Learning Pathway: Analytical Framework to Predict Learner's Learning Effectiveness and Performance

Vinayak Hegde, Sahana Patil, Smruthi G

Abstract: *In recent years, computerized online tests play a significant role to judge the learning of the students in their academics. The conventional process of examining students for academic performance take a long time and thus, identifying and understanding the learning behaviors at that time is very crucial and difficult. The pen and pencil tests that are used to assess the student's learning way by the teachers take a lot of time. Thus, a computerized test can serve as an early mechanism for testing the students wherein the evaluation results are produced immediately and thus, reducing the time for the teachers for further assessment. The students learning and understanding behaviors are reflected through their scores and performances in these tests. The present model helps the teachers to predict the weak learners and their area of weakness in studies so that special coaching and interests can be given on such students to improve their academics strength. In recent years, the JAVA programming language is extensively used for several kinds of examinations. This web-based testing environment is a JAVA based application through which the results are analyzed and are visualized to students and the reports are provided as feedback to the teachers. The analysis uses regression to predict the performance based on the y value calculated. The correlation coefficient between the average score and average time spent on chapters was found to be 0.474, which indicates the moderate correlation between them. This kind of examination helps students to self-evaluate their learning results and thus, can improvise on weaker areas.*

Index Terms: *Correlation, Education data mining, Online test, Performance prediction, Regression*

I. INTRODUCTION

With the increase in trends in the recent education system, online learning and tests are emerging wherein the students learn and improve their knowledge through these mechanisms. The online education sector has evolved over a period and is now growing with various innovative ideas as they see enormous potential in this business [1].

For years now, schools have worked with two predefined

    Vinayak Hegde*, Department of Computer Science, Amrita School of Arts & Sciences, Amrita Vishwa Vidyapeetham, Mysuru, India.
    Sahana Patil, Department of Computer Science, Amrita School of Arts & Sciences, Amrita Vishwa Vidyapeetham, Mysuru, India.
    Smruthi G, Department of Computer Science, Amrita School of Arts & Sciences, Amrita Vishwa Vidyapeetham, Mysuru, India.

purposes in mind – to customize learning experiences for individuals based on their ability, and to personalize instructions. However, to accomplish this objective, schools need to have more data that can predict a student's ability/potential in a particular task [2]. These educational models help students to self-evaluate themselves and learn concepts on their own through predictions done on the data [3]. And it also gives educators and students an edge in understanding where and how improvements can be made in the learning process. The information captured from each student is helpful to analyze the student's academic performance, involvement in other extra-curricular events and attendance patterns. The results can be collected to predict which students are expected to become detached, drop out of school and failures at the right time. Data mining helps in generating meaningful information on the data provided [4]. The data mining in education is referred to as Educational Data Mining (EDM) which is a subject related to data mining, machine learning and statistics to provide information that is generated by the educational backgrounds [5]. This field is closely related to learning analytics for assessing the learners based on their learning behaviors.

The data analytics helps in providing:
- Measure, Monitor, and Respond
- Personalize the learning experience

The conventional method of assessing students in educational institutions had several shortcomings. The time taken to assess and to provide feedback to the students was not efficient and the students could not self-assess them on their subject knowledge. These drawbacks resulted in high rates of failure and inefficient learning behaviors. To overcome this, the students must be provided with a platform to self-evaluate themselves to improve in their weak areas. The system must also provide a feedback mechanism to the tutors to identify students who are lacking in academics and provide an extra tutorial so that the failure rates are reduced [6]. Online learning systems have the capability to capture learner's behavior and can work on the data to provide a variety of stakeholders with feedback to improve teaching, learning, and educational decision making [7]. These systems help in providing a test environment wherein the students are assessed and proper feedbacks are given for improvisation.

Learning analytics [8] is mainly used for:

a) Prediction: to recognize the students who are at risk in their academics [9]
b) To provide students with a learning pathway and tutoring them
c) Personalization and adaptation, to deliver students with tailored learning pathways through materials
d) Information visualization: provides an overview of learning data through data visualization tools [10]

The prediction of the data captured can be devised through several prediction mechanisms [11],[12]. These algorithms include:

- Decision tree
- SVM
- Neural networks
- Regression
- Random forest and so on.

II. RELATED WORK

C.S. Koong, H.Y. Tsai, Y.Y. Hsu, and Y.-C. Chen [13] proposed a model that was generated using the algorithms like SVM, neural network, random forest and Naive Bayesian Model to predict students' semester results for a year. The analysis was also done for student's behaviors that were logged in the system. This was used to perform behavioral progressive analysis from which a study was made on the variances in the behaviors of students with poor results with that of other students. The outcome of the study exhibited that the automatic grading model could certainly venture a projection on semester end results w.r.t midterm grades, having a precision rate of 60%.S. C. Cheng, Y. T. Lin, and Y. M. Huang [14] proposed a dynamic question creation system for web-based exams using the unique methodology of particle swarm optimization (PSO). This model selects tailored questions from the question bank repository for each learner to satisfy various evaluation requirements. It was proved that the method is suitable for the selection of near-optimal questions from huge question banks when it was compared with the other approaches. A. Abu Saa [15] aimed to explore multiple factors that were theoretically presumed to have an effect on student's performance in higher education and to discover a qualitative model which best categorizes and forecasts the student's performance based on associated personal and social aspects. In the analysis, it was noticed that CART had an excellent accuracy of 40%, which was more than the expected correctness, CHAID was next with 34.07% and the least accurate was ID3 with 33.33%. Interesting patterns were formulated using the Naive Bayesian classification method on the dataset. Also, in the study, it was observed that not only the student's performance was reliant on their academic factors, but also on many other aspects that had equal to higher impacts. Using the same dataset, many other data mining methods can be applied. S. K. Yadav [16] analyses the classification algorithms that are effective for students' performance prediction in academics. Classification approaches were applied to the educational data for predicting the student's performance. This prediction will support to determine the weak students and provision them to score better grades in their academics. In this study, C4.5, ID3 and CART decision tree algorithms are used. The result of the decision tree shows the prediction of student's performance who were expected to pass/fail. The experiment showed that the C4.5 method has maximum accuracy when it was compared to other methods with 67.7778% accuracy. ID3 and CART algorithms also exhibited a satisfactory level of correctness.

Ahmad, F., Ismail, N. H., & Aziz, A. A [17] provides a framework for students' academic performance prediction and learning behavior. The study uses Naive Bayes, Decision Tree and Rule-Based classification techniques that are used on the students' data that are collected to yield the best students' academic performance prediction model. The framework demonstrates the steps involved in developing models to predict Student Academic Performance (SAP) of the first semester BCS students at the FIC, UniSZA. The experiment outcome shows that Rule-Based technique is the efficient model when it was compared with other methods by having the maximum accuracy value of 71.3%. The information from the prediction will be used to recognize and profile the student to govern the students' academic success level in the first semester.Cortez, P., & Silva, A. M. G. [18] predict student achievement and to identify the vital factors that affect educational success/failure. The outcome of the study shows that the predictive accuracy is high when the grades of the first and/or second term is known. This proves that previous performances of the student in the school highly influence the achievement. Binary classification, 5-Level classification, Regression, Decision tree, Random forest, Neural networks, SVM are the algorithms that are used in the study.

III. DESIGN
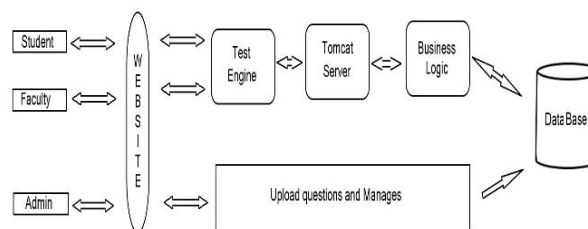
Fig. 1. Overall design

The web application provides Java subject tests that are in a chapter-wise manner. The admin has the privileges to add, delete and view the questions that are stored in the database. Once the student takes up the test, the test questions are randomly picked from the database and are given to the students. On completion of each test, topic-wise results are visualized to the students and teachers. In addition to this, Java chapter-wise content is provided to the students and the time spent on each topic is also recorded and tracked. The system helps in providing the feedback results of the student which in turn help the students as well as the teachers to track the performances. Thus, making self-evaluation easier and decreases the evaluation time. The present system provides a framework to capture the data of each student to predict their performances. In this study, regression analysis is used to predict the performance of individual student.
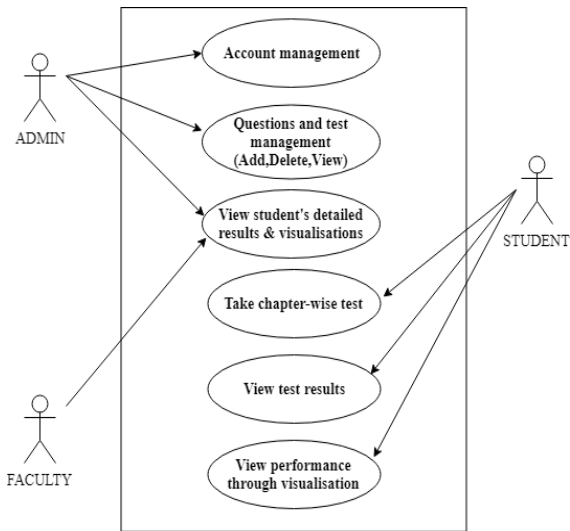
Fig. 2: Use case diagram

Fig 1 & 2 depicts the overall architecture and the roles of each actor in the system respectively.
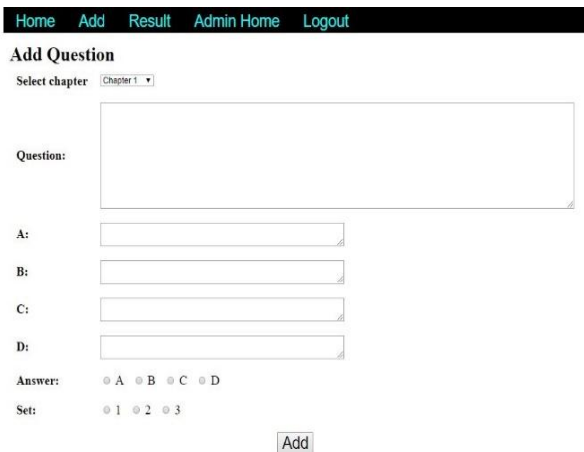

Fig. 3.Admin privileges to add questions

Fig. 3. Shows the interface wherein, the admin can add the questions and provide the correct answer for each chapter according to the subject.
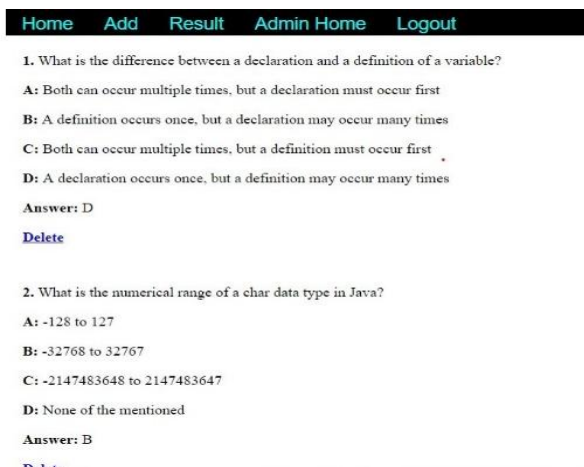

Fig . 4.Admin privileges to view or delete the questions

The admin has the privilege to view and delete the questions from each chapter which is shown in Fig. 4.

Fig . 5. Test process

Fig. 5. Shows the interface provided to the students for taking up tests in a chapter-wise manner. The questions are randomly picked from the database and is provided to the students and the answers are collected back in the database and verified with the right answers.

IV. METHODOLOGY

A. Algorithm

Linear regression is a type of predictive analysis used to find the relationship between two variables. The outcome of this technique can be used to make predictions about the data. The relationship between the variables can be defined on a straight line. The simple linear regression plots one independent variable against the other dependent variable wherein, the independent variable is termed as the predictor variable and the dependent variable is termed as the criterion variable. This analysis is used to find equations that fit the data, which can be further used to make predictions.

A linear regression line has the equation:

$$y = b_0 + b_1 * x \tag{1}$$

where,

x- the explanatory variable
y- the dependent variable
b_0- the slope of the line and
b_1- the intercept

To perform a regression analysis, b_0 and b_1 must be solved wherein, the regression coefficient b_1 has the following formula:

$$b_1 = \frac{\sum[(x_i - \bar{x})(y_i - \bar{y})]}{\sum[(x_i - \bar{x})^2]} \tag{2}$$

And b_0 is calculated using the below formula:

$$b_0 = \bar{y} - b_1 * \bar{x} \tag{3}$$

The linear regression is done on the trained dataset which is collected from 30 students of Amrita School of Arts and Sciences, Mysuru.

The test data is predicted using (1) which is calculated from the trained dataset wherein the x value is substituted for prediction results for the test data. Based on the y values calculated the performance is categorized into excellent, good, below average and bad.

Table 1. Performance categorization

y value	Performance
$0 \leq y \leq 2.5$	Bad
$2.6 \leq y \leq 3.5$	Below average
$3.6 \leq y \leq 4.5$	Good
$4.6 \leq y \leq 5$	Excellent

Table 1 shows the performance categorization based on the predicted value using linear regression method. Correlation coefficient: It is a numerical value between -1 and 1 which depicts the strength of the linear relationship between the two variables. The value of r closer to 1 indicates that the relationship is strongly positive, closer to -1 says that the relationship is strongly negative and 0 indicates that there is no relationship between the variables.

Correlation coefficient r is calculated using:

$$r = \frac{Sxy}{\sqrt{Sxx}\sqrt{Syy}} \tag{4}$$

where,

$$Sxx = \frac{\sum(x_i - \bar{x})^2}{n} \tag{5}$$

$$Syy = \frac{\sum(y_i - \bar{y})^2}{n} \tag{6}$$

$$Sxy = \frac{\sum[(x_i - \bar{x})(y_i - \bar{y})]}{n} \tag{7}$$

The correlation coefficient r is calculated using (4) between the average scores obtained and the average time spent on the chapters. The student's performance is grouped into excellent, good, bad and below average. Through this analysis, the students are categorized and based on it the learning methods can be enhanced and proper attention can be taken on them.
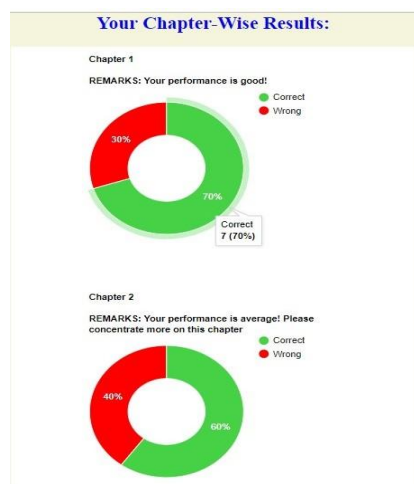

Fig . 6.Performance visualisation and remarks

The test results and performance of the students are provided through visualizations and the remarks are also given to each student based on the performance. Both students and teachers can view performance visualization.
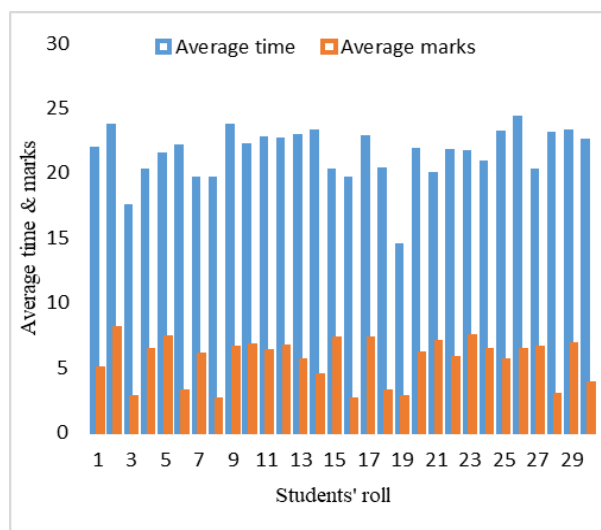
V. RESULTS

Fig. 7. Performance chart

The correlation coefficient value was found to be 0.474 between the average score and the average time spent. Thus, there is a moderate correlation between these factors which indicates the dependency between them. The prediction for student performance is done using linear regression wherein the performances are categorized into excellent, good, below average and bad based on the y values calculated. This can be used as an evaluation mechanism and also for the self-assessment of the student in academics.

VI. CONCLUSION

The model helps in tracking the student learning progress in academics. The newly discovered knowledge from the regression analysis help students and also teachers in carrying out better enhanced educational excellence, by recognizing probable poor performers at an early stage, and apply additional attention on them to help in their academics and attain better grades. It also permits students to self-monitor the progress, deliver regular feedback, provision peer learning and assessment and design self-evaluation practice.

REFERENCES

1. Jassó, J., Milani, A., & Pallottelli, S. (2008, September). Blended e-learning: a survey of on-line student assessment. In *Database and Expert Systems Application, 2008. DEXA'08. 19th International Workshop on* (pp. 626-630). IEEE.
2. Hegde, V., & Pallavi, M. S. (2015, December). Descriptive analytical approach to analyze the student performance by a comparative study using Z score factor through the R language. In *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)* (pp. 1-4). IEEE.
3. Winters, F. I., Greene, J. A., & Costich, C. M. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review*, *20*(4), 429-444.
4. Delavari, N., Shirazi, M. R. A., & Beikzadeh, M. R. (2004, May). A new model for using data mining technology in higher educational systems. In *Information Technology Based Proceedings of the FIfth International Conference onHigher Education and Training, 2004. ITHET 2004.* (pp. 319-324). IEEE.

5. Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society*, *17*(4), 49-64.
6. Bhardwaj, B. K., & Pal, S. (2012). Data Mining: A prediction for performance improvement using classification. *arXiv preprint arXiv:1201.3418*.
7. Devasia, T., Vinushree, T. P., & Hegde, V. (2016, March). Prediction of students performance using Educational Data Mining. In *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)* (pp. 91-95). IEEE.
8. Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, *47*, 168-181.
9. Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *40*(6), 601-618.
10. Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. Computers & Education, 113, 177-194..
11. Ahmed, A. B. E. D., & Elaraby, I. S. (2014). Data mining: A prediction for student's performance using classification method. *World Journal of Computer Application and Technology*, *2*(2), 43-47.
12. Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., & Punch, W. F. (2003, November). Predicting student performance: an application of data mining methods with an educational web-based system. In *Frontiers in education, 2003. FIE 2003 33rd annual* (Vol. 1, pp. T2A-13). IEEE.
13. C.-S. Koong, H.-Y. Tsai, Y.-Y. Hsu, and Y.-C. Chen, "The Learning Effectiveness Analysis of JAVA Programming with Automatic Grading System," *2018 IEEE 42nd Annu. Comput. Softw. Appl. Conf.*, vol. 1, pp. 99–104, 2018.
14. S. C. Cheng, Y. T. Lin, and Y. M. Huang, "Dynamic question generation system for web-based testing using particle swarm optimization," *Expert Syst. Appl.*, vol. 36, no. 1, pp. 616–624, 2009.
15. A. Abu Saa, "Educational Data Mining & Students' Performance Prediction," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 5, pp. 212–220, 2016.
16. S. K. Yadav, "Data Mining : A Prediction for Performance Improvement of Engineering Students using Classification," vol. 2, no. 2, pp. 51–56, 2012.
17. Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. *Applied Mathematical Sciences*, *9*(129), 6415-6426.4
18. Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance.

AUTHORS PROFILE

Mr. Vinayak Hegde Assistant Professor Department of Computer Science. His area of research is Educational Data mining and Learning Analytics and developer in Advanced Java and Java Enterprise Edition.

Sahana Patil a student of Amrita School of Arts & Sciences, Mysuru. She is a BCA graduate from Amrita Vishwa Vidyapeetham University (2017) and is currently pursuing her Masters in Computer Applications. Her area of research includes Data mining, Learning Analytics and Networking.

Smruthi G a student of Amrita School of Arts & Sciences, Mysuru. . She has completed her BCA at Amrita Vishwa Vidyapeetham University (2017) and is currently pursuing her Masters in Computer Applications. Her area of research is Learning Analytics and grid computing.