational environments [3]. Educational Data Mining uses many techniques such as Decision Trees, Neural Networks, Naïve Bayes, K- Nearest neighbor, and many others.

Using data mining techniques many kinds of knowledge can be discovered such as association rules, classifications and clustering. The discovered knowledge can be used for prediction regarding

enrolment of students in a particular course, alienation of traditional classroom teaching model, detection of unfair means used in online examination, detection of abnormal values in the result sheets of the students, prediction about students' performance and so on.

The main objective of this paper is to use data mining methodologies to study students' performance in the courses. Data mining provides many tasks that could be used to study the student performance. In this research, the classification task is used to evaluate students' performance and as there are many approaches that are used for data classification, the decision tree method is used here. Information like Attendance (ATT), Class Test Grade (CTG), Mid Semester Examination Marks (MSM), Seminar Performance (SEM), Practical Evaluation (PEV), Previous Semester Marks (PSM), Lab work and Assignment Marks (ASS) were collected from the student management system, to predict the performance at the end of the semester. This paper investigates the accuracy of Decision tree techniques for predicting student performance.

II. RELATED WORK

Data mining in higher education is a recent research field and this area of research is gaining popularity because of its potentials to educational institutes. Data Mining can be used in educational field to enhance our understanding of learning process to focus on identifying, extracting and evaluating variables related to the learning process of students as described by Alaa el-Halees [4]. Mining in educational environment is called Educational Data Mining.

Pandey and Pal [5] conducted study on the student performance based by selecting 600 students from different colleges of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayes Classification on category, language and background qualification, it was found that whether new comer students will performer or not.

Galit [6] gave a case study that use students data to analyze their learning behavior to predict the results and to warn students at risk before their final exams.

Al-Radaideh, et al [7] applied a decision tree model to predict the final grade of students who studied the C++ course in Yarmouk University, Jordan in the year 2005. Three different classification methods namely ID3, C4.5, and the NaïveBayes were used. The outcome of their results indicated that Decision Tree model had better prediction than other models.

Bhardwaj and Pal [8] conducted study on the student performance based by selecting 300 students from 5 different degree college conducting BCA (Bachelor of Computer Application) course of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayesian classification method on 17 attributes, it was found that the factors like students grade in senior secondary exam, living location, medium of teaching, mother's qualification, students other habit, family annual income and student's family status were highly correlated with the student academic performance.

Umamaheswari and Niraimathi[9] has used the data mining to categorize the students into grade order in all their education studies and it helps in interview situation. This study explores the socio-demographic variables (age, gender, name, lower class grade, higher class grade, degree proficiency and extra knowledge or skill, etc). It examines to what extent these factors helps to categorize students in rank order to arrange for the recruitment process. Due to this, all students get benefitted and it also reduces the short listings. Here, clustering, association rules, classification and outlier detection has been used to evaluate the students performance.

III. DATA COLLECTION

Here only few attributes have been selected which are required for data mining. The information was selected from the student information system and stored in an excel file which then was converted into .csv file. The student related attributes have been shown in the table 1 given below with the necessary description and the domain values.

Table 1: Attribute description (with Domain values)

Variable	Description	Possible Values
ATT	Attendance	{Poor , Average, Good}
CTG	Class Test Grade	{Poor , Average, Good}
MSM	Mid Semester Examination Marks	{First > 60%, Second >45 & <60%, Third >36 & <45%, Fail<36%}
SEM	Seminar Performance	{Poor , Average, Good}
PEV	Practical Evaluation	{Poor , Average, Good}
PSM	Previous Semester Marks	{First > 60%, Second >45 & <60%, Third >36 & <45%, Fail<36%}
ASS	Lab work and Assignment	{Yes, No}

As shown in the table 2 we have collected 72 records of the students of CSE department of PIT college who are currently in 6th semester and we want to check their performance. In this out of 72 students there are 28 female and 44 male students available.

Table 2: Student performance table

Sr.No.	Attendance (ATT)	Class Test Grade (CTG)	Mid Semester Exam Marks (MSM)	Seminar Perf (SEM)	Practical Eval (PEV)	Previous Sem Marks (PSM)	Lab work and Assig Marks (ASS)	Gender
1	Average	Poor	Third	Poor	Poor	Fail	Poor	M
2	Good	Poor	Third	Poor	Poor	Fail	Poor	M
3	Good	Poor	Second	Poor	Poor	Fail	Poor	F
4	Good	Poor	Third	Poor	Poor	Fail	Poor	M
5	Average	Poor	Fail	Poor	Poor	Fail	Poor	M
6	Good	Good	First	Good	Good	First	Good	F
7	Good	Good	Second	Good	Poor	Second	Poor	F
8	Good	Good	First	Good	Good	First	Good	M
9	Average	Average	Second	Average	Average	Second	Average	F
10	Good	Good	Second	Good	Poor	First	Poor	F
11	Poor	Poor	Third	Poor	Poor	Fail	Poor	M
12	Good	Good	First	Good	Good	First	Good	M
13	Average	Average	Second	Average	Average	Second	Average	M
14	Good	Good	First	Good	Poor	First	Poor	F
15	Average	Good	First	Good	Good	First	Good	M
16	Good	Good	First	Good	Average	First	Average	F
17	Good	Good	First	Good	Good	First	Good	F
18	Average	Good	Second	Good	Good	Second	Good	M
19	Average	Average	Third	Average	Poor	Fail	Poor	F
20	Average	Average	Third	Average	Good	Second	Good	M
21	Good	Good	Second	Good	Average	First	Average	F
22	Average	Average	Second	Average	Good	Third	Good	M
23	Good	Good	Second	Good	Good	First	Good	M
24	Good	Good	First	Good	Average	Second	Average	F
25	Average	Average	Second	Average	Average	Second	Average	M
26	Poor	Poor	Third	Poor	Poor	Fail	Poor	M
27	Average	Average	Third	Average	Average	Third	Average	M
28	Average	Average	First	Average	Average	Second	Average	F

29	Good	Average	Second	Average	Average	Fail	Average	M
30	Average	Average	First	Average	Poor	First	Poor	F
31	Average	Average	Second	Average	Poor	Second	Poor	F
32	Good	Good	Second	Good	Average	Second	Average	F
33	Average	Good	First	Good	Good	First	Good	M
34	Good	Average	Second	Average	Poor	Second	Poor	F
35	Poor	Poor	Second	Poor	Poor	Third	Poor	M
36	Good	Good	First	Good	Good	First	Good	M
37	Average	Average	Second	Average	Good	Second	Good	M
38	Poor	Poor	Fail	Poor	Poor	Fail	Poor	M
39	Poor	Poor	Fail	Poor	Poor	Fail	Poor	M
40	Average	Average	Second	Average	Average	Third	Average	M
41	Good	Average	Second	Average	Poor	Second	Poor	M
42	Average	Average	Second	Average	Average	Fail	Average	M
43	Average	Average	Second	Average	Average	Fail	Average	M
44	Good	Average	Second	Average	Average	Second	Average	M
45	Good	Average	First	Average	Average	First	Average	M
46	Good	Average	Third	Average	Poor	Third	Poor	F
47	Good	Average	First	Average	Average	Second	Average	F
48	Average	Good	First	Good	Good	Second	Good	M
49	Good	Average	Second	Average	Average	Second	Average	M
50	Good	Good	Third	Good	Good	Second	Good	F
51	Average	Good	First	Good	Poor	First	Poor	F
52	Good	Average	Second	Average	Average	Second	Average	M
53	Good	Average	Second	Average	Average	Second	Average	M
54	Good	Good	First	Good	Average	First	Average	F
55	Good	Good	First	Good	Average	First	Average	F
56	Good	Average	Second	Average	Average	Second	Average	M
57	Good	Good	First	Good	Good	First	Good	M
58	Good	Good	First	Good	Good	First	Good	M
59	Good	Good	First	Good	Good	First	Good	F
60	Good	Good	First	Good	Good	First	Good	F
61	Good	Good	First	Good	Good	First	Good	M
62	Average	Average	Second	Average	Good	Fail	Good	M
63	Poor	Good	First	Good	Good	Second	Good	M
64	Average	Average	Second	Average	Average	Fail	Average	M
65	Good	Good	First	Good	Good	First	Good	M
66	Average	Good	First	Good	Good	First	Good	M
67	Good	Good	First	Good	Good	First	Good	M
68	Good	Good	First	Good	Average	First	Average	F
69	Good	Good	First	Good	Average	Fail	Average	F
70	Average	Good	First	Good	Average	First	Average	F
71	Average	Good	First	Good	Average	First	Average	F
72	Good	Good	Second	Good	Average	Fail	Average	M

IV. DATA PREPROCESSING

To get better input data for data mining techniques, we did some preprocessing for the collected data. After we integrated the data into one file, to increase interpretation and comprehensibility, we discretized the numerical attributes to categorical ones. For example, we grouped all result marks into four groups first, second, third and *fail*, attendance as good, average and poor. In the same way, we discretized other attributes such as practical and class test marks and so on. The figure 1 below has shown the different attributes with the visualization.

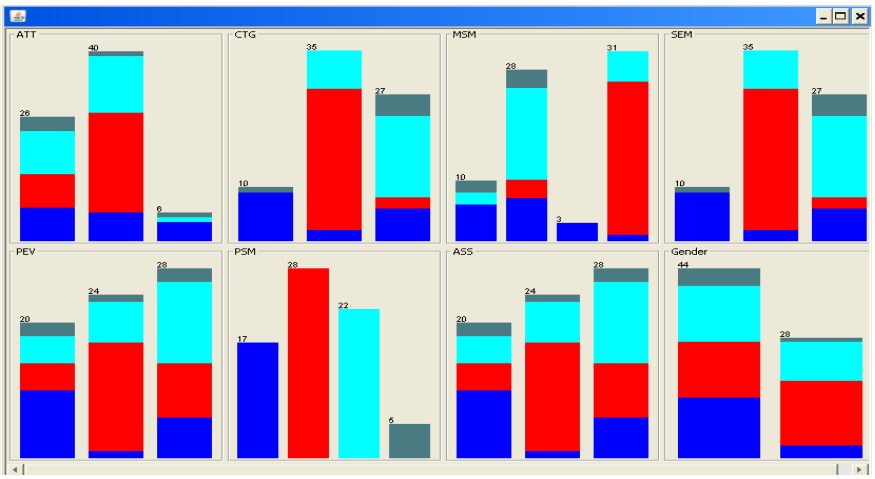

Figure 1: Data Visualization using Weka tool

V. DATA MINING ON EDUCATIONAL DATA

Data mining used advanced techniques to discover patterns from data. The data mining tasks are the kinds of patterns that can be mined. There are many tasks in data mining, the most common ones are: Association, classification, clustering and outlier detections. In the following sections describes the results of applying data mining techniques to the data of our case study for each of the four tasks.

In our case study we have used Apriori algorithm for finding the association among the different attributes of the student performance. As shown in the figure 2 given below there are total 10 best rules have been found using weka, for this the minimum support was 0.25 (i.e. 25 instances) and minimum confidence was 0.9. There are total 3 rules which has large itemset of size 3.

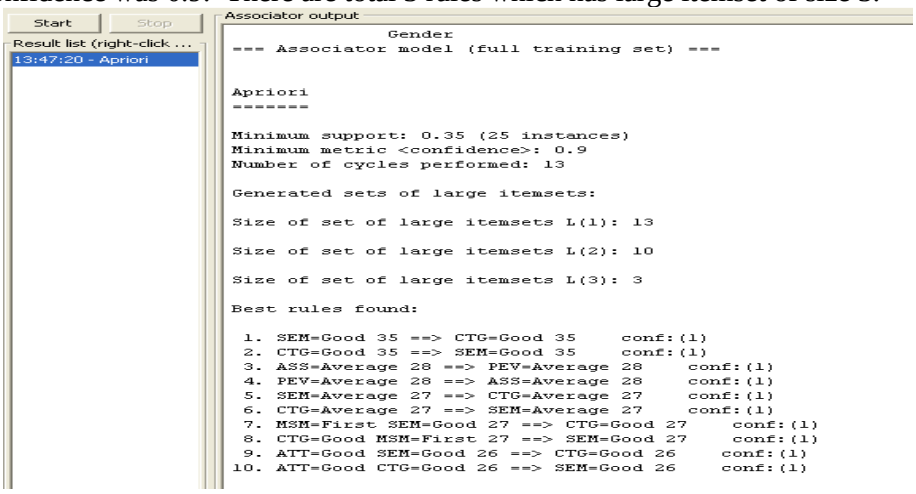

Figure 2: Associations rules for student data

VI. CONCLUSION AND FUTURE WORK

The current paper has used the association rule mining task on student database to predict the students division which is based on the previous performance. In this paper the simple Apriori algorithm has

been used in WEKA for association rule generation, this may helpful to predict the future performance of the student and based on the necessary care can be taken for the students having poor performance.

In future we will work on large data set of educational data and will also implement the same in distributed environment, which may do the processing much faster than the existing algorithm. Many other data mining techniques will be applied to improve the quality of the education.

REFERENCES

[1]. Heikki, Mannila, Data mining: machine learning, statistics, and databases, IEEE, 1996.
[2]. U. Fayadd, Piatesky, G. Shapiro, and P. Smyth, From data mining to knowledge discovery in databases, AAAI Press / The MIT Press, Massachusetts Institute Of Technology. ISBN 0–262 56097–6, 1996.
[3]. J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2000.
[4]. Alaa el-Halees, "Mining students data to analyze e-Learning behavior: A Case Study", 2009.
[5]. U. K. Pandey, and S. Pal, "Data Mining: A prediction of performer or underperformer using classification", (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), pp.686-690, ISSN: 0975-9646, 2011.
[6]. Smitha .T, V. Sundaram, "Comparative Study of Data Mining Algorithms for High Dimensional Data Analysis", International Journal of Advances in Engineering & Technology, Vol. 4, Issue 2, pp. 173-178, Sept 2012.
[7]. Galit.et.al, "Examining online learning processes based on log files analysis: a case study". Research, Reflection and Innovations in Integrating ICT in Education 2007.
[8]. Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar, "Mining student data using decision trees", International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 2006.
[9]. E.V.Ramana and P. Ravinder Reddy, "Data Mining Based Knowledge Discovery for Quality Prediction and Control of Extrusion Blow Molding Process", International Journal of Advances in Engineering & Technology, Vol. 6, Issue 2, pp. 703-713, May 2013.
[10]. B.K. Bharadwaj and S. Pal. "Data Mining: A prediction for performance improvement using classification", International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2011.
[11]. K.R.Lakshmi, Y.Nagesh and M.VeeraKrishna, "Performance Comparison of Three Data Mining Techniques for Predicting Kidney Dialysis Survivability", International Journal of Advances in Engineering & Technology (IJAET), Volume 7 Issue 1, pp. 242-254, Mar. 2014.
[12]. Umamaheshwari k and S. Niraimathi, "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, ISSN: 2277 128X

AUTHORS BIOGRAPHY

Dineshkumar Bhagwandas Vaghela PhD Scholar, Gujarat Technological University, Chandkheda. National journals and also presented the papers in many national Conference. His area of research is Distributed data mining and machine learning scholars.

Priyanka Sharma PhD Guide, GTU, Chandkheda. She has published many research papers in national and International Journals and she is currently guiding many research.