Enhanced E-Learning Using Data Mining a Rule Based Classification Approach

Gaurav Kothari, Anup Ahuje, Amish Patel, Abhinandan Khilari, Prof. Jyothi Rao

***Abstract*— E-learning environments are becoming increasingly popular in educational establishments. A popular form of e-learning is online exams and quizzes. Most of the online exams and quizzes are inadequate as they fail to provide a thorough analysis of the results due to which students are unacquainted of their weak areas in the respective subjects. Educators are finding it harder to focus on students performing poorly in respective subjects and learners are finding it difficult to make a decision about which learning materials best meet his/her situation due to more and more learning material available online. This paper presents a design of a proposed system known as E-Tutor. The proposed system offers an e-learning application integrated with an online quiz system which would predict the understanding level of students and automatically recommend favourable learning content by analysing the performance of students in online quizzes using rule based classification technique in data mining.**

***Index Terms*— e-learning, quiz, data mining, faculty, student, PACR, recommendations, classification, rules, performance level, grades, analysis, NNGE, WEKA, score.**

I. INTRODUCTION

Due to a tremendous advancement in Internet technology and World Wide Web, the concept of online learning has become quite popular among students and teachers throughout since last decade. As a result, online learning systems are extensively used in schools and colleges integrated with traditional classroom coaching to promote learning and improved understanding of concepts using audio/visual aids. These online learning systems consist of audio & video lessons, tutorials, online books, text lessons, journals, short quizzes & exams. All these contents are easily accessible to students on demand.

Online quizzes and exams are one such popular and vital form of online learning tools. It provides an array of benefits to both students as well as teachers. Some of the advantages of online quizzes are as follows:

Reduced efforts
E-learning relieves the need for testing to be done at a specific hour, but it also makes testing a hassle-free task as corrections are automated. In the cases of "Essay Question" tests, e-learning systems are usually equipped with keyword tracking tools that grade depending on what has been mentioned in the essays.

Uniqueness
Testing and quizzing can be made unique by randomizing question and answers picked from large pool of questions rather than just recycling through same sets of questions all over again.

Instant Grading and feedback
Grading manually is probably the most time consuming task for the instructor. However, in online learning systems it's just a matter of few seconds. It's where the instructor has the ability to comment on the strengths and weaknesses of a learner and enable learning to actually take place.

In-depth analysis
In-depth analysis can be done on results to extract useful information about students.

Eco-friendly
Going from hard-copy tests/ quizzes to offering the same capabilities online reduces consumption of goods such as paper.

Powerful self-assessment tool
Testing and quizzing online will usually provide the user with results instantly. This is good for students because it allows them to know what they did wrong immediately, what they need to focus on, and how to improve should they have to retake the test.

However the quizzes available online are inadequate as they fail to provide a comprehensive analysis of results. Such inadequate quizzes only notify to student, the score achieved and the questions which went right/wrong.

The task of the proposed system is to take existing system to the next level. In proposed system, the quiz section is divided into different subjects and each subject is further divided into 4 Modules wherein each module consists of quiz questions related to topic. Every question in quiz has a text explanation, video explanation, reference link of the most appropriate study material on web and PDF content associated with it. The scores of a student for all the modules of a respective subject will be analysed and fed into classification algorithm which will predict performance level of the student. If the performance of the student is poor, then the respective faculty and student will be notified and recommendations for study material will be provided to student by system as well as faculty. Further the proposed system can be used in schools and colleges along with classroom coaching to enhance learning experience.

Manuscript received April 24, 2015.
Gaurav Kothari, Department of Computer Engineering, KJ Somaiya College of Engineering, Vidyavihar, Mumbai, India
Anup Ahuje, Department of Computer Engineering, KJ Somaiya College of Engineering, Vidyavihar, Mumbai, India
Amish Patel, Department of Computer Engineering, KJ Somaiya College of Engineering, Vidyavihar, Mumbai, India
Abhinandan Khilari, Department of Computer Engineering, KJ Somaiya College of Engineering, Vidyavihar, Mumbai, India
Prof. Jyothi Rao, Department of Computer Engineering, KJ Somaiya College of Engineering, Vidyavihar, Mumbai, India

Abbreviations
PACR Performance Analyser and Content Recommender

II. LITERATURE SURVEY

A background study is done to review similar existing systems used to perform student performance analysis. Three existing system are chosen because these systems are similar to the proposed system.

A. Faculty Support System (FSS)

Shana and Venkatacalam has proposed a framework named Faculty Support System (FSS) which is low in cost as it uses cost effective open source analysis software, WEKA to analyse the students' performance in a course offered by Coimbatore Institute of Technology of Anna University [1]. FSS is able to analyse the students' data dynamically as it is able to update of students' data dynamically with the flow of time to create or add a new rule. The update of new rule is possible with the help from domain expert and the rule is determined by data mining technique such as classification technique. Classification technique is used to predict the students' performance. Besides, FSS focus on the identification of factors that contribute to performance of students in a particular course.

B. Student Performance Analyser (SPA)

SPA is existing secure online web-based software that enables educators to view the students' performance and keep track of the school's data. The SPA is a tool designed for analysing, displaying, storing, and getting feedback of student assessment data [3]. It is a powerful analyser tool used by schools worldwide to perform analysis and displays the analysis data once raw student data is uploaded to the system. The analysis is done by tracking the student or class to get the overall performance of student or class. It helps to identify the students' performance which is below the expected level, at expected level or above the expected level. This would allow the educators or staffs to identify the current students' performance easily. Other than that, it enables various kinds of students' performance report such as progress report and achievement report to be generated.

C. Intelligent Mining and Decision Support System (In Minds)

In Minds helps University Malaysia Sarawak (UNIMAS) to monitor the performance of various areas in every UNIMAS's departments [2]. The system enables top and mid-management in UNIMAS to have a clear look on the areas that needed attention by looking at the figures, revenues and risks. The features, ease of use and flexibility provided by the system makes the performance analysis in UNIMAS to be performed in an ideal solution. Charts are provided by the system for ease of student performance's interpretation.

D. Student Analysis Performance system (SPAS)

Chew Li Sa, Dayang Hanani bt. Abang Ibrahim, Emmy Dahliana Hossain, and Mohammad bin Hossin have proposed a system which offer a predictive system that is able to predict the students' performance in course "TMC1013 System Analysis and Design",. In University Malaysia Sarawak (UNIMAS) using BFTree classification algorithm which predicts whether the student is going to pass or fails in the current semester based on overall academic performance during the course of the entire semester [4].

From the reviews on these existing systems, useful techniques and features could be applied into the proposed system for a better system performance. The proposed system uses WEKA, an open source tool for data mining activities.

III. PROPOSED SYSTEM

Few features from the existing systems are included in the implementation of the proposed system. These include the simple user interface, students' performance prediction, illustration displays and report generation. In addition, the proposed system consists of quiz section and recommender which will recommend study material to students. With all these features in the proposed system, all the user requirements will be satisfied.

The proposed system is divided into following modules/entities and these modules are integrated together to form a complete application.

i. Student
ii. Faculty
iii. Quiz
iv. Performance Analyser and Content Recommender (PACR)

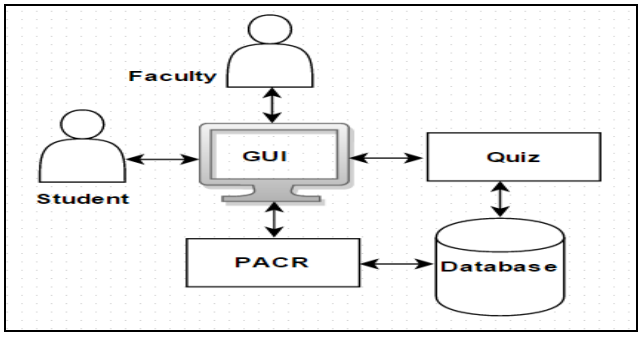

Figure 1. Architecture of proposed system

A. Student Module

Student module includes student entity. All the students of an institution will have to register themselves on the application to access its features. While registration following attributes of student are taken into consideration

Attribute Name	*Description*
Student_ID	*Id of student*
Student_Name	*Name of Student*
Roll_No	*Roll Number*
Class	*Current Class of study*
Division	*Division*
Email	*Institutional Email*
Contact	*Contact Number*

Table 1. Student Registration

Registered Students can perform following tasks

1. Access the Online quiz by selecting respective subject and module.
2. Student can check their performance module-wise/subject-wise and overall rank among other students easily with help of graph and pie charts.
3. If student is performing poor in particular chapter/subject, he/she will be automatically provided recommendations for learning material from system as well as respective subject's faculty.
4. Students can email their queries directly to respective subject's faculty. Responses will be provided within 24 hours. Student has to specify the subject name and topic followed by his doubt.
5. Student can provide his valuable feedback about the application.

B. Faculty Module

Faculty module consists of faculty entity. The proposed system is faculty controlled application hence faculty has vital role in working of application. Faculty has to register to work on application. Following attributes of faculty are taken into consideration during registration.

Attribute Name	*Description*
Faculty_ID	*Id of Faculty*
Faculty_Name	*Name*
Faculty_Subject	*Subject which faculty Teaches*
Email	*Institutional Email address*
Contact	*Contact Number*

Table 2. Faculty Registration

Faculty performs following Task

1. Faculty can add/modify/delete quiz questions.
2. Faculty can monitor performance of students' module-wise/subject-wise easily with help of graph and pie charts.
3. If student is performing poor in particular chapter/subject, respective subject's faculty will be alerted so that faculty can provide necessary tips and recommend best study content through email.
4. All of the student queries/doubts can be answered by faculty through email.
5. Faculty can assess the feedback of students.

C. Quiz Module

Quiz module consist of multi-course quiz. Quiz section is divided into subjects and each subject consists of multiple modules. Each module has pool of questions. The Questions of the quiz are added into the database by faculty only.

Creation of Quiz

Quiz questions are created, deleted or modified by faculty members using faculty panel.

Working of Quiz

Quiz section is divided into Subjects and each subject consists of 4 modules. Each module has a pool of easy, medium and hard questions. Student has to select subject and then chapter to start quiz consisting of 15 questions containing easy, medium and hard questions (5 each) picked in random order. After submitting the quiz, a detailed report will be provided to student consisting of following

1. Grade (A , B , C)
2. Questions Correct & Wrong
3. Text Explanation for every question
4. Video explanation for every question
5. Reference link of online learning material
6. PDF/Image files

Quiz Scoring

Each module consists of 15 multiple choice quiz questions containing 5 easy, 5 medium and 5 hard questions in random order. Easy question has 1 point, medium question has 2 points and hard questions have 3 points each. Hence Total score is 30.

Difficulty	***Points***	***Questions***	***Total Points***
Easy	1	5	1 x 5 =5
Medium	2	5	2 x 5 =10
Hard	3	5	3 x 5 =15
Total Points for a quiz = 5 +10+ 15 = 30			

Table 3. Total Score of each quiz

Now if a student get *x* easy, *y* medium and *z* hard questions correct out of 15 questions, then score in module *i* is of subject is calculated as

$$S_i = (x * 1) + (y * 2) + (z * 3)$$

Where Si is student's performance in module *i* for a particular subject. Subsequently, Si values for all modules in subject are calculated and converted into grades using below table.

Grade's Value	***Si Value out of 30***
A	*Above 25*
B	*15-24*
C	*Below 15*

Table 4. Grades Classification

Attribute Name	*Description*
Question_ID	*Specifies ID of Question*
Course_Name	*Name of Subject*
Chapter_Name	*Name of Chapter*
Question_Text	*Question*
Question_Type	*Specifies question Type Easy/Medium/Hard*
First	*First Option*
Second	*Second Option*
Third	*Third Option*
Fourth	*Fourth Option*
Final	*Correct Option First/Second/Third/Fourth*
Reference_Link	*Link of the study material*
File_Path	*File path of the Study material Video /Image/PDF*

Table 5. Quiz database attribute

D. Performance Analyser and Content Recommender (PACR)

This component analyses the performance grade of student in all the 4 modules of a subject and then classifies the student's performance grade in that subject as either A, B or C using Nearest-neighbor-like algorithm using non-nested generalized exemplars (which are hyper rectangles that can be viewed as if-then rules). If the final performance grade of the student is below *A* then system generates an alert to student as well as the respective subject's faculty and recommends study material to student stored in database.

Attributes	*Values*
Module1	*{A,B,C}*
Module2	*{A,B,C}*
Module3	*{A,B,C}*
Module4	*{A,B,C}*
Final_Grade	*{A,B,C}*

Table 6. Classifier Input

IV. METHODOLOGY

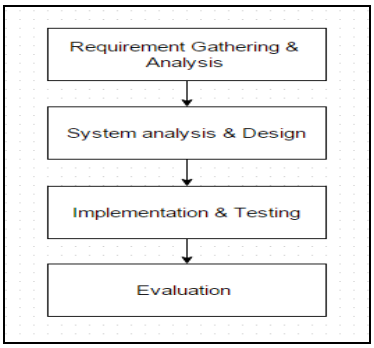

Figure 2. Development phases

A. Requirement gathering and analysis

Requirement gathering and analysis is critical in determining the success of this application. Before the system development, problems and data understanding is performed to define the project goal and objectives. The problems of the existing systems are identified and analysed for its effectiveness and efficiency in term of functionality. After the problems are identified, the solutions to solve each problem is identified and collected through literature survey. Besides, other similar systems are studied and analysed for its features, strengths and weaknesses. This helps to identify the needs and opportunities for the proposed system.

B. System analysis and design

In this phase, the overall flow of the system is planned, analysed and designed. The system and user requirements are analysed. All the required project documentation such as Software requirements specification (SRS), Project plan documentation and design documentation is done. Furthermore, all the design diagrams such as class diagram use case diagrams, data flow diagrams and ER diagrams are prepared in IBM Rational Rose to capture important aspects and functionalities of the system. Design of the proposed system also includes the design of database and user interface.

C. Implementation and testing

During the implementation phase, the grades of the student in quizzes of all the 4 modules of a particular subject is collected and student's overall performance grade in that subject is determined using IF-ELSE rules generated by Nearest neighbour algorithm students using WEKA.

During the generation of IF-ELSE rules, the test dataset is formed consisting of 92 instances which include all the possible combinations of A, B and C grades. The test dataset is pre-processed, cleaned and re-sampled before applying classifiers on it. A comparison of accuracy between different rules based classification techniques are tested to ensure that highest prediction of accuracy is achieved. Table 2 shows the accuracy comparison between five different rule based classification techniques found in WEKA.

Classifier	*Accuracy*
NNge	*98.7805%*
oneR	*81.7073 %*
PART	*85.3659 %*
zeroR	*81.7073 %*
Rudor	*85.3659 %*
JRip	*92.6829 %*

Table 7. Comparing accuracy of rule based classifiers

Nearest-Neighbor-Like Algorithm

Nearest neighbour classifiers are a class of non-parametric methods used in statistical classification (or pattern recognition)[6]. The method classifies objects based on closest training examples in the feature space. Among the various methods of supervised statistical pattern recognition, the Nearest Neighbour rule achieves consistently high performance, without a *priori* assumptions about the distributions from which the training examples are drawn. It involves a training set of both positive and negative cases. A new sample is classified by calculating the distance to the nearest training case; the sign of that point then determines the classification of the sample. The *k*-NN classifier extends this idea by taking the *k* nearest points and assigning the sign of the majority. It is common to select *k* small and odd to break ties (typically 1, 3 or 5). Larger *k* values help reduce the effects of noisy points within the training data set, and the choice of *k* is often performed through cross-validation.

```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances          81               98.7805 %
Incorrectly Classified Instances         1                1.2195 %
Kappa statistic                          0.9601
Mean absolute error                      0.0081
Root mean squared error                  0.0902
Relative absolute error                  3.7248 %
Root relative squared error             27.8072 %
Total Number of Instances               82

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.833    0          1         0.833     0.909      0.917    A
                 1        0.067      0.985     1         0.993      0.967    B
                 1        0          1         1         1          1        C
Weighted Avg.    0.988    0.054      0.988     0.988     0.987      0.967

=== Confusion Matrix ===

  a  b  c   <-- classified as
  5  1  0 |  a = A
  0 67  0 |  b = B
  0  0  9 |  c = C
```

Figure 3. NNGE Classifier Output in WEKA

```
NNGE classifier

Rules generated :
        class C IF : m1 in {C} ^ m2 in {C} ^ m3 in {C} ^ m4 in {A,B,C}  (3)
        class C IF : m1 in {A,B} ^ m2 in {C} ^ m3 in {C} ^ m4 in {C}  (2)
        class C IF : m1 in {C} ^ m2 in {C} ^ m3 in {A,B} ^ m4 in {C}  (2)
        class C IF : m1 in {C} ^ m2 in {A,B} ^ m3 in {C} ^ m4 in {C}  (2)
        class B IF : m1 in {A,B,C} ^ m2 in {A,B,C} ^ m3 in {B} ^ m4 in {B}  (9)
        class B IF : m1 in {A,B,C} ^ m2 in {B,C} ^ m3 in {B} ^ m4 in {A}  (6)
        class B IF : m1 in {C} ^ m2 in {A,B} ^ m3 in {A,B} ^ m4 in {C}  (4)
        class B IF : m1 in {A,B} ^ m2 in {A,B,C} ^ m3 in {A,B} ^ m4 in {C}  (12)
        class B IF : m1 in {A,B,C} ^ m2 in {B,C} ^ m3 in {A} ^ m4 in {B}  (6)
        class B IF : m1 in {C} ^ m2 in {A,C} ^ m3 in {A} ^ m4 in {A}  (2)
        class B IF : m1 in {B,C} ^ m2 in {B} ^ m3 in {A} ^ m4 in {A}  (2)
        class B IF : m1 in {C} ^ m2 in {A,B} ^ m3 in {C} ^ m4 in {A,B}  (4)
        class B IF : m1 in {A,B} ^ m2 in {A,B} ^ m3 in {C} ^ m4 in {A,B,C}  (12)
        class B IF : m1 in {B,C} ^ m2 in {A} ^ m3 in {B} ^ m4 in {A}  (2)
        class B IF : m1 in {B,C} ^ m2 in {A} ^ m3 in {A} ^ m4 in {B}  (2)
        class B IF : m1 in {A,B} ^ m2 in {C} ^ m3 in {A} ^ m4 in {A}  (2)
        class B IF : m1 in {A,B} ^ m2 in {C} ^ m3 in {C} ^ m4 in {A,B}  (4)
        class A IF : m1 in {B} ^ m2 in {A} ^ m3 in {A} ^ m4 in {A}  (1)
        class A IF : m1 in {A} ^ m2 in {A} ^ m3 in {A} ^ m4 in {A,B}  (2)
        class A IF : m1 in {A} ^ m2 in {B} ^ m3 in {A} ^ m4 in {A}  (1)
        class A IF : m1 in {A} ^ m2 in {A} ^ m3 in {B} ^ m4 in {A}  (1)
```

Figure 4. IF-THEN rules generation in WEKA

These rules are implemented into the IF-ELSE condition using JAVA. This classification assists the faculty members to identify those students whose overall grade in a particular subject is below A. If the grade in the subject is B or C then recommendation for corresponding modules consisting referral links and video tutorials are displayed to student. Students can take quiz as many times as possible to improve their grade. However, questions won't be repeated next time the student takes quiz. Since the questions are designed and uploaded by the faculty, there is no restriction kept on number of question in quiz database.

Entire project is developed in eclipse IDE tool and written in JSP. HTML, CSS, JavaScript is used for front end development and My SQL is used for database operations. Application is hosted on localhost using Apache tomcat server.

D. Evaluation of System

For the evaluation of the system, 20 end-users are requested to evaluate the usability of the application. This is to ensure the objectives of the proposed system are achieved as well as to ensure the ease of navigation across the interfaces of the proposed system. Moreover, the evaluation is performed to ensure that high effectiveness of the proposed system is achieved.

From the evaluation performed, a list of users' recommendation is stated as shown below

1. Including online tutorials and assignments in the application.
2. Including overall rank indicator of student in a subject among other students.

V. FUTURE WORK

Future work includes addition of online tutorials and assignment in addition to quiz so that more advanced classifiers like ID3 can be applied to get more accurate prediction results. Further the system can be used on commercial scale and the respective response and operational efficiency can be analysed.

VI. CONCLUSION

The project focuses on the enhancing of user learning experience through online quizzes. A data mining technique known as Nearest Neighbour classification algorithm which generates If-Else rules is applied in this project to ensure accurate prediction of the student's performance level in a subject by thoroughly analysing the quiz scores. The key contribution of the proposed system is that it assists the faculty in conducting student performance analysis. The system assists faculty in identifying the students whose performance is poor. Furthermore, system automatically recommends study material and even allows the faculty to recommend study material to students having low performance taking the online quiz experience to a new level.

ACKNOWLEDGEMENT

This work is project on "Enhanced E-learning using Data Mining: A rule based classification approach" under guidance of Prof. Jyothi Rao, Department of computer Engineering, K.J Somaiya College of Engineering, Vidyavihar, Mumbai.

REFERENCES

[1] J.Shana, and T.Venkatacalam, "A framework for dynamic Faculty Support System to analyse student course data", International Journal of Emerging Technology and Advanced Engineering, Vol. 2, No. 7, 2012, pp.478-482.

[2] Quality Assurance Division (2010). InMinds: Intelligent Mining and Decision Support System. Universiti Malaysia Sarawak, UNIMAS[Online].Available:http://www.qad.unimas.my/Function/ICTCompliance/inminds.html

[3] SPA (2013). What is SPA Standard? SPA Student Performance Analyser [Online]. Available:https://www.studentperformanceanalyser.com.au/spa/about.html

[4] Student Performance Analysis System (SPAS) Chew Li Sa, Dayang Hanani bt. Abang Ibrahim, Emmy Dahliana Hossain, Mohammad bin Hossin Faculty of Computer Science and Information System Universiti Malaysia Sarawak (UNIMAS) 94300, Kota Samarahan, Sarawak, Malaysia chwlisa@gmail.com,hananii@fit.unimas.my,hedahliana@fit.unimas.my,hmohamma@fit.unimas.my

[5] Rule-Based Data Mining Methods for Classiation Problems in Biomedi al Domains Jinyan Li Limsoon Wong Institute for Info omm Resear h 21 Heng Mui Keng Terra e, Singapore 119613 Email: fjinyan, limsoongi2r.a-star.edu.sg July 26, 2004

[6] Nearest Neighbor Classification Charles Elkan elkan@cs.ucsd.edu January 11, 2011