Improving Learning Experience of Students by Early Prediction of Student Performance using Machine Learning

Radhika Lagad Department of Computer Engineering, SKNSITS, Lonavala, India, radhikalagad555@gmail.com

Saurabh Dhotre Department of Computer Engineering, SKNSITS, Lonavala, India, saurabhdhotre5154@gmail.com

Rahul Pise Department of Computer Engineering, SKNSITS , Lonavala, India, rahulpise137@gmail.com

Vedant Chachane Department of Computer Engineering, SKNSITS , Lonavala, India, chachanevedant09@gmail.com

Guide - Prof Balasaheb Tandale Computer Department , SKNSITS, Lonavala, India, bctandale.sknsits@sinhgad.edu

Abstract :. Early indicators of students' progress enable teachers to focus on a variety of teaching methods and optimize their learning tactics in order to create an effective learning environment. Academics can proactively engage students in more effective learning experiences by using machine learning applications to anticipate potential shortcomings in learning processes. Using historical data of student grades in one of the undergraduate courses, we applied logistic regression, linear discriminant analysis, K-nearest neighbors, classification and regression trees, gaussian Naive Bayes, and support vector machines. From there, we developed a model to predict the grades of students taking the same course in the following term. Based on our research, we have shown that the best method for accurately predicting students' performance in final exams is linear discrimination analysis.

Keywords—Machine Learning, Predictive Analysis, Student Performance Prediction, Linear discrimination analysis.

I. INTRODUCTION

In academia, there has been a noticeable transition from traditional instruction to a student-centered curriculum. The traditional method of teaching involves giving lectures to students, who then passively take in the information. Teachers then assess their students' learning using various methods. On the other hand, the student-centric approach to learning promotes an effective learning environment for students. As a result, it becomes imperative that teachers constantly assess students' development and adjust their instructional methods. As a result, formative assessments are now an essential technique used by teachers to evaluate how well their students are learning.

Through the use of an early warning system, machine learning applications have the potential to greatly assist instructors in identifying pupils who are performing poorly. This allows the teachers to concentrate more on these underachieving pupils.

II. system requirements and design

A.Hardware Requirements

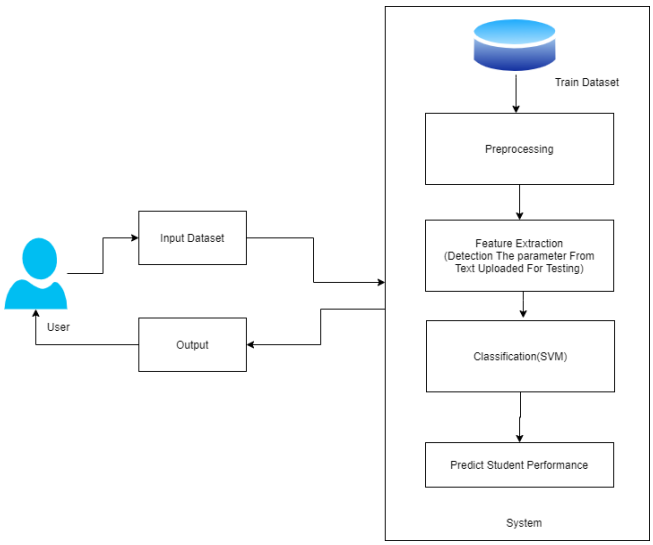
Processor	: Pentium-IV
RAM	:512 MB(min)
Hard Disk	: 40 GB
Key Board	: Standard Windows Keyboard
Mouse	: Two or Three Button Mouse
Monitor	: LCD/LED

B. Software Requirements

IDE	: Spyder
Coding Language	: Python Version 3.8
Operating System	: Windows 10

C. Design

 DOI : 10.35291/2454-9150.2024.0198



III. MATH

The current system has poor accuracy. As a result, the suggested system suggests an improved system. In addition to other student variables, behavioral attributes can be included in the dataset. To increase accuracy even further, other algorithms can utilize parameter adjustment. It is currently impossible to forecast and suggest student achievement with any degree of accuracy. On the other hand, the suggested approach will help students perform better by emphasizing their strengths or pinpointing their areas of weakness as they study.

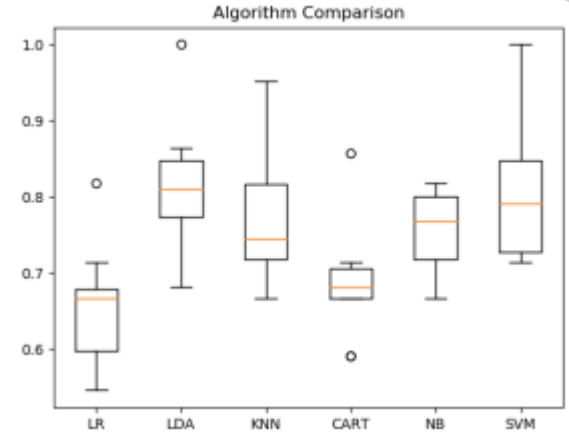
(LR) Logistic Regression

K-Nearest Neighbors (KNN) using Linear Discriminant Analysis (LDA).

Regression and Classification Trees (CART).

Naive Bayes Gaussian (NB).

SVMs, or support vector machines.



IV. UNITS

Precision, Recall, F1-Score and Kappa:

TABLE I. MEASURES AND THEIR VALUES

Algorithm	Precision	Recall	F1-Score	Kappa
LR	68.89	68.89	71.11	60.73
LDA	80	80	77.78	74.75
KNN	82.22	82.22	82.22	79.36
CART	71.11	73.33	68.89	77.18
NB	82.22	82.22	82.22	77.12
SVM	86.67	86.67	86.67	86.26

V. HELPFUL HINTS

Figures and Tables

Actual past data from students enrolled in undergraduate programs is gathered for the classes of 2016, 2017, 2018, and 2019. There are 250 undergraduate students in the dataset. Each student's data includes grades as the aim and the marks earned in examinations (Quiz 1: 5 marks, Quiz 2: 5 marks, Midterm Exam: 20 marks, Project: 15 marks, Lab: 15 marks). The distribution of each student's marks for each attribute/parameter is displayed in Figure 1:\

The first step in the implementation process is gathering the data set needed for the project. The methodology is applied to a dataset that contains student data. By recognizing the distinctive characteristics of the data gathering, it can reduce the scope of our analysis. Analysis can't be done with it. The data is transformed into the desired format after it has been collected. This is the mechanism at work. This procedure is known as data pre-processing. This is the most important phase. The higher the rate of raw material pre-processing accuracy, the better for extracting the necessary information from the raw data. It is preferable if the admissible data has a higher accuracy rate. The current system has poor accuracy. As a result, the suggested system suggests an improved system .In addition to other student variables, behavioral attributes can be included in the dataset. To increase accuracy even further, other algorithms can utilize parameter adjustment.

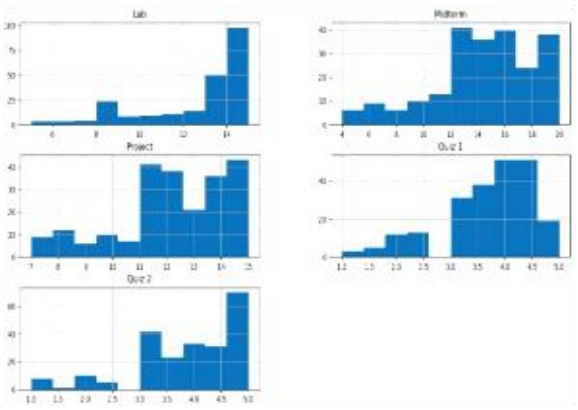


Fig. 1. Marks Distribution

VI. LITERATURE SURVEY

The suggested approach forecasts students' academic performance, allowing instructors to intervene in a timely

 DOI : 10.35291/2454-9150.2024.0198

manner to encourage students to become more involved in their studies and achieve better results. This study offered a model for predicting student performance that is based on the supervised learning approach of linear regression. This approach will help numerous educational institutions raise their students' performance levels. These should employ performance data to quickly identify various student profiles. This will allow content to be quickly modified to meet the needs of the students. It is also a helpful tool for teachers and students to evaluate the development of their pupils. Making sure that students have a good learning experience is something that the student-centered learning approach promotes. Therefore, it's critical for teachers to monitor their students' development and modify their teaching strategies as necessary. Formative assessments have therefore grown in importance as a method for educators to measure the effectiveness of the teaching and learning process. It is crucial to predict a student's academic achievement since it can warn teachers of students who may be at risk. Forecasting can assist you if you drop out of the event. More assistance for students who need to raise their GPA and their academic successes. This dataset should be seen as a tabular format containing student information (such as age, gender, academic records, and medical information). Several algorithms can be used to develop a model that produces results for this dissertation. Algorithms are utilized in a variety of ways. models that can be predicted. The focus of this research is on how linear the student's academic performance has regressed and the performance based on the dataset provided by the students

To several research articles that are related to the thesis in order to specify the thesis as a well-structured thought. The following are the details of a few of the publications' conclusions. The linear regression technique is described in detail in this research report. This method is used to forecast a student's academic achievement. In this research paper, a model named Augmented is proposed to predict the academic performance of college students. These experimental results demonstrate that Augmented can predict academic performance with quite a high accuracy, which helps to formulate personalized feedback for at-risk (or unself-disciplined) students . This research paper proposed various predictive models trained on several ML and DL algorithms for predicting students' performance based on demographics variables, demographics and clickstream variables, and demographics clickstream and assessment variables.

VII. Conclusion

Highly helpful in enhancing students' educational experiences in the proposed system. By identifying low-performing kids early on, algorithms for data mining can support this process and help teachers adjust their teaching methods while maintaining the attention of these students. To evaluate the accuracy of students' achievement in our research, we employed a course's data set. After that, we detected student performance using an SVM classifier. The suggested approach forecasts students' academic performance, allowing instructors to intervene in a timely manner to encourage students to become more involved in their studies and achieve better results. This study offered a model for predicting student performance that is based on the supervised learning approach of linear regression. This approach will help numerous educational institutions raise their students' performance levels.

VIII. Acknowledgment

The suggested approach forecasts students' academic performance, allowing instructors to intervene in a timely manner to encourage students to become more involved in their studies and achieve better results. This study offered a model for predicting student performance that is based on the supervised learning approach of linear regression. This approach will help numerous educational institutions raise their students' performance levels. These should employ performance data to quickly identify various student profiles. This will allow content to be quickly modified to meet the needs of the students. It is also a helpful tool for teachers and students to evaluate the development of their students.

The predictive power of DM approaches to forecast students' performance following their preparation year at the degree level in higher education was the main focus of this study. Based on the Grade Point Average (CGPA), which can be classified as high, average, or below average, the students' performance is evaluated. Three SVM classifiers—J48, RT—have been used throughout the trial to predict the students' achievement by graduation year using the student dataset. The J48 classifier outperforms in predicting students' achievement with a respectable 69.3% accuracy, according to the data. Furthermore, the CGPA for the prep year, the computer skills course, the communication skills course, and the mathematics course were significant factors that predicted the academic progress of CCSIT students. The outcomes will aid in forecasting the students' ultimate performance.

IX. References

[1] Web 2.0 Usage by Saudi Female Students for Information Sharing in Public Sector Universities: A Pilot Study, Int. J. Public Adm. Digit. Age, 2017, doi: 10.4018/ijpada.2017070106. [1] S. Saeed, H. Gull, and S. Z. Iqbal.

[2] Student Outcomes Assessment Methodology for ABET Accreditation: A Case Study of Computer Science and Computer Information Systems Programs, A. Shafi, S. Saeed, Y. A. Bamarouf, S. Z. Iqbal, N. Min-Allah, and

 DOI : 10.35291/2454-9150.2024.0198

M. A. Alqahtani, IEEE Access, 2019, doi: 10.1109/ACCESS.2019.2894066.

[3] In VFAST Trans. Educ. Soc. Sci., vol. 3, no. 2, 2014, R. K. Siddiqui, S. Saeed, and F. Wahab, "Understanding Role Of Student Feedback In Quality Assessment: A Case Study."

[4] Mumtaz, H., Saeed, S., and Wahab, F., "Perception of Pakistani Students on the Quality of University Computing Education," vol. 2, no.

[5] A collection of reflections on teaching human-computer interaction courses to undergraduate students is presented in Proc. - 2018 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2018, pp. 659–662, doi: 10.1109/CSCI46756.2018.00132, by H. Gull, S. Saeed, S. Z. Iqbal, M. Saqib, Y. A. Bamarouf, and M. A. Alqahtani.

[6] Saeed, Saeed, Aamir, and Mahmood, R., "Observations on instructing undergraduate students in database management systems," Int. J. Educ. Econ. Dev., vol. 2, no. 4, 2011.

[7] Embedding Information Systems Environment Modules in Information System CurriculumNo Title, S. Saeed, M. Saqib, and A. A. Salam, VFAST Trans. Educ. Soc. Sci., vol. 15, no. 1, 2018.

[8] Int. J. Teach. Case Stud., vol. 3, no. 2, S. Saeed, R. Aamir, and M. Ramzan, "Plagiarism and its implications on higher education in developing countries."

 DOI : 10.35291/2454-9150.2024.0198