

# Evaluating the Performance of the Students in the Technical Education through Data Mining Technique

Shiv Kumar Gupta<sup>1</sup>, Dr. Ritu Vijay<sup>2</sup>

<sup>1</sup>*School of Engineering and Technology, Jaipur National University, Jaipur (India)*

<sup>2</sup>*Department of Computer Science and Electronics, Bansthali University, Rajasthan (India)*

skgupta50@yahoo.com

**Abstract:**-In order to understand how and why data mining works, it's important to understand a few fundamental concepts. First, data mining relies on four essential methods: Classification, categorization, estimation, and visualization. Classification identifies associations and clusters, and separates subject under study. Categorization uses rule induction algorithms to handle categorical outcomes, such as "persist" or "dropout," and "transfer" or "stay". Visualization uses interactive graphs to demonstrate mathematically induced rules and scores, and is far more sophisticated than pie or bar charts. Visualization is used primarily to depict three-dimensional geographic locations of mathematical coordinates. Higher education institutions can use classification, for a comprehensive analysis of student characteristics, or use estimation to predict the likelihood of a variety of outcomes, such as transferability, persistence, retention, and course success. In this paper we have evaluated about the performance of the students through their scores obtained in different years as well as their dropout percentage of each year.

**Keywords:** Data Mining, Technical Education, Student Retention, Classification, Visualization.

## 1. INTRODUCTION:

Today there is an urgent need for well trained Engineers and Scientists. Very less Institution pay proper attention to maintain the minimum requirement of each resource related to each program. This paper addresses the capabilities of data mining and its applications in the technical education. This demonstrates how data mining saves resource while maximizing efficiency and increases productivity without increasing cost. Data mining uses a combination of an explicit knowledge based sophisticated analytical skill and domain knowledge to uncover hidden change and patterns. These trained patterns form the bases of predictive models and enables an analyst to produce new observation from the existing data. It is powerful tool for academic intervention through data mining. A university could predict dropout percentage of the students as well as their performance in discipline wise. Data mining technique could help us to provide personalized education, maximize educational system efficiency and reduces the cost of education processes. This will further help us to increase student's retention rate increases educational improvement ratio, increase student's learning result. In one work to predict learning behavior of the student's data mining techniques have been applied. In another work minimum numbers of features have been used to train and classify the students to evaluate their performance. One more work present an approach to classify the students in order to predict their final grade based on different features extracted from logged data in an education

system.

It has designed implemented and evaluated a, series of pattern classifiers and compare their performance on an online course dataset. A combination of multiple classifier leads to a significant improvement in classification performance. [4] This paper has discussed about two ways that researchers have attempted to categorize the diversity of research in educational data mining, and review the types of research problems that these methods have been used to address.[2] In this paper we are using data mining simulator to evaluate the performance of the students. The rest of the paper is structured as follows: Data mining Techniques, in Section 3 proposed model is shown in section 4, section 5 contains conclusions while the references are shown in last section.

## 2. DATA MINING TECHNIQUE:

Various data mining techniques have been used for different applications.

### Decision Tree:

Decision tree is one of the data mining technologies which have been used in this work. This technique is originally developed for statisticians to automate the process of determining which field in their database is actually useful or correlated with the particular problem that they are trying to understand. This is also particularly adept at handling raw data with little or no preprocessing. This technology can be used for exploration of the data set and future prediction. The general idea behind decision tree technology is that this decision tree can be built from the historical data. This is a form of supervised learning, although this is often used for exploratory analysis as well. [7]

### Data Clustering:

A particular type of data mining problem can be examined by data Decision tree. Given a large set of multidimensional data points the data space is not usually uniformly occupied. Data clustering identifies the sparse and the crowd places and discover the overall distribution pattern of dataset and cluster can be visualized more effectively. A good quality clustering will provide high intra-clustering and low inter clustering. A good quality clustering can help to find hidden patterns without scanning data. Balanced Interactive Reducing and Clustering using Iteractions performs faster than existing algorithms on large datasets in quality, speed, stability and scalability. Its I/O cost is linear in the

size of dataset: a single scan of dataset yields a good clustering; including one or more additional passes can improve quality further. [8]

**Association Rules:**

Association rules of data mining are employed today in many application areas including web usage mining, intrusion etc. This methodology known as association analysis which is useful for discovering interesting hidden relationships in large datasets. A record in such a data typically consists of the transaction.

Association rules are usually required to satisfy a user specified minimum support and a user specified minimum confidence at the same time. Association rule generation is usually split up into two separate steps.

1. First minimum support is applied to find all frequent item sets in a database.
2. Second, these frequent item sets and the minimum confidence constraint are used to form rules.

While the second step is straightforward and the first step needs more attention. [9]

**3. PROPOSED MODEL:**

The technical education system gets affected by the different reasons. One is from the Government governing organization like AICTE (All India Council Of Technical Education) in India, second is private Universities have

different internal policies and thirdly the outside industrial environment which require trained manpower. The outside environment plays a great role to uplift the status of the student. This motivates the educational organization to stay ahead of their business competitors. Our proposed model is developed to serve as a prerequisite to today's need. Technical education systems have to look for a new and faster solution to overcome the educational problems in order to achieve a high standard academic level. This model can be used by the technical institution to identify that which part of their educational processes can be improved by data mining technique and how they can achieve this goal by implementing data mining system. We have considered two main processes in technical education system which is the admission of the students and their evaluation.

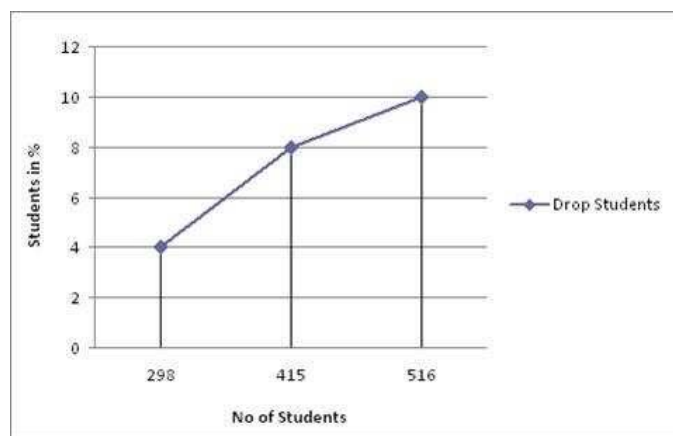
**4. EXPERIMENTS AND RESULTS:**

We have collected three years data from one of the reputed technical institution. This includes data of 1229 students with 5 different disciplines. Data contains the number of students admitted in each year discipline wise as well as the students who drops from the course. The complete details of students are shown through the following table.

**Experiment 1: Admitted and Dropout percentage of students.**

Discipline	Year 2009 – 10		Year 2010 – 10		Year 2011 – 12	
	Students admitted	Students dropped	Students admitted	Students dropped	Students admitted	Students dropped
Civil Engg.					117	7
ComputerEngg.	77	3	74	9	70	10
ElectricalEngg.	74	3	64	3	67	61
Electronics & Commucation Engg.	84	2	56	9	60	50
Information Technology	63	3	62	2	96	86
Mechanical Engg.			147	11	163	20
Total	298	11	415	34	573	516
% of drop students	4%		8%		10%	

**Table 1 indicates the students enrolled and left the courses.**



**Fig 1: Students and their dropouts percentage.**

The table 1 clearly specified the dropout percentage of each year. It seems clear that the dropout percentage is increasing in the subsequent year.

In order to increase students retention, we should understand why students dropout. A survey is conducted among why student’s dropout students of each year by picking the attributes gender, loan from the bank, score of previous semester, regularity in attending the classes, academic involvement. The decision tree algorithms have been used to show that the dropout becomes high due to the one or more reasons specified below:

1. Due to being failed in two or more than two papers.
2. Due to the dissatisfaction of the institution either in terms of resources or in terms of level of faculties.
3. Due to financial position of the parents.

An experiment for student dropout is also conducted based

on the student profile. Two set of data have been collected for this purpose one was for training and another one was for testing.

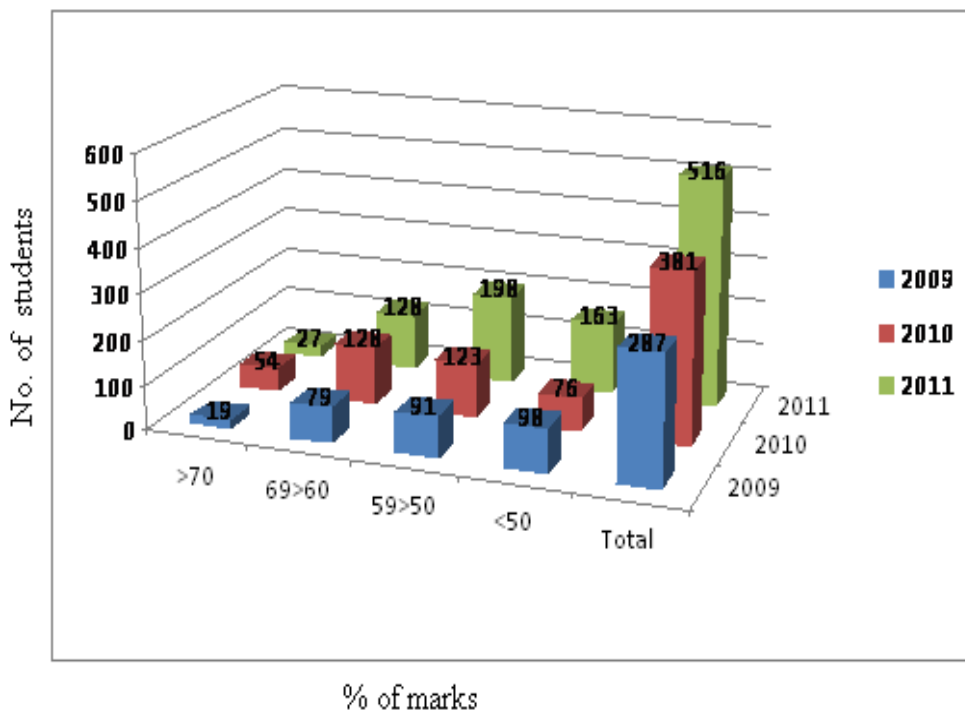
In the next experiment performance of the students for three years are evaluated. The following table shows the detailed report of the students.

From the table 2 /figure2, it looks clear that in all three years the number of students who scores between the range 50% to 59% in grade 3 are higher. In grade 2, 3 & 4, we have these students who are in majority and score less than 70%, In Grade 1 excellent students are few in the range 5-14% and this analysis continues in subsequent years. Therefore some corrective measures are required to incorporate.

**Experiment 2: Performance of the students for the years 2009 -12**

Batch	Exam. Held	No. of Student Appeared	% of Marks Secured			
			>70% Grade 1	60 % to 69% Grade 2	50% to 59% Grade 3	<50% Grade 4
2009 - 10	Dec 2009	287 (100%)	19 (7%)	79 (28%)	91 (32%)	98 (34%)
2010- 11	Dec 2010	381 (100%)	54 (14%)	128 (34%)	123 (32%)	76 (20%)
2011-12	Dec 2011	516 (100%)	27 (5%)	128 (25%)	198 (38%)	163 (32%)

**Table 2: Indicate % of marks secured verses % of the students**



**Fig 2: Performance of the students for the years 2009 -12**

#### 4. CONCLUSION:

In this paper we have evaluated the performance of the students by considering their percentage of marks /scored and also tried to find the reasons of their dropouts. Our survey in the current works in data mining shows that this work could be helpful to improve the quality of technical education.

#### REFERENCES:

- [1]. Jing Luan, “*Data Mining Applications in Higher Education*”, 2004 SPSS.
- [2]. Ryan S.J.D. Baker and Kalian Yacef, “*The State of education data mining in 2009: A Review and Future Visions*”, 2009.
- [3]. Alaa El-Halees, “*Mining student’s data to analyze learning behavior: A case study*”.
- [4]. Behrouz Minaei-Bidgoli, Deborah A. Kashy, Gerd Kortemeyer and William F. Punch, “*Predicting student performance: An application of data mining methods with the educational web-based system Lon-Capa*”, 2003 IEEE.
- [5]. Brijesh Kumar Baradwaj and Saurabh Pal, “*Mining educational data to analyze students’ performance*”.
- [6]. Ying Zhang and Tony Clark, “*Use data mining to improve student’s retention in higher education – A Case study*”.
- [7]. Alex Benson, Stephen J. Smith, “*Data Warehousing, Data Mining, & OLAP*”, The McGraw Hill 21<sup>st</sup> reprint 2011 ISBN: 978-0-07-058741-0.
- [8]. Zhang, Ramakrishnan, Citations, “*An efficient data clustering method for very large databases*”, 149 et.al.-1996.
- [9]. [http://en.wikipedia.org/wiki/Association\\_rule\\_learning](http://en.wikipedia.org/wiki/Association_rule_learning) - cite\_ref-apriori\_5-1Agrawal, Rakesh; and Srikant, Ramakrishnan; “*Fast algorithms for mining association rules in large databases*”, in Bocca, Jorge B.; Jarke, Matthias; and Zaniolo, Carlo; editors, Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), Santiago, Chile.
- [10]. Beikzadeh M., and Delavari, N., “*A new Analysis Model for Data Mining Processes in Higher Educational Systems*”, On proceedings of the 6<sup>th</sup> Information Technology Based Higher Education and Training 7-9 July, 2005.