Data mining Application to Design a System for Performance Improvisation of Students in Their Academic Studies

Kalyani. M. Moroney[#], Dr. Sanjay U. Makh[*]

[#] *Assistant Professor P.G. Department of Computer Science & Technology,*
Degree College of Physical Education, Hanuman Vyayam Prasarak Mandal,
Amravati (Maharashtra)
kalyani_ph@rediffmail.com

[*] *Head, Department of Electronics Technology,*
Takshashila Mahavidyalaya, Amravati.(Maharashtra)
sanjaymakh@gmail.com

Abstract— **This paper presents a system that can be used for a performance improvisation of students in their academic studies. The system accumulates a vast amount of information which is very valuable for analyzing the student's performance and could create a gold mine of educational data. Data generated by the student and tutor is recorded in this system. However, due to the large quantity of data daily generated by the ssystem, it is very difficult to analyze this data manually. A very promising approach towards this analysis objective is the use of data mining techniques.**
This system help tutors to provide the study material related to the topic. The outcome of the result indicates that some students are good in subject or topic but some are really poor and need the help to improve in it. Using this system we can suggest Admin to necessarily concentrate on these students and to appoint experts in these subjects for teaching or preparing better study material than present.
Data mining is a tool that can extract predictive information from large quantities of data, and is data driven. It uses mathematical and statistical calculations to uncover trends and correlations among the large quantities of data stored in a database. Data mining started with statistics. Statistical functions such as standard deviation, regression analysis, and variance are all valuable tools that allow people to study the reliability and relationships between data. Much of what data mining does is rooted in statistics, making it one of the cornerstones of data mining technology.

Keywords— **Classification, Data mining, Education system, Examination system, Regression**

I. INTRODUCTION

The education process of human being is going on since its existence on earth. From pictures to action and sounds to expressions, mankind went on communicating and educating itself.

India also has rich tradition learning, teaching and overall education. The oral method of teaching is the peculiarity of Indian education system. Scholars and teachers went to guiding and teaching students orally. Knowledge, information and wisdom were passed from one generation to another only through spoken words for hundreds of years. The development of writing skills increased the efficiency of passing on knowledge and information from one person to another. Writers used to write on palm leaves and the barks of trees. The Gurukul system in ancient India is unique in itself which was prevalent in this country for centuries. At Gurukul, student had to leave at the place of Guru, who was much revered than just a teacher. At Gurukul, students used to learn different skills to increase their intellect and knowledge. Different subjects necessary for day to day life as well as to earn bread and butter were also taught in this Gurukul. In return, students used to give Gurudakshina to Guru. [1]

The present university structure came into existence in India, more than hundred years back during the British rule. For all these years, Indian education system is following a typical pattern of education. In last few years, the focus on technical education has increased much more than past.

The universities offer students the skills and knowledge they will need to work in a large number of different environments. With the global challenges and modern conditions as well as requirements, education system in India needed some changes desperately. New disciplines, decentralized system of education, small universities are the necessity of hours. Students should be inspired to use their thoughtfulness, skills and intellect, instead of focusing on fame dump recitations of prescribed syllabus used for different courses for years with only minor modifications.

In present day's educational system, a student's performance in any college is determined by the combination of internal assessment and external marks. The internal assessment is generally based upon student's performance in various evaluation methods such as tests, viva voce, seminars, class quiz, lab work, attendance, previous semester grade and his/ her involvement in extracurricular activities. External assessment of a student is based on marks scored in final examination. The external marks are totally based on the knowledge of the student. When college conducts tests or seminars certain rules may be produced. Usage of these rules may identify poor students and hence the teacher may concentrate on these students and take necessary steps to improve their performance in final examination.

Tutors have deep knowledge to apply it correctly to the learning process. Data mining is a tool that can extract predictive information from large quantities of data, which can help in developing correlations and relationships among the data to improve performance of students.

Application of Data Mining Technology to determine the plan for better performance of a student is studied in the present research. A scientific tool will be made available to the tutor, which will help him to improve the performance of the student.

II. NECESSITY OF DEVELOPING A COMPUTATIONAL TOOL

It is well known that each student has different grasping power. As every student has different abilities, on one side the fellow feels easy to understand a particular topic or a subject, just could not swallow anything of the another topic or subject even after trying hard and repeatedly. If students fail to understand the basics of the subject, it become more difficult for them to acquire the further knowledge which results in the less scoring in that particular subject, and In our present education system student's performance in any college is determined by the combination of internal assessment and external marks. The internal assessment is carried out by the teachers based upon student's performance in various evaluation methods such as tests, assignments, seminars, attendance and extension activities. The external mark is one that is scored by the student in examination. When college is going to conduct tests or seminars certain rules may be produced. Usage of these rules may identify poor students and hence the teacher may concentrate on these students and take necessary steps to improve their performance in final examination. In such condition, it is needed to facilitate the students to understand a particular topic or subject in which they are poor.

Therefore it is necessary to study, develop and design a system that can suggest a better technique to improve the performance of a student. The main objective of any higher educational institution is to impart quality education. One way to reach the highest level of quality in examination systems is by improving the decision making procedures on various processes such as assessment, evaluation, counselling and so on which requires knowledge. Tutors must have deep knowledge to apply it correctly to the learning process. In an Examination system, the candidates are required to submit the answers of all questions within the prescribed time limit. Large number of students will appear for the examination, so that the large quantity of data will be recorded in the database. Data mining is a tool that can extract predictive information from large quantities of data, and is data driven. Data mining technique can be used to improve effectiveness, efficiency and speed of the process. This research will definitely provide a correct technique of providing to tutors and help the students to improve their performance.

III. LITERATURE REVIEW

Moodle compared different data mining techniques for classifying students based on both students' usage data in a web-based course and the final marks obtained in the course [2]. He developed a data mining tool for making this task easier for instructors. Finally, the he claimed that a classifier model appropriate for educational use has to be both accurate and comprehensible for instructors in order to be of use for decision making.

An innovative knowledge-based methodology for terrorist detection can be described by using Web traffic content as the audit information [3]. This methodology can be useful for detecting terrorists and their supporters using a legitimate ways of Internet access to view terror related content at a series of evasive web sites. The detection methodology presented during the research can be applied to detecting other types of criminals surfing the Web such as paedophiles accessing child pornography sites.

Using selected classification algorithms, the potential classification techniques can be developed for human talent forecasting through some experiments [4]. The research is focussed on one of the talent management challenges i.e. to identify the existing talent regarding the key talent in an organization by predicting employee's performance using previous employee performance records in databases by using classification technique in data mining. The significance of the study is described by using data mining for talent management especially for classification and prediction. C4.5 classifier algorithm is the potential classifier in this experiment. These generated classification rules can be used to predict the potential talent for the specific task in an organization. In HRM, there are several tasks that can be solved using this approach, for examples, selecting new employees, matching people to jobs, planning career paths, planning training needs for new and senior employee, predicting employee performance, predicting future employee and etc.

In the comprehensive review of different classification techniques in data mining, it has been mentioned that the Classification methods are typically strong in modeling interactions [5]. Several of the classification methods produce a set of interacting loci that best predict the phenotype. However, a straightforward application of classification methods to large numbers of markers has a potential risk picking up randomly associated markers.

Michigan State University (MSU), the Learning Online Network with Computer-Assisted Personalized Approach (LON-CAPA). LON-CAPA servers developed the data mining aspects of the latest online educational system that is recording student's activities in large logs [6]. This tool can help instructors to design courses more effectively, detect anomalies, inspire and direct further research, and help students use resources more efficiently. An advantage of this developing approach is its broad functionality in many data mining application domains. Specifically, it allows for contrast rule discovery with very low minimum support.

IV. PERFORMANCE OF STUDENT

A Introduction to performance

Student's performance has been the subject of ongoing debate among educators, academics, and policy makers. There have been many studies that sought to examine this issue and their findings point out to hard work, previous schooling, parent's education, family income and self motivation as factors that have a significant effect on the student's performance.

Study effort, age of student, the participation of student in class discussion all has positive effect on student's performance. Student's performance is significantly correlated with satisfaction with academic environment and service received. The existence of professional development programs and internship opportunities are associated with better academic performance.

If we came to know the factors that negatively affect the student's performance then it should help policy makers to design and implement policies to improve students' performance on one hand and improve the efficiency of education on the other hand.

B. Academic performance

1) Definition: In educational institutions, success is measured by academic performance, or how well a student meets standards set out by local government and the institution itself. As career competition grows ever more fierce in the working world, the importance of students doing well in school has caught the attention of parents, legislators and government education departments alike.

2) Significance: Education has proved immensely useful over the years to attain progress and achieve success in the world. Progress of students in schools and colleges is the matter of concern for all and efforts are already taken to identify, evaluate and boost the progress of students in schools or colleges. Good academic results have helped to attain better careers, better jobs and job security. For many, good career has become synonymous with the good and higher education. As a result of which parents are increasingly caring about their child's academic performance.

Methods of evaluating performance of students also have changed over the years. Earlier, there were verbal and oral methods of judging the performance of a particular student. Many things were largely depended on the observations of teachers. Now, in recent past various summation or numerical methods are evolved to gauge the performance of the students. The evaluation of learning of students varied more or less as the teachers changed. Attempts of standardization were made in the evaluation of performance of students still variations continued. Recently, methods are found out to analyze abilities of each student differently and alternate methods are also devised for that. The process of finding newer such methods is still going on and researchers are consistently studying that.

3) Function: The evaluation of the academic performance of a student helps to find out areas of academic achievement and failures of a student. Such evaluation gives scope to improve the performance, strengthen the weaker areas and making maximum use of learning process. Results obtained after the evaluation provide a framework about the level of performance of students and on the basis of which some decisions can be taken to further improve the performance. Sorting, classification and ranking of students can be done on the basis of results. Numeric evaluation using right evaluation processes also helps to reduce errors in evaluation and thereby the complaints regarding it. It also increases the accountability of teachers, schools and colleges.

4)Feature: There can be number of ways of evaluation of performance of students. Written and oral tests give opportunity to students to exhibit their knowledge of the subject. Classroom activities like making presentations, group discussions and others also helps students to demonstrate knowledge of students.

5) Consideration: The subjectivity in the evaluation of academic performance of students has not been eliminated entirely. Due to human involvement, it may not be possible to fully remove subjectivity from the current evaluation methods, since most are based on traditional teaching methods. The evaluation of students is based on the ability of students of responding to these traditional methods.

Standardized testing is best responded to by students that excel in reading, mathematics and test-taking. But that is not self sufficient to indicate the academic excellence of students. The standardized test fails to recognize students with learning and physical disabilities. Evaluations from classroom teachers may still retain bias if individual is not judged differently, individually and different learning styles are not considered. [8]

V. RESEARCH METHODOLOGY

A. Importance of study of regression

The study of Regression is of immense use in this research work because of the following reasons:

1. Regression is used to examine the relationship between one dependent and one independent variable. After performing an analysis, the regression statistics can be used to predict the dependent variable when the independent variable is known. In this study we used knowledge of student (independent variable) to predict the performance in examination (dependent variable).

2. Once we know that two variables are closely related, we can estimate the value of one variable when the value of another variable is known. This is known with the help of regression analysis. For example, a regression model could be used to predict the value of a house based on location, number of rooms, total size, and other factors.

3. In actuality regression takes a set of data and fit the data to a formula. Regression when used for classification, the input values are the values from database and output values represent the classes. In this study we used knowledge of topic in each subject as an input from database and predict the performance of student.

Regression goes beyond correlation by adding prediction capabilities. People use regression on an intuitive level every day. In business, a well-dressed man is thought to be financially successful. A mother knows that more sugar in her children's diet results in higher energy levels. The ease of waking up in the morning often depends on how late you went to bed the night before. Quantitative regression adds precision by developing a mathematical formula that can be used for predictive purposes.

B. Linear Regression

Regression can be performed using many different types of techniques. One of the techniques is simple linear regression. In this study we used simple linear regression.

[9]. Simple linear regression fits a straight line through the set of n points in such a way, which makes the sum of squared residuals of the model as small as possible.

The formula for Simple linear regression is :

y = c0 + c1x1 + ……..+ cnxn

By determining the regression coefficients c0 , c1,……… cn the relationship between the output parameter y, and the input parameters x1, x2……. xn can be estimated.

Regression can be used to perform classification using two different approaches:

1. Division: The data is divided into regions based on the class.
2. Prediction: Formulae are generated to predict the output class value.

The first case views the data as plotted in n-dimensional space without any explicit class values shown. Through regression, the space is divided into regions-one per class. With second approach, a value for each class is included in the graph and using regression, the formula for a line to predict class value is generated.

In this study we used the first technique i.e division process. The class to which a student is assigned is based only on the numeric value of his or her marks obtained in examination. We used linear regression concept to determine how to distinguish between the poor and good performing classes. We used the linear regression formula y = c0 + ∈. In this we found the value for c0, that best partitioned the marks, from which we differentiated between poor and good performer.

If the linear regression formula is y = c0 + ∈, then the sum of the squares of the error is :

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - c_0)^2$$

To minimize the error taking the derivative with respect to c0 and setting equal to zero we get

$$-2 \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} C_0 = 0$$

And solving for c0 we get,

$$C_0 = \frac{\sum_{i=1}^{n} y_i}{n}$$

VI SYSTEM OVERVIEW

The system developed using SQL is to be used for conducting online examinations for different courses. In other words, it will be easier to conduct online examinations using this system and the system will be applicable to all kind of courses.
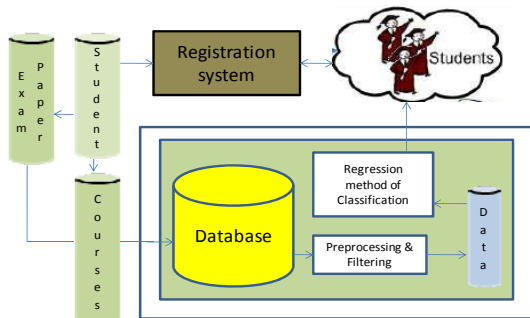

Fig. 6.1 Architecture of the system

The user of the system can add new course name in the format provided in the system. Along with addition of courses, the user can also edit the list of course provided. As like courses, subjects in respective courses can also be added. Two options, General and Course-wise are provided in the format. Clicking on General options, two sub options, Technical and General Aptitude will be displayed, selecting any of those sub-options, will display subjects under those heads. E.g. Clicking on General option and technical sub option, subjects which are part of General-Technical will be displayed.
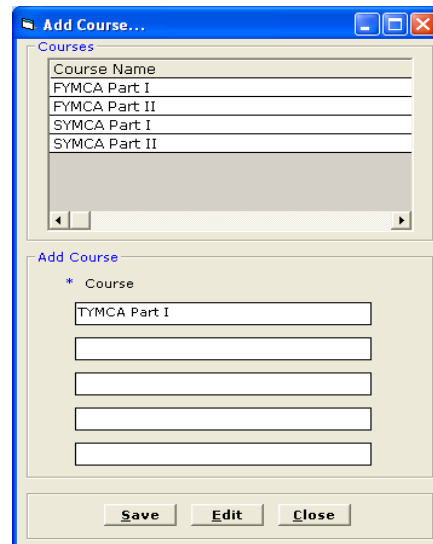
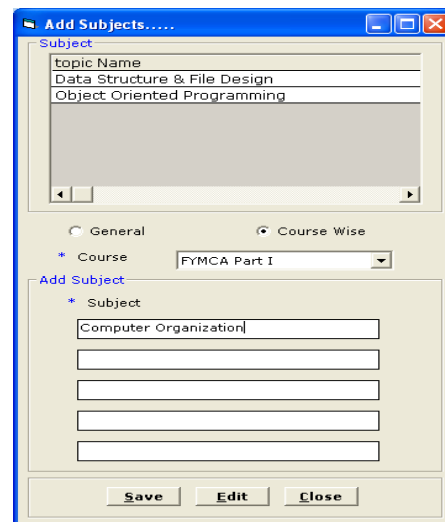

Fig. 6.2 Enter new Course

Fig.6.3 Enter new Subject

On selecting Course-wise option, a list of various courses will be displayed and the user will have to select the course. On selecting a specific course, subjects of that course will be displayed. User can add or edit the subjects.

User can add subjects in the courses as well as topics corresponding to subjects . Even user can add technical and aptitude subjects and topics. Subsequently   questions related to the topic can also be included. Topics other than already provided in the list can also be entered here. This system also has question bank facility which has questions related to different courses, subjects, topics and difficulty

levels. User can set difficulty level up to 4. These questions can be used for setting question papers. Here also, there are two sub options, General and Course wise along with their sub options as like earlier. User can enter the questions subject wise and topic wise. The questions are of objective and subjective type. User can create paper sets of any subject. Objective and subjective type question papers of different difficulty level can be set and the time required for the paper can also be defined. After setting paper of any subject user can conduct the examination of student.
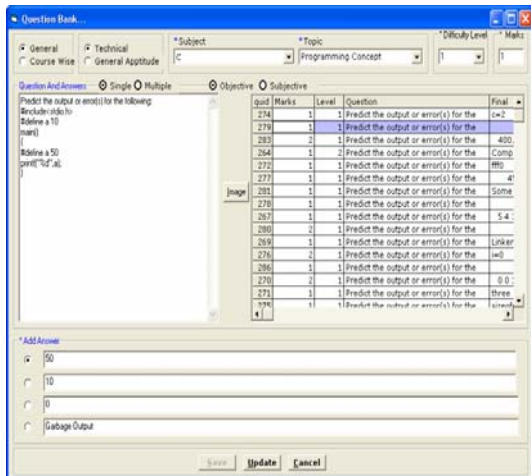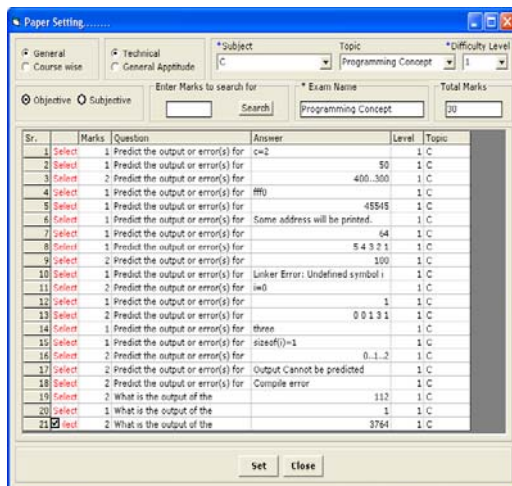

Fig. 6.4 Question Bank

Fig.6.5 Set a Question Paper

While starting the exam, the system will provide the options, Practice Examinations and Scheduled Examinations. Practice examinations are for the practice of students. Scheduled examinations include different examinations conducted by formal educational establishments. A system will show marks obtained in both scheduled and practice examinations but marks obtained in scheduled examinations only will be considered for assessing the performance of an individual. The timing for the exam can be set and altered. At the end of the given time, the exam will end automatically. If any student completes the exam before allotted time, student will have to submit the paper and to click on the option End exam. On ending the exam, the system will show, the score achieved by the examinee. The user can conduct the same

test 'n' number of times. When user will conduct the test based on same topic 'n' number of times then it is possible that the performance of student will improve. If the performance of student is poor even after repeated tests, then there is a need of teacher intervention.

This system will provide the score sheet of student after each examination. From this score sheet user can classify the students whose performance is poor and excellent. Date wise score sheet can be displayed. The list of poor performers as well as good performers can also be displayed. When the exam is conducted 'n' number of times the graphical representation of individual is recorded in this system so that user gets an idea about the individual whether he/she is poor or good in that topic. In this graphical representation report, date of examination, average marks of that examination which is calculated using regression method of classification, marks obtained by the student in that examination and the result whether he/she is poor or good. Sometimes student is poor in first attempt but perform better in second and again give bad performance in third. Using this system user can distinguish the students using classification.

VII RESULT OF THE IMPLEMENTATION

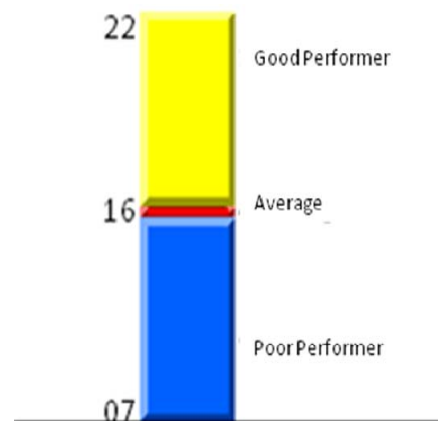Classification of marks using Division technique

Fig. 7.1 Report showing the classification of marks using division technique
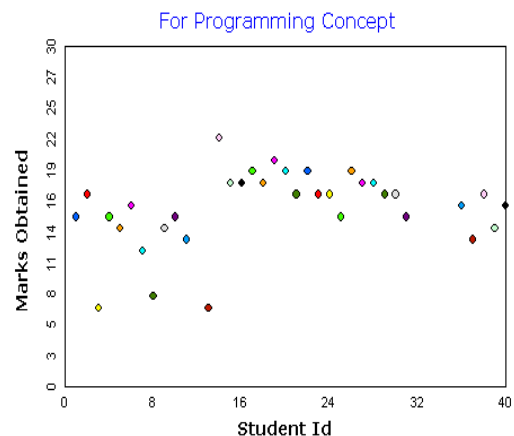
Scatter Plot

Fig.7.2 Report showing the scatter plot of marks obtained by student

TABLE 1 STUENTS WITH POOR PERFORMANCE

20-August-2011

Sr. No.	Roll No.	Name of the Student
1	F-03	Priyanka Atmaram Sawale
2	F-04	Neha Ravindra Nishandar
3	F-07	Sonia Gopal Mundada
4	F-08	sarika ravishankar yadav
5	F-09	Sneha Nandkishor Borode
6	S-04	Apurva Gajanan Gangatre
7	S-07	BHAGYASHRI SHRIPAD SONTAKKE
8	S-20	Ketaki Dipak Gogarkar
9	S-29	Payal Vilasrao Thakre
10	S-30	PRAGATI GOPAL RAHANGDALE
11	S-33	Priti Radheshyam Lakhade
12	S-34	Priyanka Kashikar
13	S-35	Puja Chandrahasya Gujar
14	S-39	roshan ramankumar rathi
15	S-41	Rujuta Ashok Palwekar
16	S-46	Shital Jaykumar Deshpande
17	S-52	Tejaswini Mahadevrao Choudhari
18	S-55	Vaishali P. Darokar

TABLE 2 STUENTS WITH GOOD PERFORMANCE

20-August-2011

Sr. No.	Roll No.	Name of the Student
1	F-05	swati vijay ganvir
2	S-03	Aniket Ganesh Yadao
3	S-14	Gauri Rajesh Jajoo
4	S-21	Kunal Panjab Gajbhiye
5	S-22	Myuri Dilip Kulkarni
6	S-23	Namrata Subhashchandji Lodha
7	S-27	Nivedita Nidhi
8	S-28	Payal Vikas Lokhande
9	S-32	Priti Dnyaneshwar Kamdi
10	S-36	Rajshri Pandurang Rauts
11	S-37	Reena Manohar Agrawal
12	S-38	Rohini Pramod Bankapure
13	S-40	Roshni Amrut Bari
14	S-44	Sangharsh Kamlakar Meshram
15	S-45	Santoshi Ramkumar Yadav
16	S-48	Sneha Shailendra Sharma
17	S-50	Supriya Panjabrao Shirbhate
18	S-51	Suraj Sahebrao Puyad

In this system Date wise list of poor performers as well as good performers can be displayed according to the regression coefficient C_0.
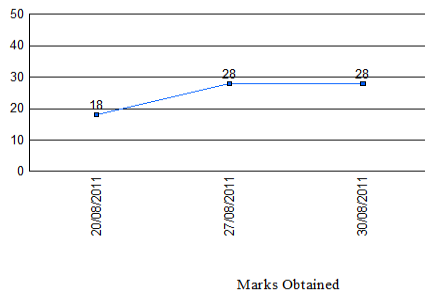
Fig. 7.3 Report showing date wise individual performance of student

VIII. CONCLUSION

In the proposed system, Regression is used for classification. Division technique of Regression is used to differentiate between poor and good performers. The solution set for every question paper is provided so that students can understand where they are wrong.

On analyzing data the following results are drawn

1. From the data generated by the system we get the list of students who are poor in any particular topic of any subject so that they can improve their performance.
2. The use of data mining in the system is useful for retrieving data from large database so that it is faster and reliable.
3. Regression is used to examine the relationship between one dependent and one independent variable. After performing an analysis, the regression statistics can be used to predict the dependent variable when the independent variable is known.
4. It is concluded that the knowledge of each topic of subject is need of today. Using this system student can appear for the exam based on the topic of any subject or entire subject.
5. Performance of students can be evaluated minutely at different levels of subjects by conducting 'n' number of tests of the students and remedies can be suggested depending on the performance.

LIMITATIONS:

- In this application only score i.e knowledge parameter is used to measure the performance of the student.
- The existing system has no options available by which a physically disabled user can use the system.

SCOPE OF FURTHER RESEARCH:

➢ The Application can be modified for different courses.
➢ Very Specific Statistical Method Regression is used for this research. Better or other method may be used.
➢ the scope of this system may be extended to increase the performance of students depending on various parameters like time factors, IQ, minimizing errors etc

REFERENCES

[1] http://education.newkerala.com/india-education/Brief-History-of-Education-in-India.html
[2] http://sci2s.ugr.es/keel/pdf/specific/congreso/Data%20Mining%20Algorithms%20to%20Classify%20Students.pdf
[3] http://www.ise.bgu.ac.il/faculty/mlast/papers/JIW_Paper.pdf
[4] http://www.intechopen.com/books/knowledge-oriented-applications-in-data-mining/data-mining-classification-techniques-for-human-talent-forecasting
[5] Proceedings of the International MultiConference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, March 18 - 20, 2009, Hong Kong
[6] http://www.lon-capa.org/papers/BehrouzThesisRevised.pdf
[7] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.107.2384&rep=rep1&type=pdf
[8] http://www.ehow.com/about_4740750_define-academic-performance.html
[9] Margaret H. Dunham, Data Mining Introductory and Advanced Topics, Pearson Education, New Delhi, India, 2003
[10] Arun K Pujari, Data mining techniques, Universities (India) press private limited, Hyderabad.