Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 12 (15): 1572-1579, 2012
ISSN 1812-5654 / DOI: 10.3923/jas.2012.1572.1579

A Performance Evaluation Tool for Behavioral Analysis of Students

K.K. Mishra, Ashish Tripathi and A.K. Misra
Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology, Allahabad, Allahabad, Uttar Pradesh, 211004, India

Abstract: The role of the examination in the current education system is very important. It helps the faculty to assess the knowledge of students in various subjects. It has a decisive role in building student's career too. But the problem is that faculties can't predict the hidden information about the students' behavior in the examination. Classification methods like Bayesian network, rule mining and decision trees can be used to extract the hidden knowledge about the students behavior. These methods can be applied on the educational data to identify the weak students and can also be used to predict the students' behavior and performance in the examination. This paper has applied the concept of decision tree algorithm on the students' first mid-semester Multiple Choice Questions (MCQs) test data to predict their performance in the next mid-semester as well in the final examination. A fully automated Performance Evaluation Tool (PET) has been proposed in this study for evaluating the behavior and performance of the students while answering MCQs i.e., whether they are confident of their marked answer or only doing guess-work to arrive at the answer. This tool will maintain the status of students in the database on the basis of the way of answering the question. Two parameters have been taken for measuring the behavior of the student i.e., confident and guess-work. It is found that in the study, PET is a result oriented and user friendly software that helps the faculties to predict the number of students who will not perform well in the upcoming lab examination and it helps the faculties to improve and bring out betterment in the result of students.

Key words: Educational data mining, classification, knowledge discovery in database, ID3 decision tree, performance evaluation tool

INTRODUCTION

It is very difficult to ignore the importance of quality education. Quality education means to provide education to the students in an efficient manner so that they can learn without any problem. Educational Data Mining (EDM) is one of the new arising techniques of data mining and it is used to extract useful information from the raw data available in the education sector. EDM can be used to take decisions for improving the quality of education and also helpful for predicting the students' behavior and their performance in the examination (Heiner *et al*., 2006; Bray, 2007). EDM contains many tasks like classification, regression, association, sequential mining, and correlation (Romero and Ventura, 2007) these tasks can be applied to bring out the various hidden knowledge from the educational database. Knowledge Discovery in Database (KDD) is a well known branch of Data Mining. KDD is automatic extraction of hidden knowledge from large volumes of data for better decision making. The aim of KDD is to discover useful information from large collection of data (Mannila, 1996; Bharadwaj and Pal, 2011a). The objective of data mining is to discover and extract patterns of stored data by applying several methods and algorithms (Fayadd *et al*., 1996; Bharadwaj and Pal, 2011a). Due to the significance in decision making, data mining and knowledge discovery applications have got a rich attention into new fields of Statistics, machine learning, artificial intelligence and databases. Decision trees induction, Bayesian classification, classification by back propagation, classification based on concepts from association rule mining other classification methods like K-nearest neighbor classifier, class-based reasoning, genetic algorithm and fuzzy set approach are some classification methods used for data analysis in Data Mining. Out of these methods like decision trees and Bayesian trees can be used to predict the student's behavior, performance and interest for a subject and outcome in the examination (Kumar and Vijayalakshmi, 2011a). The main objective of this study was to apply Data Mining techniques through which the software can analyze the student behavior and

Corresponding Author: K.K. Mishra, Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, 211004, India

also it helps to improve the overall performance of the students. There are many approaches available for data classification. Some literature review have been taken for understanding the importance of data mining techniques in the prediction of students' performance based on educational data. El-Halees (2008a) described the importance of classification and clustering in the improvement of the students' performance. Chandra and Nandhini (2010) identified students' failure pattern using mining techniques. Abu Tair and El-Halees (2012), conducted a study on how to use educational data mining to improve students' performance. Yadav and Pal (2012), presented a case study on how data mining techniques can be used to select students for admission in a particular course. Bharadwaj and Pal (2011a), checked the accuracy of decision tree techniques for predicting students' performance at the end of the semester. Bharadwaj and Pal (2011b), found that student academic performance is highly correlated with the factors like students' living location, mother's qualification, medium of teaching, students' grade in exam and other habits. Yadav *et al.* (2012), found in their study that CART is the best algorithm for the classification of data among ID3, C4.5 and CART algorithms. Ayesha *et al.* (2010) applied k-means clustering algorithm is one of the effective result oriented data mining technique to predict students' learning activities in a students' database. In this study, PET has been used for evaluating the behavior of the students while answering multiple-choice questions (MCQs) i.e., whether they are confident of their marked answer or only doing guess-work to arrive at the answer. Students immediately see their results after submitting the answer. Also, PET has a facility through which teacher can upload, view and give marks for lab assignment submitted by every student and students can see and upload their assignments for evaluation. Finally students can see their assignment marks. Other features have been discussed in the results and discussion section.

MATERIALS AND METHODS

This study has used classification method like decision tree to implement the proposed approach for the performance evaluation and behavior analysis of students. Following sections are used to describe the same.

Classification: It is a popular data mining technique for finding hidden patterns in data and it assigns items in a collection to target classes. Classification predicts

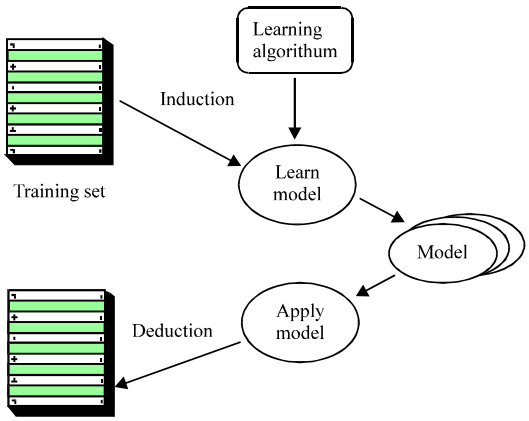

Fig. 1: Working model for classification technique

categorical data and classifies data based on the training set classification goal is to predict the accurate target classes for each case in the data. For example, loan applicants can be identified as low, medium, or high credit risk using classification model. Analysis of data using classification model can also help to identify which loan applicants are safe and which are unsafe for the bank. Other examples like data analysis on a customer profile can help the marketing manager to guess whether customer will buy a new mobile. Data related to chronic diseases like cancer and HIV/AIDS, a medical researcher use data in order to predict which one of the available specific treatments is suitable for the patient. It can be seen in the examples given above, the data analysis task is classification, where a classifier is used to predict categorical labels, such as issuing loan application is safe or unsafe, yes or no for the marketing data and treatment a, treatment b, or treatment c for the medical data. Binary classification is the simplest type of classification in which the target attribute has only two possible values like high and low credit rating. Where in the case of multiclass targets have multiple values i.e., low, medium, high or other unknown credit rating. Classification has a two step process, (1) model construction and (2) model usage. Model construction describes a set of predetermined classes in which training sets are used for construction of the model. Model usage is used for classifying future or unknown objects. It estimates the accuracy of the model using actual test set. Figure 1 shows the working model of classification technique.

Decision tree: It is a flow-chart like tree structure, in other words it is a tree-like graph, where each internal node denotes a test on an attribute and represented by rectangles, branches represents an outcome of the test

and leaf nodes represents class level and represented by ovals. In decision tree all internal nodes accommodate splits and it is used to test value of expression of the attributes. Label on the branches from an internal node to its child node represents the outcomes of the test and a class label is associated with each leaf node. The roll of decision tree is to gain information for the purpose of decision making. The working of decision tree starts with a root node. According to the decision tree learning algorithm users split each node recursively starting with root node. In the end of this whole process, each branch denotes a possible scenario of decision tree and its outcome. ID3 (Quinlan, 1986), classification and Regression Tree (CART) Bharadwaj and Pal (2011a), Chi-squared Automatic Interaction (CHAID) (El-Halees, 2008b) and C4.5 Bharadwaj and Pal (2011b) are the four widely used Decision tree learning algorithm but in this study the concept of ID3 algorithm has been used.

The ID3 decision tree: ID3 algorithm is invented by Quinlan (1986) at the University of Sydney in 1975. It is a non-incremental algorithm because it derives its classes form a fixed set of training instances. The basic idea of ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node. The idea behind developing a decision tree using ID3 algorithm is to employ a top-down, greedy search on the given sets to test each attribute at each tree node until the attributes perfectly classifies the training sets, otherwise it recursively operates on the n partitioned subsets to find their best attribute, where n is number of possible values of an attribute. ID3 algorithm uses information gain to decide which attribute is suitable for entering into a decision node. More details can be found for ID3 decision tree and algorithm in the paper written by Yadav and Pal (2012).

RESULTS AND DISCUSSION

Here, the approach for behavioral analysis and performance evaluation of students in the lab examination using Performance Evaluation Tool is discussed. Other technical details are discussed below:

Performance evaluation tool (PET): It is created by using PHP, MySQL and WAMP server. PHP is a powerful server side scripting language. MySQL is used for database management and it is an open-source database software. WAMP server is a Windows web development

Fig. 2: Login screen

environment. Apache 2, PHP 5 and MySQL are preinstalled in WAMP server. When PET is accessed through a web browser, the login screen appears as shown in Fig. 2. As noted earlier, there are three users in this software i.e., administrator, Teachers and students. Each user has different work.

There are two major objectives that can be solved by PET:

- It allows teachers to do students' behavioral analysis and performance evaluation i.e., (i) what the students think while solving the MCQ question in the lab examination, (ii) whether students are confident or only doing guess-work for the answers given by them and (iii) maintain the chapter wise status of students. When students will give their next lab exam, PET will generate questions randomly on the basis of the status maintained in the database. There are two parameters (i) confident and (ii) guess have taken for PET to evaluate student performance. There are two flags has been taken to maintain the status of the students in each chapter of a particular subject, 1 for confident and 0 for guess-work. For the second lab exam the maximum questions will be randomly generated from the guess portion of the status maintained in the database for evaluating the improvement of the students' status in comparison to the first semester lab examination. If the result is negative then the teacher will arrange a counseling session for helping weak students. Time to time teacher conducts separate lab exam for weak students to improve their performance before the end semester lab examination

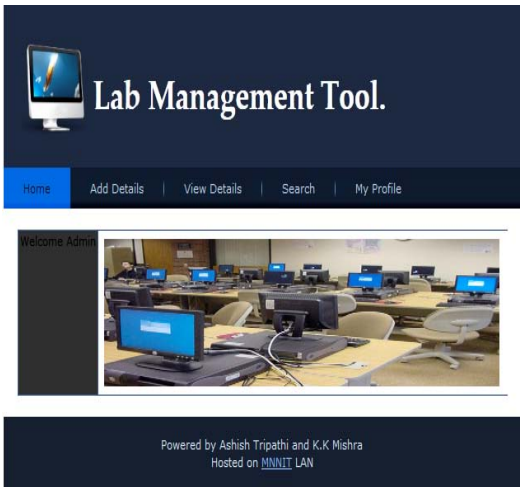

Fig. 3: Administrator module

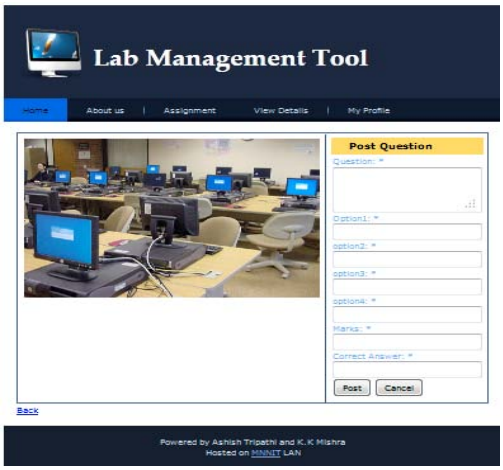

Fig. 4: Teacher module

Fig. 5: Student module

- Success and failure rate regarding the MCQs attempted can be seen by the teacher and the student. This tool is helpful for teachers as they can easily categories the students on the basis of their performance in the lab examination and also makes efforts to improve the performance level of the weak students before the end semester lab examination. PET can be applied in all educational institutions, where the paper work is a headache, only some modifications are required according to the requirement of the institutions. PET has three different modules:

- Administrator Module (Fig. 3)
- Teacher Module (Fig. 4)
- Student Module (Fig. 5)

Working of each module is different from other modules. These modules are briefly discussed bellow.

Administrator module

Add details

Add teacher: Administrator add teacher id into administrator module. This id will be matched with the id given by teachers at the time when they register for this software.

Add student: Administrator can add new student using student id and this id will be checked with the id given by the new student at the time of registration.

Add subject: Administrator can also add the subject name, total no of chapters and name of chapters in that subject.

Add programme: Administrator can add programs like M.Tech, B.Tech and MCA.

View details Edit Records: Administrator can view, edit records related to students and teachers by id. Administrator can also see the number of subjects and number of chapters in each subject by using subject id and can also update courses.

Delete records: Administrator can delete student and teacher account, subjects and no of chapters in each subject.

Search

Search student: Administrator can search student by student name.

Search teacher: Administrator can search teachers' record by using their id.

Search subject: Administrator can search students' record by using their id.

My profile
Edit profile: Administrator can view or edit his/her profile.

Logout: Administrator can logout from his/her account.

Teacher module
Assignments
Post assignments: Teacher can post assignments for students.

Check assignments: Teacher can check assignments and give marks to students.

View assignments: Teacher can edit and delete the assignments.

View details
View questions: Teacher can edit and delete questions.

View student performance: Teacher can also view the chapter wise performance of a student in a subject.

My profile
Edit profile: Teacher can also edit his/her profile.

Change password: Teacher can easily change his/her password.

Logout: Teacher can logout from his/her account.

Student module
On line test: PET will generate randomized multiple choice questions to students and they can give answer.

View result: Student can immediately view the lab examination result whether he/she is pass or fail.

View assignment: Student can view the assignment posted by teacher.

Post and edit assignment answer: Student can post and edit the assignment answer. If the answer posted by student is wrong, he/she can edit answer before the last date of assignment submission.

Edit profile: Student can edit his/her profile.

Other important features-At a Glance:

- New user can create his/her new account and also they can request for a new password
- A valid user can login with valid user name and password. Unauthorized user can not login on PET
- Individual student accounts. Students can use the same account to access any lab examination on the PET as long as they have valid id
- Password protected pages (accessible only to registered teachers and students)
- Professionally designed website template
- Extremely user friendly
- Students may view assignments and upload their answer through the browser
- PET produces automatic result, success and failure rate to the student but the teacher can also see the status of the student weather he/she is confident or not
- All registered PET users can modify their profile after registration

Technical aspects: Lots of parameters are required to be considered during the prediction of the academic outcome of a student. There are some characteristics like data related to students' background knowledge, the ability of understanding a question and the ability as well as accuracy to complete the examination in time plays a vital role in predicting their performance. Outcome of the main examination can be predicated on the basis of the marks obtained by the student in the internal examination (Kumar and Vijayalakshmi, 2011b). It can be seen in various examinations. The role of previous academic results is very important in predicting their current outcome. The author has proposed many variables that are used to predict the students who can be failed in the examination (Kovacic and Gree, 2010). Percentage of success can be improved by having knowledge about the reason of failure of students. It can help teachers to take some necessary steps for improving the success rate of students in the examination. In this proposed software the data is stored in the database. PET will automatically take data from the database, it generates random question to the students and takes feedback from the answers given by students. As discussed earlier the main objective of this Performance Evaluation Tool is to analyses students' behavior and to improve performance in the lab examination. Other related objectives are also important like less use of paper work, immediate result, fully automated assessment of the examination can be achieved by PET.

For achieving the objectives written above, decision tree has given a great support. Actually the concept of ID3 algorithm of decision tree technique has used for making this software comes true.

Data collection and preprocessing: The data is collected by the PET, when first time student gives there lab examination. When students submit their answer, PET stores the students answer in the database as well as it also store the status of the students. There are four parameters have been taken for a student to evaluate the performance and behavioral analysis. These parameters are correct, incorrect, guess and confident. Records are stored in the database in terms of these parameters. Whenever, it is required to use data for a student, PET automatically collects data from the database.

Strategy for capturing the student's behavior in the 1st lab exam:

- Quiz panel screenshot is shown in Fig. 8
- Each question of quiz panel has a drop down menu to check the student surety level.
- Surety level has for four choices:

- 0-25%
- 26-50%
- 51-75%
- 76-100%

If the student tick the answer on radio buttons but does not click on surety level that is treated as an unconfident answer or a guess or an un-attempted question:

- Value 0 and 1 is used for the flag. 0 for guess and 1 for confident

Actually the status of the student response to each question is stored as fully confident via a flag if and only if the following two conditions hold:

- If the student selects 76-100% in surety level
- The response given to question by student matches the actual answer stored in the database

In all other cases tool will identify the student response to the question as guess-work identified by the flag value 0.

After first test taken by this software the status will be stored in the data base for each student that he/she is confident or not.

Strategy for capturing the students' behavior in the 2nd lab exam.

- In the 2nd round of quiz, maximum questions will arrive from the students' status that was recorded as guess-work through the database information collected during the 1st semester lab examination. Basically the aim of PET is to identify the chapter wise weakness as well as confidence level of each student
- In the 2nd lab exam PET will give another set of questions for evaluating the improvement of students from the previous lab examination. PET will also maintain the status of each student in the chapters given in the lab examination

The benefit of this software is that it can predict weak students on time. For example the PET can predict those students whose performance is not good before the end semester lab examination and teacher may work on them to improve their performance before the final lab examination.

This software is also helpful to check the performance of the lab instructor. Whether he/she is tacking the lab classes properly or not. After each lab exam student can see his/her result immediately whether he/she is pass or fail and teacher can get the performance of each student. It is previously mentioned above that the concept of ID3 algorithm is used in PET. A table and model have been proposed below for the evaluation of the questions generated by PET. This evaluation is based on decision tree.

Information given below is used for Table 1 and Fig. 6.

- **CL:** Confidence Level, S: Sure and G: Guess are used in decision tree as mentioned in Fig. 6
- **Nil:** There is no selection on surety level
- Marks are given for each question on the basis of correct answer and surety level

Information given below is used for Table 2 and Fig. 7.

Table 1: Table for correct answer

Surety level	Confidence level	Surety (%)	Guess (%)	Marks
Nil	0	0	100	0
0-25	1	25	75	1/4
26-50	2	50	50	1/2
51-75	3	75	25	3/4
76-100	4	100	00	1

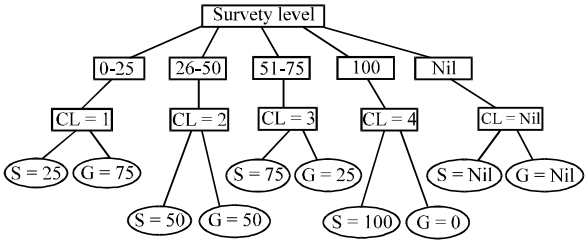

Fig. 6: Decision tree for correct answer

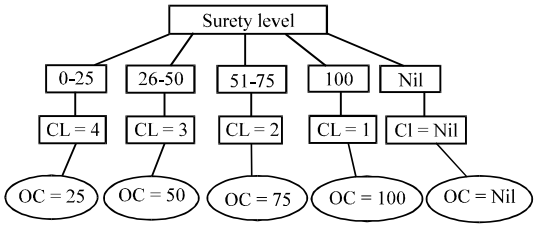

Fig. 7: Decision tree for incorrect answer

Table 2: Table for incorrect answer

Surety level	Confidence level	Over confident (%)	Marks
Nil	0	-	-
0-25	4	25	-
26-50	3	50	-
51-75	2	75	-
76-100	1	100	-

- **CL:** Confidence Level, OC: Overconfident is used in the decision tree given above
- **Nil:** There is no selection on surety level

Experimental result: In Motilal Nehru National Institute of Technology, the usage of PET is in the initial stage. This software has been applied on 80 students of B.Tech IT third year and it has been found that response of this software is very good. This tool is providing benefit to both the faculty members and students because it has given such type of platform where it reduces paper work and gives a user friendly environment in which students and teachers can interactively works together. This software is helping teachers for learning students' behavior and they are easily posting quizzes and assignments. Students are also getting benefit from this software because after giving exam they can evaluate themselves. Table 3 shows the marks given on surety level of all correct answers.

The performance of student represented in the form of IF-HEN rules of ID3 algorithm in Table 4 is given below.

Below some snap shots have been taken to show the actual implementation of the rule in PET software (Fig. 8 and 9).

Calculation of Net Confidence Level of five correct answers is presented in Table 5.

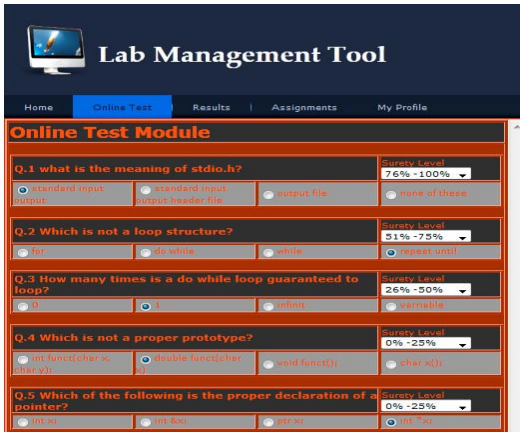

Fig. 8: Shows the lab examination quiz panel

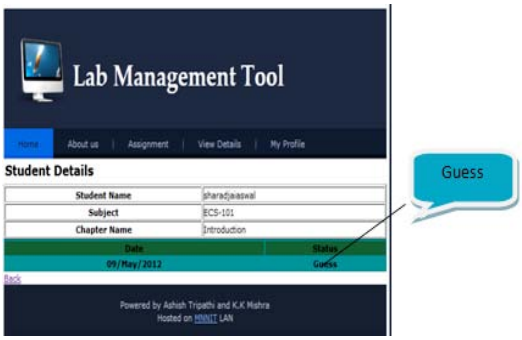

Fig. 9: Net status after the answer submitted by student

Table 3: The marks on surety level of correct answers

Question	Surety level	Surety (%)	Guess (%)	Status	Marks
1	4	100	0	1	1
2	3	75	25	1	3/4
3	2	50	50	0	1/2
4	1	25	75	0	1/4
5	Nil	-	-	0	0

Table 4: Rule set generated by ID3

Rule for surety level = 4
IF Surety (%) = 100 and Guess (%) = 0, then Status = 1 and Marks = 1
Rule for surety level = 3
IF Surety (%) = 75 and Guess (%) = 25, then Status = 1 and Marks = 3/4
Rule for surety level = 2
IF Surety (%) = 50 and Guess (%) = 50, then Status = 0 and Marks = 1/2
Rule for surety level = 1
IF Surety (%) = 25 and Guess (%) = 75, then Status = 0 and Marks = 1/4
Rule for surety level = Nil
IF Surety (%) = 0 and Guess (%) = 100, then Status = 0 and Marks = 0

Table 5: Calculation of net confidence level

				Status	
Question	Surety level	Surety (%)	Guess (%)	Confident	Guess
1	4	100	0	1	-
2	3	75	25	1	-
3	2	50	50	-	0
4	1	25	75	-	0
5	Nil	-	-	-	0

Net confidence level in five questions is 2

CONCLUSION

This study shows that the performance in first mid-semester MCQ test can be used to create the model using ID3 decision tree algorithm for the prediction of student's performance and behavior. This study help teachers to identify the weak portion of the student in a subject and it will help to improve the performance of the student in coming mid-semester and final examination. PET possesses four vital strengths: (1) it is flexible (2) it is very simple and user friendly (3) it is easy to use and (4) it is customizable. The ideas behind making this tool were to reduce the paper work, to fully automate the lab examination, to show the results immediately, to check the student performance and behavior analysis during the lab examination.

REFERENCES

Abu Tair, M.M. and A.M. El-Halees, 2012. Mining educational data to improve students performance: A case study. Int. J. Inf. Communi. Technol. Res., 2: 140-146.

Ayesha, S., T. Mustafa, A. Sattar and I. Khan, 2010. Data mining model for higher education system. Eur. J. Sci. Res., 43: 24-29.

Bharadwaj, B.K. and S. Pal, 2011a. Data mining: A prediction for performance improvement using classification. Int. J. Comput. Sci. Info., Security (IJCSIS), 9: 136-140.

Bharadwaj, B.K. and S. Pal, 2011b. Mining educational data to analyze students performance. Int. J. Adv. Comput. Sci. Applicat., 2: 63-69.

Bray, M., 2007. The shadow education system: Private tutoring and its implications for planners. 2nd Ed., UNESCO, Paris, France.

Chandra, E. and K. Nandhini, 2010. Knowledge mining from student data Eur. J. Sci. Res., 47: 156-163.

El-Halees, A., 2008a. Mining students data to analyze E-Learning behavior: A case study. Comput. Inf. Sci.,

El-Halees, A., 2008b. Mining students data to analyze learning behavior: A case study. Proceeding of the 2008 International Arab Conference on Information Technology, December 15-18, 2008, University of Sfax, Tunisia.

Fayadd, U., G. Piatesky-Shapiro and P. Smyth, 1996. From Data Mining to Knowledge Discovery in Databases. AAAI/ MIT Press, Massachusetts, USA.

Heiner, C., R. Bakery and K. Yacef, 2006. Domain specific interactive data mining educational data mining. Proceedings of the 8th International Conference on Intelligent Tutoring Systems, June 26-30, 2006, Jhongli, Taiwan,.

Kovacic, Z.J. and J.S. Green, 2010. Predictive Working Tool for Early Identification of at Risk Students. National Center for Tertiary Teaching and Learning, New Zealand.

Kumar, S.A. and M.N. Vijayalakshmi, 2011a. Efficiency of decision trees in predicting student's academic performance. Proceedings of the 1st International Conference on Computer Science, Engineering and Applications, October 15-18, 2011, Orlando, FL., USA.

Kumar, S.A. and M.N. Vijayalakshmi, 2011b. Prediction of the student's recital using classification technique. IFRSA's Int. J. Comput., 1: 305-309.

Mannila, H., 1996. Data mining: machine learning, statistics and databases. Proceedings of the 8th International Conference on Scientific and Statistical Database Systems, June 18-20, 1996, Stockholm, Sweden, pp: 2-9.

Quinlan, J.R., 1986. Introduction of decision tree. Machine learn., 1: 81-106.

Romero, C. and S. Ventura, 2007. Educational data mining: A survey from 1995 to 2005. Expert Syst. Appl., 33: 135-146.

Yadav, S.K. and S. pal, 2012. Data mining application in enrollment management: A case study. Int. J. Comput. Appl., 41: 1-6.

Yadav, S.K., B.K. Bharadwaj and S. Pal, 2012. Data mining applications: A comparative study for predicting student's performance. Int. J. Innovative Technol. Creative Eng., (IJITCE), 1: 13-19.