

# A Crucial Engineering Students Academic Performance Evaluation Data Mining Techniques

Stephenraj. M<sup>1</sup> Rajanand. S<sup>2</sup> Dhayalan. D<sup>3</sup> Sivakumar. S<sup>4</sup>  
<sup>1,2,3,4</sup>Department of MCA

<sup>1,2,3,4</sup>Vel Tech High Tech Dr Rangarajan Dr Sakunthala Engineering College, Chennai,600062, India

**Abstract**— Data mining and knowledge engineering is to predict the analysis of data for the purpose of verdict certain informational patterns that can be used for the benefits of an educational institution. Data mining methods are often implemented at higher education engineering colleges / universities today for analyzing available data and extracting information and knowledge to support decision-making. Data mining is used to improve graduate students' performance and overcome the problem of failures, due to the problematic subject in low grades marks and failures for engineering graduate students. The main scope of this paper are to identify the data analysis for engineering students evaluation performance based on skills in academic subject and process, the information gathered through this research can be used by the educational institution to improve its services and became top engineering college and universities in Tamil Nadu. Here, the classification, association, clustering are used to evaluate student's performance. By using these methods to extract the knowledge description database that describes students' performance in the end of the university examination. It also helps earlier in identifying the dropouts engineering course and students who need special attention and allow the professor to provide appropriate advising and counseling.

**Key words:** Data Mining, categorization, Association, Clustering and Educational Data Mining

## I. INTRODUCTION

Data mining is a type of sorting technique which is actually used to extract hidden patterns from large databases[2]. Data mining concepts and methods can be applied in various fields like, Infrastructure, enterprise resource planning, engineering, information technology, advertising, healthcare, educational data mining, etc. Educational data mining is a new emerging technique of data mining that can be applied on the data related to the field of education[4]. It uses many techniques such as decision trees, classification, relationship, clustering and many others. Using these techniques different aspects of knowledge can be discovered using association rules, categorization, decision tree and clusters. By using this we extract knowledge that describes students' performance in the end of the semester examination and all their details. In the face of huge amounts of information, the first task is to sort them out, cluster analysis is to categorize the raw data in an efficient way. The so called clustering is a group of physical or abstract objects, according to the same degree between them, divided into more groups and makes the same data objects within a groups of high similarity and different groups of data objects which are not similar.

## II. RELATED WORK

Educational data mining in higher education is a latest research field and this area of research is gaining popularity

because of its potentials to educational institutes[11]. In this paper that use students data to analyze their learning behavior to predict the results. Mohammed M.AbuTair, Alaa M.EIHalees had a survey on educational data mining [1993-2007] they collect engineering graduate students information and applied data mining techniques to discover knowledge. Using discovered association rules, they sorted the rules using raise metric. Then they used two way classification methods which are rule induction and naive Bayesian classifier to predict the grade of engineering graduate students [3]. They also clustered the students into groups using k -mean clustering algorithm. Finally, they used an outlier detection to detect all outliers in the data. Two outlier methods which are distance based approach and density -based approach were used. Each one of these tasks goes hand in hand to improve the performance of engineering graduate students. We have survey on educational data mining between 2005 and 2010.They concluded that educational data mining is a promising area of research and it has specific requirements not presented in other domains. Thus, the work should be oriented towards educational domain of data mining. We have conducted study on the students performance based on selecting 500 students from 5 different degree college conducting B.E and B. Tech courses of it was found that the factors like students grade in higher secondary exam, address, medium of instruction, parents qualification, students characteristics and other habit, family annual income and parent occupation were highly correlated with the student academic performance.

## III. PREPROCESSING TECHNIQUES AND DATA MINING METHODOLOGY

Romero&Ventura (2007) according to Baker & Yacef (2009) identifies the following types of educational data mining [5]:

Statistics and visualization

### A. Web Mining:

- Clustering, classification, and outlier detection
- Association rule mining and sequential pattern mining

Baker & Yacef (2009) then summarize a new typology defined in Baker (2010):

- 1) Prediction
  - Classification
  - Regression
  - Density estimation
- 2) Clustering
- 3) Relationship mining
  - Association rule mining
  - Correlation mining
  - Sequential pattern mining

Calders & Pechenizkiy (2011) associate basic EDM tasks to traditional data mining problems;

Data Mining Problems	Educational Example	Author
Classification	Categorizing and profiling students, determine their learning styles and preferences	Cha Et Al, 2006
Predictive Modeling	Inducing models that can predict whether (and when) A Student will pass a course or not or will eventually graduate or drop out	Hämäläinen & Vinni, 2006; Dekker Et Al, 2009
Clustering	Grouping similar students (based on behavior, performance, etc) or grouping similar courses, assignments, etc together, exploring collaborative learning patterns	Perera Et Al, 2009
Frequent Pattern Mining	Finding (elective) courses often taken together or popular paths in study programs or actions.	Zaiāne, 2001
Collaborative Filtering and Recommendations	Recommending suitable learning objects, based on the analysis of the performance of other learners, recommend remedial classes to	Perera Et Al., 2009.

Table 1: Traditional Data Mining Problems.

The dataset is a collection data item of final year engineering graduate students information. To group the student’s data using clustering technique, it may hold the academic, non academic and personal record of the student. It contains students whole study details from its beginning. When the data is being taken from the educational field then this new emerging field is termed as “Educational Data Mining[1]”. The major aim of this higher educational institutions is to provide proper educational facilities to the student[13]. For this reason they categorize the student based on their skill level. Skill level is ranked in the form of CGPA grade taking into account end semester marks and also based technical skill test .These are used to evaluate the results. According to students level, they categorized them into groups.

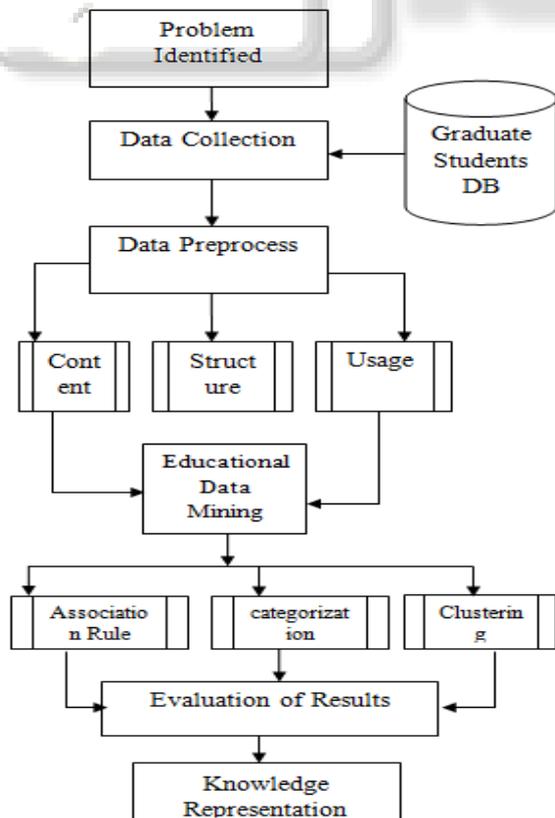


Fig. 1: Data Mining Methodology

**B. Preprocessing:**

Data mining algorithms can be used, and target data set must be assembled. As data mining can only be uncover patterns that are actually present in the data, the target data set must be huge enough to contain these patterns while remaining concise enough to be mined within an acceptable time limit[3]. A common source for data is a data mart or data warehouse. Pre-processing is an essential to analyze the multivariate data sets before data mining. The target set is then cleaned, then the Data cleaning removes the observations that contains noise and those with missing data. There are three phases of preprocessing techniques; first one is called as content pre-processing is the process of converting text, image, scripts and other documents into the forms that can be used by the usage mining. Second one as called a structure preprocessing is the structure of a Website is formed by the hyperlinks between page views; the structure preprocessing can be done by parsing and reformatting the information. Finally usage preprocessing is the data cleaning techniques to eliminate the impact of the irrelevant items to the analysis result.

**C. Association Rule:**

Association rule learning is a popular and well researched method for discovering interesting relations between variables in huge databases. An 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.

First, minimum support is applied to identify all the recurrent item sets in a database. Secondly, these frequent item sets and the minimum confidence constraint are used to form rules. By finding all frequent item sets in a database is difficult since it involves searching all possible item sets (item combinations). The set of likely item sets is the power set over I and has size  $2^n - 1$  (excluding the empty set which is not a valid item set). It depicts a sample of association rules discovered from data students with average grade, the support and confidence.

**D. Procedure:**

- If students CGPA is between 9 -10 THEN performance = “Excellent”
- If students CGPA is between 8-9 THEN performance = “Very Good”
- If students CGPA is between 7-8 THEN performance = “Good”
- If students CGPA is between 6-7 THEN performance = “Average”
- if students CGPA is below 5 THEN performance = “Poor”

**E. Classification:**

Classification is the process of finding a model that describes and distinguishes data concepts, for the purpose of being able to use the model to predict the class of objects whose class label is indefinite. The derived model is based on the analysis of a set of instruction data. It is important to know that classification rules are different than rules generated from association. Association rules are featured rules, but categorization rules are prediction rules.

F. Clustering:

Clustering is a method in which we make cluster of objects that are somehow similar in characteristics. The criteria for checking the similarity is implementation dependent [5]. Clustering is often confused with categorization, but there are differences. In classification the objects are assigned to pre-defined classes, where as in clustering the classes are also to be defined. This clustering methods may be divided into two categories based on the cluster structure which they produce hierarchical cluster and partitioning cluster.

IV. SIMULATION RESULTS

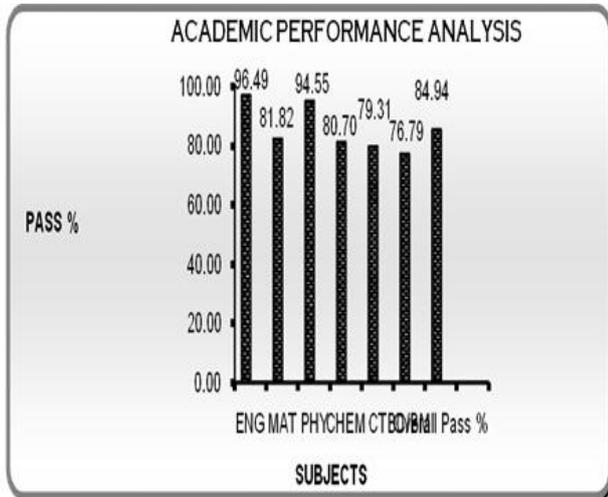


Fig. 2: Academic Performance Analysis

The Figure.2 shows that Academic performance are calculated based on the percentage of mark is respective subjects and minimum pass is 50%, and is overall percentage of respective subjects such As English 93.49, maths81.8,physics94.55,chemistry80.70,computerseience79.38,biology76.79,other languages-84.94.

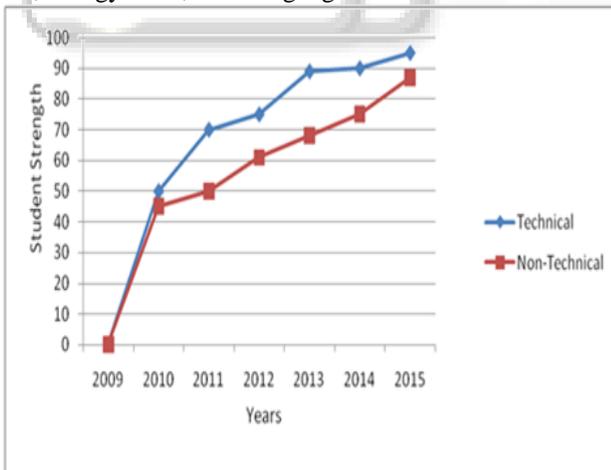


Fig. 3: Yearwise Report Of Technical and Non-Technical Student Strength

The Figure.3 shows that Student strength yearwise report, the student strength can be splitted into, two ways technical and non-technical from 2009-2015 years and is each year the students strength have been increased rapidly,and among themwhile comparing technical student strength have been increased in 2015.

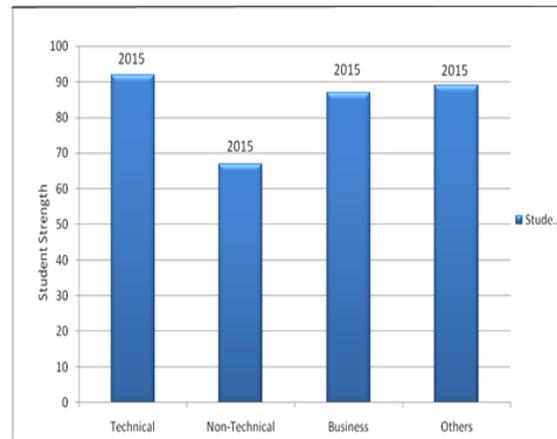


Fig. 4: Student Strength Ratio Yearly Activities Of Student In Engineering College.

The Figure.4 Shows that student strength ratio. The Student strength ratio among different divisions is Technical ,Non-technical,Business,others. The technical ratio in student strength is high in 2015.

V. CONCLUSION

In this paper data mining techniques are efficiently used to categorize the level of engineering students. One of the data mining techniques that is classification accurately classifies the data for categorizing engineering student based on the levels. As one important function of data mining, clustering analysis is either as a separate tool to discover data sources distribution of information, as well as other data mining algorithm as a pre-processing step, the cluster analysis has been into the field of data mining is an important research topic. Clustering is used to the group the engineering students according to their academic grade and proficiency. This goes a long way to help how define the academic result process in an easier manner. This work used the classification, association, clustering are used to evaluate student's performance. By using these methods we extract knowledge description database that describes students' performance in the end of the university examination. The prediction of the paper proposed any student's without dropout engineering course in future and students who needed special attention and allow the professor to provide appropriate advising and counseling to the students and also students performance to be intimate to the parents.

REFERENCES

- [1] Er.Rimmy Chuchra,'Use of Data Mining Techniques for the Evaluation of Student Performance: A case study", International Journal Of Computer Science and Management Research,vol 1, 140-146,2012.
- [2] Han.J and Kamber.M,"Data Mining: Concepts and Techniques,2<sup>nd</sup> edition. The Morgan Kaufmann series in Data Management Systems, Jim Gray, series editor, 129-134,2006.
- [3] El-Hales-A,"Mining Student s Data to Analyze Learning Behavior: A Case Study", The 2008 International Arab Conference of Information Technology(ACIT2008) -Conference Proceedings, University of Sfax,Tunisia,Dec 15-18,2008.

- [4] B.k.Bharadwaj and s-pal, "Data Mining: A Prediction for Performance Improvement using Classification", (IJCSIT) International Journal of Computer Science and Information Security (IJCSIS) vol.9,no.4,pp-136-140,2011.
- [5] M.I. López and J.M Luna, "Classification via clustering for predicting final marks based on student participation in forums, 1-4,2009.
- [6] R.Rabbany, M. Takaffoli and O. Zaïane, "Analyzing participation of students in online courses using social network analysis techniques", Proceedings of educational Data Mining, 21-30, 2011.
- [7] S. K. Yadav, B.K. Bharadwaj and S. Pal, "Data Mining Applications: A comparative study for predicting students performance", International Journal of Innovative Technology and Creative Engineering, vol1, No. 12, ISSN:2045-8711, 51-56, 2011.
- [8] Q. A. Al-Radaideh, E. W. Al-Shawakfa, and M. I. Al-Najjar, "Mining student data using decision trees", International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 31-37,2006.
- [9] Minaei-Bidgoli B.; Kashy D.; Kortemeyer G.; . "Predicting student performance: an application of data mining methods with an educational web based system". In Proc. of IEEE Frontiers in Education. Colorado, USA,13-18,2003.
- [10] Available[online]  
[http://en.wikipedia.org/wiki/Data\\_mining](http://en.wikipedia.org/wiki/Data_mining).
- [11] Mohammed M.Abutair, AlaaM.El-Halees, "Mining Educational data to improve students" performance: A case study", International Journal of Information and Communication Technology Research, Volume 2 no2,2012.
- [12] Daniel B.A., Ping Chen Using Self-Similarity to Cluster Large Data Sets[J].Data Mining and Knowledge Discovery,7(2):123-152,2003.
- [13] Calders, Toon & Mykola Pechenizkiy Introduction to The Special Section on Educational Data Mining, SIGKDD Explorations 13 (2),2011.
- [14] Yujie.zhang. "Clustering Methods in Data Mining with its Applications in High Education", International conference on Education Technology and Computer (ICETC 2012) IPCSIT vol.43,2012.