A Survey Paper On Academic Performance Evaluation of Students Using Educational Data

Kiran Joseph[1], Nixon Joseph[2], Prathap Prasannan[3,] Sandeep Damanan[4], Justin Mathew[5]

[1] Computer Science and Engineering Department, Saintgits College of Engineering

Kottukulam Hills, Pathamuttom, Kottayam, India

Abstract— **Educational Data Mining is one of the emerging disciplines which includes the process of analysing the students' details using different attributes. The attributes such as students' name, roll number, internal test marks, previous semester marks, attendance, assignment, seminar performance are used to evaluate the students' performance. It helps the educational institution to predict and make decisions related to the students' academic status. With a higher education, now a days dropping out of students' has been increasing, it affects not only the students' career but also on the reputation of the institute. This paper presents a survey based on the concepts, technologies and tools used in educational data analysis. This paper also aims in providing a survey on different kinds of techniques used for analysing students' performance. With this survey a solution is provided for analysing students' performance using only the internal test marks which is analysed according to the importance of topics, their occurrence in previous question papers and using a weightage evaluation algorithm.**

Keywords— *Educational Data Mining, Dropout, Educational data analysis, Academic status, Weightage Evaluation Algorithm*

I. INTRODUCTION

Analysing the huge amount of data to form summarized useful information is a tedious task for human kind. Data Mining is the area which analyses huge repositories of data to extract necessary or useful information. Computers can process any kind of data like numbers, texts, images and facts. This task performs the analysis based on the patterns, association, relations among all these data so as to get the information. The prediction with high accuracy in students' performance is beneficial as it helps in identifying the students with low academic achievements at the early stage of academics.

In universities, student retention is related to academic performance and enrolment system. Data mining is used to extract the meaningful information from large data using some patterns. It has been used in many applications such as educational data mining, web mining and text mining. Educational data mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students' and the settings in which they learn. In most colleges and universities students' academic performance is evaluated by the marks they scored for internal exams.

In this paper a solution is provided for analysing the knowledge acquired by students in each topics during their academic session. It generates list of students with their understandability in percentage.

II. LITERATURE REVIEW

Baradwaj, et al. [1] implemented an ID3 decision tree learning algorithm. In their study, they described the reasons for using decision tree algorithms. They calculated the entropy values for each attribute in the data set and then calculated information gain for each attribute. Then they selected the root node as an attribute which has the highest information gain and found which attribute was the next decision node until they ran out of attributes. Finally, their ID3 classification algorithm has generated the decision tree for weather data set.

J K Jothi and K Venkatalakshmi conducted the students' performance analysis on the graduate students' data collected from the Villupuram college of Engineering and Technology. The data included five year period and applied clustering methods on the data to overcome the problem of low score of graduate students, and to raise students' academic performance [2].

Adhatrao, et al. [3] implemented an ID3 and C4.5 Decision tree technique on educational data mining. First-year students' data was collected. Their system was used to predict the result of same students in second year. The accuracy of the ID3 algorithm is 75.145% and that of C4.5 is 75.145%. They concluded study can be used to predict the students' result based on previous semester marks.

Mythili M S and Shanavas A R applied classification algorithms to analyze and evaluate school students' performance using weka. They came with various classification algorithms, namely J48, Random Forest, Multilayer perception, IBI and decision table with the data collected from the student management system [4].

Z. J. Kovacic et al. [5] presented a case study on educational data mining to identify up to what extent enrollment data can be used to predict students' success. They used CART and CHAID decision trees and the accuracy of classifiers obtained was 59.4 and 60.5 respectively.

Osmanbegovic and Suljic conducted a study for investigating students' future performance in the end semester results at the University of Tuzla. They considered 11 factors and used classification model with highest accuracy for naive Bayes [6].

Ramaswamy et al. [7] designed a technique on students' data which has 33 features including class label. 6 feature selection techniques were applied on the data set, for selecting the relevant attributes. These attributes have a relevant value to the students

result. They used different classification algorithms. Voted perception showed the highest predictive accuracy of 89%.

Suyal and Mohod applied the association and classification rule to identify the students' performance. They mainly focused to find the students who need special attention to reduce failure rate [8].

Acharya et al. [9] applied feature selection techniques and data mining algorithms on students' data. The data have been collected from St Xavier's College, Kolkata. Different feature selection techniques were applied on the data for extracting the relevant attributes and discarding irrelevant attributes. They got 79% accuracy.

Jeevalatha, Ananthi, and Saravana Kumar presented a case study on performance analysis for placement selection for undergraduate students. They applied decision tree algorithm by considering the factors like HSC, UG marks and communication skills [10].

Tair M M A and El-Halees presented a case study with a set of data collected from degree holders of college 'Science and Technology, Khanyounis', during the period of 1993 to 2007. They used two classification methodologies such as Rule Induction and Naive Bayesian classifier to forecast the grades of the students [11].

Dinesh A and Radhika V targeted on the techniques and strategies of instructional data processing for data discovery from the information collected from various universities. This paper stated that relationship mining was leading between 1995 and 2005 and in 2008 to 2009 it slipped to 5th place. During the period 2008 to 2015 45% papers are moving to prediction. The prediction model acts like a warning system to improve their performance [12].

Remesh, Parkavi, and Yasodha conducted a study on the placement chance prediction by investigating the different techniques such as Naive Bayes Simple, MultiLayerPerception, SMO, J48, and REPTree by its accuracy. From the result they concluded that MultiLayerPerception technique is more suitable than other algorithms [13].

III. PERFORMANCE EVALUATION USING COMPREHENSIVE WEIGHTAGE EVALUATION ALGORITHM

A. Syllabus And Question Paper Upload

The teacher who login to the system, using subject code, uploads syllabus and one previous university question paper of the subject. The subject code is entered followed by the file uploading of question paper and syllabus in pdf format is done. Syllabus may contain several common words which need to be removed before next step of processing.

B. Keyword Matching

The question paper and syllabus is extracted to a text file. The extracted text from questions and syllabus are compared and keywords are found. Keywords represent the topics present in syllabus which occurs in the question paper. The number of occurrences of each keyword in the question paper is displayed with a numerical value after the keyword matching process. Multiple occurrence of any keyword can also be displayed.

C. Ranking

Ranking is done to generate an order in which the keywords are displayed based on the weightage and occurrence. The weightage is entered by the teacher according to the importance of topic. The ranking is calculated using the equation:

$$R = A \times B,$$

where A = occurrence of the topic, B= weightage given to the topic. After ranking is done, teacher hast to identify the questions in internal exams which belongs to the generated keywords. Then the corresponding question number with their total marks are entered according to each topic.

D. Analysis

This is the final step which provides the understandability of students for each topic. In the analysis phase mark sheet of students for any internal exam is uploaded. This mark sheet contains roll number, name and marks obtained by students for each question. After uploading mark sheet, it is then mapped with the question number belonging to each topic and its total mark. Result of each student for each topics is displayed in percentage using the comprehensive weightage evaluation algorithm.

The algorithm says:
- Fetch uploaded marks of students from array.
- Fetch total marks of each topic from array.
- Iteratively divide mark of each student with corresponding total mark for individual question.
- Return the percentage per topic per student.

$$G = \frac{x}{y} \times 100$$

G= grade in percentage
x = mark obtained by student
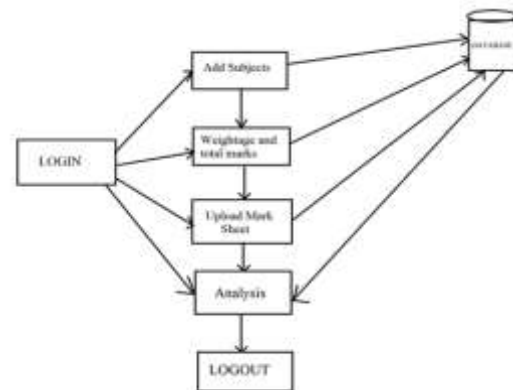y = total mark of the question
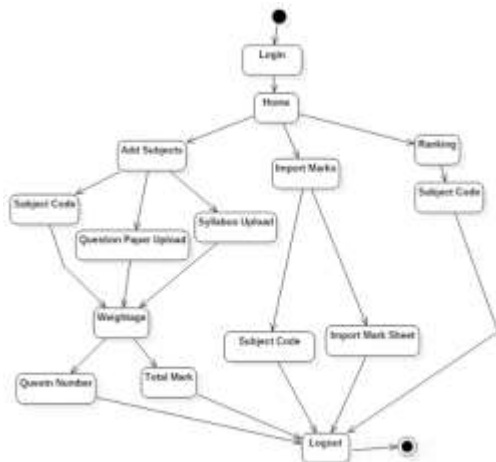

Fig. 1 Block diagram

Fig. 2 Activity Diagram

Fig. 4 Screenshot 1 Analysis

IV. RESULTS

The comprehensive weightage evaluation algorithm seems to be efficient in analysing the academic performance of student for each topic of a subject.

A group of 10 students are selected and their marks with question numbers are stored in an excel sheet which is uploaded to the system. After uploading syllabus and question paper 5 topics were generated.

Teacher gives weightage for the topics and then total marks and question numbers were also inserted. Final list displayed the 10 students with their understandability in each topic in percentage.
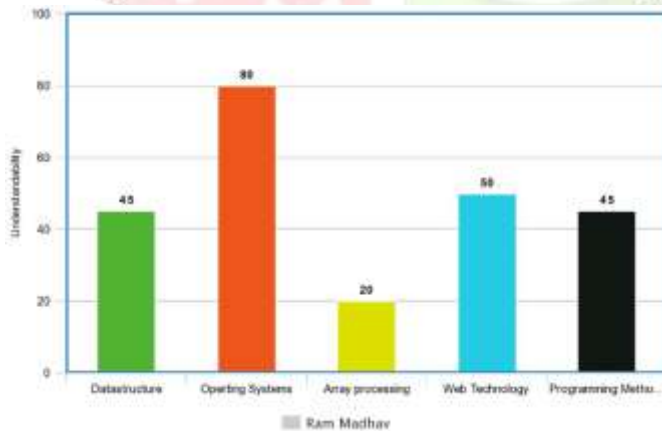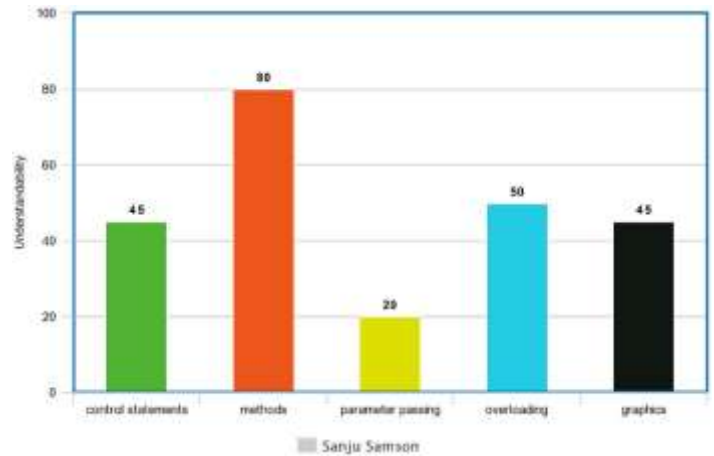
VI. CONCLUSION

Various surveys have been studied and the merits and demerits of each paper have been observed. Various techniques and algorithms have been used in the existing work in analysis of students' performance. Each technique have taken some factors for the purpose of analysing and all the considered factors does not provide an accurate output which helps the instructors in the institution to take a proper decision. Finally by observing all factors a proposal is given as a framework of analysing the students' academic performance using marks obtained for internal exams and important topics from syllabus.

The proposed system seems to be efficient and unique in a way such that it helps teachers to analyse each student's performance in each topic of a subject.

Fig. 3 Analysis per topic per student

V. SCREENSHOTS

VII. REFERENCES

[1] Baradwaj, B. K., and Pal, S. 2012. Mining educational data to analyze students' performance. arXiv preprint arXiv:1201.3417.

[2] J.K.Jothi and K.Venkatalakshmi, "Intellectual performance analysis of students by using data mining techniques", International Journal of Innovative Research in Science, Engineering and Technology, vol 3, Special iss 3, March 2014.

[3] Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R., and Honrao, V, "Predicting Students' Performance Using ID3 and C4.5 Classification Algorithms." arXiv preprint arXiv, Vol.3, No.5, September 2013, pp.39-52

[4] M.S. Mythili and A.R.Mohamed Shanavas , "An analysis of students' Performance using classification algorithms ", IOSR-JCE, Volume 16, iss1, Jan. 2014.

[5] Kovacic, Z. "Early prediction of student success: Mining students' enrolment data."

[6] E. Osmanbegović and M. Suljić, "'ata mining approach for predicting students performance", Economic Review, vol 10, iss 1, 2012.

[7] Ramaswami, M., and Bhaskaran, R. 2009. A study on feature selection techniques in educational data mining. arXiv preprint arXiv:0912.3924.

[8] Sayali Rajesh Suyal and Mohini Mukund Mohod, "Quality improvisation of student performance using data mining techniques", International Journal of Scientific and Research Publications, vol 4,iss 4, April 2014.

[9] Acharya, A., & Mukherjee, S. "Modeling Value Chain Analysis of Distance Education using UML." International Conference on Modeling, Optimization, and Computing. Vol.1298. No. 1. AIP Publishing, 2012.

[10] T.Jeevalatha, N. Ananthi and D.Saravana Kumar, "Performance analysis of undergraduate students placement selection using Decision Tree

Algorithms", International Journal of Computer Applications (0975- 8887), vol 108, December 2012.

[11] Mohammed M.Abu Tair and Alaa M.El-Halees, 'Mining Educational Data to Improve Students' Performance: A case Study', International journal of information and Communication Technology Research, ISSN: 2223-4985, vol.2 no.2, February 2012.

[12] A.Dinesh Kumar and V.Radhika, "A survey on predicting student performance", International Journal of Computer Science and Information Technologies, Vol. 5, 2014.

[13] V.Ramesh, P.Parkavi and P.Yasodha, "Performance analysis of aata mining techniques for placement chance prediction", International Journal of Scientific and Egineering Research , Vol.2, iss 8, August 2011.