

Student's Performance Measuring using Assistant Algorithm

N. Venkatesan, N. Chandru

Abstract— The main theme of school education management is always willing to impart quality education to its students. In this paper we focused higher level to apply Data Mining techniques for implementing and give a quality in technical education. Many ways to achieve to the highest level of quality in technical education as well as increase the student's academic performance and predict that performed and underperformed of the student's for providing training and placement. In our work the data set can be prepared from student's academic's (technical training or higher education) like a student's roll no., student's name, student's date of birth, student's 10th, student's 12th, academic percentage upto 7th semester student's database have to take. In this paper we propose to create a data set for five departments, we have taken (CSE, IT, ECE, EEE, MECH) each department 10 students were taken in preparing the data set, after preprocessing the data set final data can be obtained for training and placement from performing and underperformed student's from each department. In Educational Data Mining knowledge is hidden we can retrieve the knowledge through data mining techniques. By this process we extract knowledge that measure student's performance at the end of the semester examination. It helps earlier in identifying performed and underperformed student's who needs a special attention in academic wise and based on that we give training and placement for the student's. Classification method, decision tree, ASSISTANT algorithm were used. In future this study will be assisted to develop new concepts of data mining techniques in technical education.

Keywords: Data Mining, discover knowledge, Technical Education, Knowledge Discovery in databases

I. INTRODUCTION

Data Mining is a process of excerpting unknown, valid, potential useful and hidden patterns from large data sets. Nowadays data increasing fastly, vast amount of data that can be stored in educational database. Even though we get a necessary profits from such large amount of data and to find out hidden relationships between various databases and using different data mining techniques were developed and used (Han and Kamber, 2006). Even the cost of processing ability and storehouse is coming down easily, data storage became easier and cheaper so the amount of data stored in educational databases is accelerate rapidly. Despite of to get benefits from such large database to find hidden relationships between different databases using different data mining techniques has developed and processed.

Manuscript Received November, 2013.

Dr. N. Venkatesan, Professor Bharathiyar college of Engineering and Technology India.

N. Chandru, Assistant Professor Sri Venkateswaraa College Of Technology India.

II. KNOWLEDGE DISCOVERY IN DATABASES

Data mining, sometimes we can say as Knowledge Discovery in databases (KDD), Approximately It has been figured that the amount of data in the world twice every 20 months. The sized and number of databases are like increases even faster. From 1989, the total amount of data has been increased in the world was estimated at five million, even though most of them are small DBASE III databases. The high technology on business activities produces a stream of data, increasing day by day data to increase more database ever-increasing stream of data because essential simple transactions, we took an example such as telephone call, the use of a credit card, or a medical test, these types of data can be regularly stored as well as typically recorded in a computer. Let us see another database like Scientific and government, these databases are day by day storing databases. Ex National Aeronautics and Space Administration has already begun to analyze much more data. Earth observation satellites, planned all previous missions combined in 1990s, are hoped-for to generate one terabyte (1015bytes) of data every day. Especially in this missions people working day and night pictures getting at a moment in a single moment but a process behind the person took several years (working nights and weekends) just to look at the pictures generated in one day. Let us See another thing, In USA government funded biological "HUMAN GENOME PROJECT" department has stored thousand's in bytes for each genetic base to become million genetic databases every day, In 1990 census in this course million and million bytes of encode patterns (That is knowledge is hidden) ways describe the stepping stone for today's modus Vivendi and social group of today's USA. The computer Science Community people have to take obligation for both scientific and applications challenges presented and need to find the knowledge aimless in the flood of the data

Various algorithms and skillful techniques already being implemented and nowadays many of the people's and many of the areas they were using these various Data mining methods like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases. These data mining methods need concise mention to have a better understanding. Distinctly, little of the data will be seen by human eyes. If not understood at all, that data have to analyze by computers. Although some of the simple applied mathematics techniques were developed data analysis year long ago, and some of the advanced techniques were used to implement the intelligent data analysis but not yet mature those. And the result is increasing present days, the gap between data generation and data understanding to overcome the gap between data generation and data understanding, growing data and expectation of the data to be intelligently analyzed and to be

presented will be the most valuable resource for the today's competitive advantage. The computer Science Community people have to take obligation for both scientific and applications challenges presented and need to find the knowledge aimless in the flood of the data. A major European expert on machine learning foreseen that next area in data mining will be explored to be use of machine learning tools as a component of large-scale data analysis "A recent National Science Foundation workshop on the future of database" research ranked data mining among the most promising research topics for the 1990s (Silberschatz, stone-breaker, and Ullman 1990).

There is a vast amount of information in the databases of current research on knowledge discovery in databases in this particular field huge heterogeneity of databases of current research on knowledge discovery in databases in this specify topic we defining as well as explaining applicable terms in knowledge discovery, is the nontrivial extraction implicit previously unknown and possibly multipurpose information from data. Given a set of facts (data) F , a language L and some measure of certainty C , we define a pattern as a statement S in L that describes relationships among a subset FS of F with a certainty c , such that S is simpler (in some sense) than the enumeration of all facts in FS . A Pattern in nothing but interesting according to the user idea some kind of measures user has to follow to create a pattern ,again according to the user's criteria patterns applied to the databases extract the useful data is called knowledge. Let's we view an example , Wal Mart is leading and masterminding monolithic one of the largely used data mining factory in USA to transform its supplier relationships. Wal mart natural process sale transactions from over 2,900 stores in 6 countries and endlessly transmit this data to its massive 7.5 terabyte "Terabyte "data that can be stored everyday in the data warehouse. Wal mart allows more than 3,500 suppliers , to access data on their products and perform data analyzes Normally Wal mart suppliers normally use their data to identify customer buying patterns at the display level, Normally a database of Wal mart use to manage local store stock lists and identify new selling possibilities. In 1995, Wal mart computers processed over 1 million complex data queries. The output of the program getting from the database, for applying the patterns to the databases signified is getting the discovered knowledge. In the following III paragraphs we conclude the significances of these terms and intimate the connection to the problem of knowledge discovery in databases

IV. DATA MINING DEFINITION AND TECHNIQUES

Generally data mining sometimes called data or knowledge discovery is the action of examining data from assorted perspectives and iterate it into useful information, that useful information can be increased the revenue the costs or both. It permits the users to analyze data to many different conceptions of the data to categorize it and summarize the relationships identified. Technologically, data mining is the activity of finding patterns in large relational databases. Data mining also called as knowledge discovery in databases, it mentions to selection or mining knowledge from large assets of data. Using Data mining techniques we can detect large amounts of data to observe hidden patterns and relations using this approach it helpful in the decision making process, While some of them treated as data mining and knowledge discovery as synonyms , normally not like

that data mining is the part of the knowledge discovery process , procedure of steps will be followed for extracting knowledge from data diagram already shown in the Fig 1.

A. Classification

First thing we are going to see "classification", this method most remarkably applied data mining technique, which chooses primed of pre classified information to create a model using this technique we can population of records at large. This approach is often needed decision tree or neural network based classification algorithms. In this classification method data classification process involves first approach learning and second approach classification .We discuss first approach based on learning, the training data are analyzed by classification algorithm , and the second approach has taken classification test data are used to estimate the accuracy of the classification rules .Based on two approaches if the accuracy acceptable rules can be applied to the new data tuples. The classifier model uses the same pre classified examples to determine the data of parameters; these parameters can be used in classifier training algorithm

B. Clustering

Clustering can be same as determination of related classes of objects. By using this clustering method we can further promote dense and sparse regions in object space and can observe whole distribution patterns and correlations among the data attributes Compare to other methods classification approach is an effective means of distinguishing groups or classes of object , but clustering approach preprocessing approach to use for subset collection and classification

C. Predication

In prediction Regression technique can be modified. The main theme of regression analysis model used to find out the relationship between the one or more independent variables and the dependent variables. In the data mining area independent variables are considered as attributes (already known values) and the response variables is what we want to predict , Unluckily some of many real world problems are not simply to predict. Some of the complex techniques like logistic regression, decision trees, or neural nets these are the necessary things to calculate the future values. The same model types can be used for both regression and classification. For Ex (CART) it means (Classification and Regression trees) ,Decision tree algorithm that can build to find out the classification trees, regression trees used to forecast the variables. Neural Networks too can create both classification and regression models

D. Association convention

In this Association and correlation is usually finding out the common item set accumulation among large dataset. So these kind of discovery helps drudging to make certain decisions such as catalog design , cross marketing, and customer purchasing behavior analysis. Usually Association Rule Algorithms need to generate rules with confidence values less than one . Even so the number of possible Association rules of the given data set high, so likewise high balance rules can be used (if any occur like that)

E. Neural networks

The neural network has connected set, each

connection sets having input and output unit, each input and output has weight present with it, During the learnedness phase networks learns by adjusting weights in the networks , so it will able to predict the accurate class labels of the input tuples. Comparisons to other things neural networks have the extraordinary ability to deduct the meaning from complicated as well as inaccurate data it can be extracted patterns and detect courses that are too complex noticed by humans, these are the well suitable for inputs and outputs. Neural networks are the high-grade notifying patterns or various drifts in data are well suited for prediction of forecasting needs

F. Decision Trees

If we see Decision tree in data mining, Decision tree is tree shaped structure and each branch represents take it of decisions. Using these decisions we can generate rules for the classification of the particular data set. A specific decision tree method will be included two things one is classification and another another one is Regression Trees , CART method is used for classification and CHAID is used for Regression Trees.

G. Nearest Neighbor Method

The other name of this method is called k-nearest neighbor technique focusing thing of nearest neighbor method is classifying objects based on the close examples placed in the data set. A method that separates each record in a data set based on the combination of classes of the k record(s) most similar data avail in a historical dataset (where k is greater than or equal to 1). Erstwhile called the k-nearest neighbor technique.

V. RELATED WORK

Recent research field in Data mining is higher education ,particularly in this research field educational institution how much gaining knowledge to increase the quality of the education

With the help of data mining how much educational field better to enhance the quality of the learning process , using the variables present in the student databases we can extract and evaluating the learning process of the students as described as Educational Data Mining Alaael -Halees [4]

The execution of data mining methods and applying the data mining tools for examining at any educational institutions as known as Educational Data mining, is the new stream in data mining research area, research field are supplied by Romero and Ventura [15], screening the research efforts in the region between 1995 and 2005, and by B a k e r and Yacef [2], for the period after 2005

Long handles in [9] many expected data mining applications that can be used in higher education , while we use data mining techniques used in academics based on the resources , while maximizing the efficiency of data mining used in academic side it saves the resources

Hijazi and Naqvic doing as study on the basis of student performance by choosing 300 students,225 male students and 75 female students taken from colleges affiliated to punjab university of Pakistan the basic proposal they were taken as "student's attitude"various student attitude from that we choose attendance in class ,hours spent study after finish the college on daily basis ,family income,mother's age,and mother's education are importantly taken as student performance was framed ,using linear regression

analysis,we identify that two factors have correlated one is mother's education and other one is student's family income were highly matched with student's performance

Al-Radaideh, et al taken as a decision tree model the predict final grade student's studied in c++ course in the yarmulke University garden in the year 2005, In this Paper author focus three classification methods were used ID3, C4.5, and the Naive Bayes. From that author predict that Decision tree model is better than other models.

Shannaq et al. [11], applied the classification as a data mining technique to predict the numbers of enrolled students by evaluating academic data from enrolled students to study the main attributes that may affect the students' loyalty (number of enrolled students). The extracted classification rules are based on the decision tree as a classification method, the extracted classification rules are studied and evaluated using different evaluation methods. It allows the University management to prepare necessary resources for the new enrolled students and indicates at an early stage which type of students will potentially be enrolled and what areas to focus upon in higher education systems for support.

Chandra and Nandhini [4], has taken as association rule mining to find out student's failed courses from that we can recognize the student's failure patterns .The main goal of the this paper to determine the failed courses and hints the applicable cause of to improve the low grade student's performance. Using association rule it can be applied to the student's failed course which bring out the hidden patterns ,it will be steeping stone for academic planners to make academic conclusions, using the curriculum resource reconstructing and alteration to improve the student's performance and reduce the failure rate Khan activity a performance study of 400 students,400 students taken an Aligarh Muslim University from Aligarh India ,200 boys and 200 girls were taken from senior secondary school, main objective to launch in this study to predictive value of different activities of to obtain a knowledge like, attribute and statistic variables using this to obtain higher secondary level in science course, main thing in this study we taken as cluster sampling technique in which student entire population is divided into groups or clusters ,for further analysis of this study student's random sample taken from groups or clusters, from this study we can identify ,if girls having socio economic status achievement is high they get a higher action in science stream, if boys is low status of academic they get general status of science stream.

Z. J. Kovacic conferred given a case study about educational data mining (EDM) what kind of data has to be taken to identify student's success. The algorithm was used in this study CHAID and CART. These two algorithms were used in student's from open polytechnic from New Zealand classifying successful and unsuccessful student's .Using decision tree classifying students and accuracy can be obtained from CHAID (59.4)and CART(60.5)respectively. Using this study student's data can be analyze based on student learning behavior to promise the student's and give counsel to students to avoid the risk before their final exams

VI. DATA MINING PROCESS

In present day's educational data mining in educational system student's activities can be determined using various fields, like a test, seminar, assignments, and



lab work etc. can be carried out based on educational activities. End of the semester student gets scored based on various educational activities. Each student minimum marks in internal, as well he scored in semester examination (supported things can be obtained from various student activities). From this we can promise performer and under performer of the students from each department and give a training and placements to those students

A. Data Preparations

In this study dataset is obtained from Bharathiyar college of engineering and technology from karaikal,(India),5 departments were taken from the session 2005 to2009 ,each department 10 students has taken and size of the dataset is 50.In this preparations different tables is joined into single table after joining table errors is removed

B. Data selection and transformation

Data selection process in the educational data mining which field are needed which things are necessary extract the database, from the analyzing we can predict performed and underperformed from 5 departments student'

TABLE 1 STUDENT RELATED VARIABLES

Variable	Description	Possible Values
PSM	Previous Semester Marks	{First > 60% Second >45 & <60% Third >36 & <45% Fail < 36% }
CTG	Class Test Grade	{Poor , Average, Good }
SEM	Seminar Performance	{Poor , Average, Good }
ASS	Assignment	{ Yes, No }
GP	General Proficiency	{ Yes, No }
ATT	Attendance	{Poor , Average, Good }
LW	Lab Work	{ Yes, No }
ESM	End Semester Marks	{First > 60% Second >45 & <60% Third >36 & <45% Fail < 36% }

The domain values for some of the variables were defined for the present investigation as follows:

PSM – Previous Semester Marks/This marks can be obtained from 5 departments, from that data we can categorize the marks or grade can be split the marks into five different levels and investigate, the data department wise . It is split into five class values: **First** – >60%, **Second** – >45% and <60%, **Third** – >36% and < 45%, **Fail** < 40%.

CTG – Class test grade obtained. Each department 2 class test were conducted in each semester ,average marks of two

class test were taken as to calculate the test grade from each department ,we split the three things to investigate **CTG is split into three classes: Poor** – < 40%, **Average** – > 40% and < 60%, **Good** –>60%.

SEM – Seminar Performance obtained. In each semester seminar are organized by each department conducted by particular faculty based on the faculty subject based on the seminar we can predict that student's having how much presentation skill and how much student's having communication capablity can be evaluated from that idea we can get a idea check the performance of students. Seminar performance is evaluated into three classes: **Poor** – **Presentation and communication skill is low, Average** – **Either presentation is fine or Communication skill is fine, Good** – **Both presentation and Communication skill is fine.**

ASS – Assignment performance. In each semester faculty has given two 4 assignments are granted to the students by particular faculty handled that subject Assignment action is bifurcate into two classes: **Yes** – **student submitted assignment, No** – **Student not submitted assignment.**

GP – General Proficiency performance. In academic wise, much academic work is present like assignment, class test etc. Apart from that department or college conducted various general proficiency programmers like symposium, workshop etc. How much student's can be participated , from that we can categorize into two classes **Yes** –**student participated in general proficiency, No** – **Student not participated in general proficiency.**

Like seminar, in each semester general proficiency tests are devised. The General Proficiency test is divided into two classes: **Yes** –**student participated in general proficiency, No** – **Student not participated in general proficiency.**

ATT – Attendance by Student. Marginal 70% attendance is required by the students to take a part at the end of the semester examination from this attendance authentic reason divided into three classes, **Poor** - <60%, **Average** - > 60% and <80%, **Good** - >80%.

LW – Lab Work. Student's has to finish Lab work within the semester,and finish it off all experiments in particular period of time this lab work is divided into two classes: **Yes** – **student completed lab work, No** – **student not completed lab work.**

ESM - End semester Marks obtained End of the semester marks can be obtained , based on the semester marks and It is split into five class values: **First** – >60% , **Second** – >45% and <60%, **Third** – >36% and < 45%, **Fail** < 40%.

C. Decision Tree

“Decision tree” is tree shaped structure ,in this tree each branch node represents a choice between number of options ,and each leaf node of the tree corresponds a decision. The main purpose of the decision tree is to make the decision making process. Always decision starts with a root node from on which users have to take actions .Based upon nodes branch user split the data to the particular node , according to the decision tree algorithms. And the final outcome of the decision tree of each branch represents the manageable scenario of the decision and its outcome .Three widely used algorithms in the decision tree algorithm are ID 3, ASSISTANT, C4.5.



D. The ASSISTANT Decision tree

The ASSISTANT Algorithm [7] is a Descendant of Quinlan's ASSISTANT (1983), and incorporates a tree pruning mechanism for handling noisy data

Let E be the set of instances

A be a set of attributes for describing example

$T E (E)$ is being a termination criterion

$I D M(a_i, E)$ be an evaluation function where $a_i \in A$

Procedure Assistant (E) returning TREE:

If: E satisfies the termination criterion $T E (E)$ then return a leaf node for TREE, label labeled with the most common class of examples in E .

Else: determine the attribute $a_{best} \in A$ with the largest value of the function $I D M$ (at $a_{best}; E$). Then, for each value V_j of attribute a_{best} , generate subtrees using assistant (E_j) where E_j are those examples in E with value j for attribute a_{best} . Return a node labelled as a test on attribute a_{best} with these subtrees attached

TABLE II THE CORE OF THE ASSISTANT ALGORITHM

IV. CLASSIFICATION

Predictive modeling is the activity based on the activity a model is created or selected to give the best outcome. Many proceedings the possibility is selected on the basis detection theory, and outcome based on the amount of input classification is the prognostic data mining techniques makes predictions about the different values can be used to obtain different values can be obtained. Predictive models have a particular intention of permitting us to predict the unknown values from other variables. Predictive modeling can be thought of as Umesh K Umar Pandey et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (2), 2011, 686-690. Learning a mapping from an input set of vector measurements x to a scalar output. According to the predefined bulk of classes is often referred as supervised learning, the reason is, classes are discovered before analyzing the data. Before describing the classes by seeing or analyzing the data already known to belong to the classes.

V. BAYESIAN CLASSIFICATION

Bayes classification has been proposed based on Bayes rule of conditional probability. Bayes rule is the method to computation the probability of a property of a given set of data as indication or input as Bayes rule or Bayes theorem.

$$P(x_i | h_i) P(h_i)$$

$$P(x_i | h_i) = \frac{P(x_i | h_i) P(h_i)}{P(x_i | h_1) P(h_1) + P(x_i | h_2) P(h_2)}$$

$$P(x_i | h_1) P(h_1) + P(x_i | h_2) P(h_2)$$

The approach is called naive because it assumes the freedom between the various attributes values. Naive Bayes classification it can be considered as both descriptive and predictive kind of algorithm. The more chances for descriptive are used to predict the class relation for a target

tuple. The Naive Bayes approach has several advantages. One is it is easy to use, other classification rules only scan of the training data required. Another advantage of the naive Bayes classifier easily handles the missing value and small amount training data is estimated. Bayes classifiers have worked quite well in many complex and suitable for many real world situations.

E. Measuring Impurity

If we calculate the measuring impurity, a data table contains the attributes and class of attributes, we can measure the table based on the classes focused on two things either homogeneity and heterogeneity. Suppose if we say a table is pure or homogenous if it contains only a single class, another thing data table contains several class tables is impure or heterogeneous. There are several measures to calculate measuring impurity, many measures are calculated impurity one is Entropy, Gini index, Classification error,

Entropy $- \sum p_j \log_2 p_j$ Entropy of a pure table (consists of single class) is zero because the probability is 1 and the $\log_2(1) = 0$. Entropy reaches maximum value when all classes in the table have equal probability

$$\text{Gini Index} = 1 - \sum p_j^2$$

The Gini index of a pure table consists of a single class is zero because the probability is 1 and $1 - 1^2 = 0$. Similar to Entropy, Gini index also reaches maximum value when all classes in the table have equal probability

$$\text{Classification Error} = 1 - \text{Max}\{p_j\}$$

Similar to Entropy and Gini Index, a Classification error index of a pure table (consist of single class) is zero because the probability is 1 and $1 - \text{max}(1) = 0$. The value of the classification error index is always between 0 and 1. In fact the maximum Gini index for a given number of classes is always equal to the maximum of the classification error index because of a number of classes in, we set probability is equal to $p = 1/n$ and maximum Gini index happens at $1 - n(1/n)^2 = 1 - 1/n$ while the maximum classification error index also happens at

$$1 - \text{max}\{1/n\} = 1 - 1/n$$

F. Splitting Criteria

To determine the best attribute for a particular node in the tree we use the measure called Information Gain. The information gain, $\text{Gain}(S, A)$ of an attribute A , relative to a collection of examples S , is defined as

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

VI. RESULTS AND DISCUSSION

The dataset of 50 students was used in this study was obtained from Bharathiyar college of engineering and technology from Karaikal (Pondicherry), India already five departments were taken. Based on the academic data we created the dataset for preprocessing.

Sample Data set

Student's Performance Measuring using Assistant Algorithm

S.N O	PSM	CTG	SEM	ASS	GP	ATT	LW	ESM										
1	First	Good	Good	Yes	Yes	Good	Yes	First	24	Second	Poor	Poor	Yes	Yes	Good	Yes	Second	
2	First	Good	Average	Yes	No	Good	Yes	First	25	Second	Poor	Poor	Yes	Yes	Poor	Yes	Third	
3	First	Good	Average	No	No	Good	Yes	First	26	Second	Poor	Poor	No	No	Poor	Yes	Fail	
4	First	Average	Good	Yes	No	Average	No	First	27	Third	Good	Good	Yes	Yes	Good	Yes	First	
5	First	Poor	Average	No	Yes	Average	No	First	28	Third	Average	Good	Yes	Yes	Good	Yes	Second	
6	First	Poor	Average	No	No	Average	Yes	First	29	Third	Good	Average	Yes	Yes	Good	Yes	Second	
7	First	Poor	Average	No	No	poor	Yes	Second	30	Third	Good	Good	Yes	Yes	Average	Yes	Second	
8	First	Average	Poor	Yes	Yes	Average	No	First	31	Third	Good	Good	Good	No	Good	Yes	Second	
9	First	Poor	Poor	No	No	poor	No	Third	32	Third	Average	Average	Average	Yes	Good	Yes	Second	
10	First	Average	Average	Yes	Yes	Good	No	First	33	Third	Average	Average	Average	No	Average	Yes	Third	
11	Second	Good	Good	Yes	Yes	Good	Yes	First	34	Third	Average	Good	Good	No	Good	Yes	Third	
12	Second	Good	Average	Yes	Yes	Good	Yes	First	35	Third	Good	Average	Average	No	Average	Yes	Third	
13	Second	Good	Average	Yes	No	Good	No	First	36	Third	Average	Poor	No	No	Average	Yes	Third	
14	Second	Average	Good	Yes	Yes	Good	No	First	37	Third	Poor	Average	Yes	No	Average	Yes	Third	
15	Second	Good	Average	Yes	Yes	Average	Yes	First	38	Third	Poor	Average	No	Yes	Poor	Yes	Fail	
16	Second	Good	Good	Yes	Yes	Good	Yes	First	39	Third	Average	Average	No	Yes	Poor	Yes	Third	
17	Second	Good	Average	Yes	Yes	Good	Yes	First	40	Third	Poor	Poor	No	No	Good	No	Third	
18	Second	Good	Average	Yes	No	Good	No	First	41	Third	Poor	Poor	No	Yes	Poor	Yes	Fail	
19	Second	Average	Good	Yes	Yes	Good	No	First	42	Third	Poor	Poor	No	No	Poor	No	Fail	
20	Second	Good	Average	Yes	Yes	Average	Yes	First	43	Third	Good	Good	Yes	Yes	Good	Yes	Second	
21	Second	Poor	Average	No	Yes	Poor	No	Third	44	Fail	Good	Good	Yes	Yes	Good	Yes	Second	
22	Second	Poor	Poor	Yes	Yes	Average	Yes	Third	45	Fail	Average	Good	Yes	Yes	Average	Yes	Third	
23	Second	Poor	Poor	No	No	Average	Yes	Third										

46	Fail	Poor	Poor	Yes	Yes	Average	No	Fail
47	Fail	Good	Poor	No	Yes	Poor	Yes	Fail
48	Fail	Poor	Poor	No	No	Poor	Yes	Fail
49	Fail	Average	Average	Yes	Yes	Good	Yes	Second
50	Fail	Poor	Good	No	No	Poor	No	Fail

To solve the information gain for A relative to S, we first need to compute the entropy of S. Here S is a set of 50 examples are 14 "First", 15 "Second", 13 "Third" and 8 "Fail".

$$\text{Entropy}(S) = -P_{\text{First}} \log_2(P_{\text{First}}) - P_{\text{Second}} \log_2(P_{\text{Second}}) - P_{\text{Third}} \log_2(P_{\text{Third}}) - P_{\text{Fail}} \log_2(P_{\text{Fail}})$$

To determine the best attributes for a particular node in the tree we use the measure called Information Gain. The information gain, Gain(S, A) of an attribute A, relative to a collection of examples S,

$$\text{Gain}(S, \text{PSM}) = \text{Entropy}(S) - \frac{|S_{\text{First}}|}{|S|} \text{Entropy}(S_{\text{First}})$$

$$- \frac{|S_{\text{Second}}|}{|S|} \text{Entropy}(S_{\text{Second}})$$

$$- \frac{|S_{\text{Third}}|}{|S|} \text{Entropy}(S_{\text{Third}})$$

$$- \frac{|S_{\text{Fail}}|}{|S|} \text{Entropy}(S_{\text{Fail}})$$

Gain Values

GAIN	VALUES
Gain(S, PSM)	0.58
Gain(S, CTG)	0.52
Gain(S, SEM)	0.37
Gain(S, ASS)	0.22
Gain(S, GP)	0.04
Gain(S, ATT)	0.45
Gain(S, LW)	0.45

PSM has the highest gain, therefore it is used as the root node as shown in figure 2.

Gain ratio can be used for attribute selection, and split information can be calculated shown in the example, split information

Split Information

Split Information	Value
Split(S, PSM)	1.39
Split(S, CTG)	1.45
Split(S, SEM)	1.6
Split(S, ASS)	1.74
Split(S, GP)	1.92
Split(S, ATT)	1.51

Split (S, LW)	1.51
---------------	------

One classification rules can be generated for each path from each terminal node to root node. Pruning technique was executed by removing nodes with less than desired number of objects. **IF- THEN** rules may be easier to understand is shown in figure 3

Fig 3 Rule Set Generated by the by Decision tree

IF PSM = „First“ AND CTG = „Good“ AND ATT = „Good“ OR „Average“ THEN ESM = „First“
IF PSM = „Second“ AND ATT = „Good“ AND ASS = „Yes“ THEN ESM = „First“
IF PSM = „Second“ AND CTG = „Average“ AND LW = „Yes“ THEN ESM = „Second“
IF PSM = „Third“ AND CTG = „Good“ OR „Average“ AND ATT = „Good“ OR „Average“ THEN PSM = „Second“
IF PSM = „Third“ AND ASS = „No“ AND ATT = „Average“ THEN PSM = „Third“
IF PSM = „Fail“ AND CTG = „Poor“ AND ATT = „Poor“ THEN PSM = „Fail“

CONCLUSION

In this paper, the classification task is used in the student database to predict the student's division on the basis of previous database. As there are many approaches that are used for data classification, the decision tree method is used here. Information's like Attendance, Class test, Seminar and Assignment marks were collected from the student's previous database, to predict the performance at the end of the semester. This study will help to the students and the teachers to improve the division of the student. This study will also work to identify those students which needed special attention to reduce fuel ration and taking appropriate action for the next semester examination as well as give a placement and training program to the students.

REFERENCES

- Heikki, Mannila, Data mining: machine learning, statistics, and databases, IEEE, 1996.
- J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann, 2000
- Alaa El-Halees, "Mining students' data to analyze e-Learning behavior: A Case Study", 2009..
- Shaeela Ayesha, Tasleem Mustafa, Ahsaan Raza star, M Inayat Khan- "Data mining Model for higher education" ISSN 1450-216X Vol.43 No.1 (2010) pp24-29
- Thearling, K., "An Introduction to Data Mining" ,http://thearling.com/text/dmwhite/dm
- Westphal, C., Blaxton, T., 1998. Data Mining Solutions, John Wiley.
- www.educationaldatamining.org
- 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.
- S. T. Hijazi, and R. S. M. M. Naqvi, "Factors affecting a student's performance: A Case of Private Colleges", Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006.
- Z. N. Khan, "Scholastic achievement of higher secondary students in science stream", Journal of Social Sciences, Vol. 1, No. 2, pp. 84-87, 2005.
- 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.

