

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to the **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T.Sasikala M.E., Ph.D., Dean**, School of Computing , **Dr.S.Vigneshwari M.E., Ph.D.**, and **Dr.L.Lakshmanan M.E., Ph.D.**, Heads of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Mrs. D. Deepa** for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

In many organizations, data mining techniques are used for analyzing large amount of available data's, information's for decision making process. In educational sector, Data mining is used for wide variety of application's such as performance of the students like mark, attendance, staff opinion, extracurricular activities, Ragging and stress. The data mining techniques used for identifying the performance of the student using Naïve Bayes, K-means, KNN and Lasso algorithms. These two algorithm's identifies and analyses the performance of the student.

Table Of Content

No.	Title	Page No.
	Abstract	5
	Table Of Content	6
1.	Introduction	7-8
1.1	Objective	8
2	LITERATURE REVIEW	9-15
2.1	Existing Sytem	14
2.2	Proposed System	15
3	AIM AND SCOPE OF THE PRESENT INVESTIGATION	16
4	Experimental or materials and methods algorithms used	17-48
5	RESULTS AND DISCUSSION	49-51
5.1	Screenshots	49
6	CONCLUSION	50
7	References	51-52
8	Source Code	53

1. INTRODUCTION

There are many tremendous improvement research interests in using data mining in educational sector. This modern emerging sector, called educational data mining, concerned with improved methods that extract knowledge from data come from the educational sector. Data mining is a technique of sorting which is actually used to extract hidden patterns from huge databases. This concepts and methods can be applied in various fields like marketing, medicine, real estate, customer relationship management, engineering, web mining, etc. Educational data mining is a new emerging or advanced technique of data mining that can be applied on the data related to the field of education. The data can be collected from past used data and operational data reside in the databases of educational institutes. The data of students can be personal information or academic performance. Furthermore it can be achieve from e-learning database systems which have a huge amount of data and information used by most institutes . It uses many techniques for proper implementation of data mining concepts such as K-means Clustering, K-Nearest neighbour and Lasso algorithm. Using these techniques different kinds of knowledge can be discovered using association rules, classification and clustering. By using this we extract knowledge that describes students performance in examination and all their detail information. From This huge amounts of data, the first task is to sort them out, cluster analysis is to classify the raw data in a reasonable way. Clustering is a bunch of physical or abstract objects, as per the degree of similarity between them, divided into several groups, and makes the same data objects within a groups of high similarity and different groups of data objects which are not similar.

1.1 OBJECTIVE

The main aim of this project is to improvise the student performance in studies based on some important factors. Education is an essential element for the betterment and progress of a country. It enables the people of a country civilized and well mannered. Now-a-days developing new methods to discover knowledge from educational database in order to analyse student's trends and behaviours towards education. To analyse the data from different dimensions categorize it and to summarize the relationships. It motivated us to work on student dataset analysis. The data collection, categorization and classification is being performed manually.

2. LITERATURE REVIEW

Performance prediction of students using distributed data mining

The performance of students in higher education in India is a turning point in the academics for all students for their brightest career. In today's generation the amount of data stored in educational database increasing at a great rate. These databases contain secret information for improvement of students' performance; these data can be located at different nodes in distributed system. Classification and prediction are among the major techniques in Data mining and widely used in various fields. In this paper classification techniques are used for prediction of student performance in distributed environment. Data mining methods are often implemented at many advance universities today for analyzing available data and extracting information and knowledge to support decisionmaking. While it is important to have models at local level, their results makes it difficult to extract knowledge that can be useful at the global level. Therefore, to support decision making at this area, it is important to generalize the information contained in those models, specific classifier method can be used to generalize these rules for global model.

Predicting Student Academic Performance

Engineering schools worldwide have a relatively high attrition rate. Typically, about 35% of the first-year students in various engineering programs do not make it to the second year. Of the remaining students, quite often they drop out or fail in their second or third year of studies. The purpose of this investigation is to identify the factors that serve as good indicators of whether a student will drop out or fail the program. In order to establish early warning indicators, principal component analysis is used to analyze, in the first instance, first-year engineering student academic records. These performance predictors, if identified, can then be used effectively to formulate corrective action plans to improve the attrition rate.

Data Mining Approach For Predicting Student Performance

This work proposes a novel approach - personalized forecasting - to take into account the sequential effect in predicting student performance (PSP). Instead of using all historical data as other methods in PSP, the proposed methods only use the information of the individual students for forecasting his/her own performance. Moreover, these methods also encode the "student effect" (e.g. how good/clever a student is, in performing the tasks) and "task effect" (e.g. how difficult/easy the task is) into the models. Experimental results show that the proposed methods perform nicely and much faster than the other state-of-the-art methods in PSP.

A novel approach for upgrading Indian education by using data mining techniques

Education is the backbone of all developing countries. Upgrading of the education system, upgrades the country to the world top ranking level. One of the major problems that the education system facing is predicting the behaviour of students from large database. This paper focus on upgrading Indian education system by using one of the techniques in Data mining namely clustering. Cluster analysis solves the given data into some meaningful groups. Normally the performances of the students can be classified into different patterns as normal, average and below average. In this paper we attempt to analyze student's data in different angle beyond the above indicated patterns through newly proposed UCAM (Unique clustering with Affinity Measures) clustering algorithm.

A Review on Data Mining techniques and factors used in Educational Data Mining to predict student amelioration

Educational Data Mining (EDM) is an interdisciplinary ingenuous research area that handles the development of methods to explore data arising in a scholastic fields. Computational approaches used by EDM is to examine scholastic data in order to study educational questions. As a result, it provides intrinsic knowledge of teaching and learning process for effective education planning. This paper conducts a comprehensive study on the recent and relevant studies put through in this field to date. The study focuses on methods of analysing educational data to develop models for improving academic performances and improving

institutional effectiveness. This paper accumulates and relegates literature, identifies consequential work and mediates it to computing educators and professional bodies. We identify research that gives well-fortified advise to amend edifying and invigorate the more impuissant segment students in the institution. The results of these studies give insight into techniques for ameliorating pedagogical process, presaging student performance, compare the precision of data mining algorithms, and demonstrate the maturity of open source implements.

Data Mining: A prediction of performer or underperformer using classification

Now a day's students have a large set of data having precious information hidden. Data mining technique can help to find this hidden information. In this paper, data mining techniques name Byes classification method is used on these data to help an institution. Institutions can find those students who are consistently perform well. This study will help to institution reduce the drop put ratio to a significant level and improve the performance level of the institution.

Mining students data to analyze e-Learning behavior: A Case Study

Educational data mining concerns with developing methods for discovering knowledge from data that come from educational environment. In this paper we used educational data mining to analyze learning behavior. In our case study, we collected students' data from DataBase course. After preprocessing the data, we applied data mining techniques to discover association, classification, clustering and outlier detection rules. In each of these four tasks, we extracted knowledge that describes students' behavior.

Scholastic achievement of higher secondary students in science stream

The present study was conducted on 400 students (200 boys and 200 Girls) selected from senior secondary school of A.M.U., Aligarh-India, to establish the prognostic value of different measures of cognition, personality and demographic variables for success at higher secondary level in science stream. The scores obtained on different variables were factor-analyzed to get a smaller number of meaningful variables or factors to establish the predictive validity of these predictors. Factors responsible for success in science stream were identified. The prognostic value of the predictors was compared for high achievers and low achievers in order to identify the factors which differentiate them.

Mining student data using decision trees

Student performance in university courses is of great concern to the higher education managements where several factors may affect the performance. This paper is an attempt to use the data mining processes, particularly classification, to help in enhancing the quality of the higher educational system by evaluating student data to study the main attributes that may affect the student performance in courses. For this purpose, the CRISP framework for data mining is used for mining student related academic data. The classification rule generation process is based on the decision tree as a classification method where the generated rules are studied and evaluated. A system that facilitates the use of the generated rules is built which allows students to predict the final grade in a course under study.

An Empirical Study of the Applications of Data Mining Techniques in Higher Education

Few years ago, the information flow in education field was relatively simple and the application of technology was limited. However, as we progress into a more integrated world where technology has become an integral part of the business processes, the process of transfer of information has become more complicated. Today, one of the biggest challenges that educational institutions face is the explosive growth of educational data and to use this data to improve the quality of managerial decisions. Data mining techniques are analytical tools that can be used to extract meaningful knowledge from large data sets. This paper addresses the applications of data mining in educational institution to extract useful information from the huge data sets and providing analytical tool to view and use this information for decision making processes by taking real life examples.