



PREDICTING STUDENT PERFORMANCE USING MACHINE LEARNING

Sindhu¹, Raghavi², Pruthavi³, Sandhya⁴, Mrs. Shruthi Shetty J⁵^{1,2,3,4}-Student, ⁵ – Associate Professor
 Department of Computer Science and Engineering,

Alva's Institute of Engineering & Technology, Moodbidri

Abstract: This paper discusses how machine learning applications have a significant impact on teaching and learning, as well as how to improve the learning environment in higher education. We use innovative machine learning applications in teaching and learning while taking into account the students' background, academic performance in the past, and other factors. Because of the enormous class numbers, it would be difficult to support each individual student in each open learning course, which would raise the dropout rate at the end of the course. In this work, we will use linear regression, a machine learning algorithm, to predict a student's academic success.

Index Terms - Component, formatting, style, styling, insert.

I. INTRODUCTION

The Internet has opened the door to a new way of learning. The amount of information available therein exceeds that of any physical library. Although the internet can be used for a variety of purposes, students consider academic purposes to be the most desirable. In general, college students are exposed to a wide range of subjects in the core area, which can be difficult for them to comprehend at times. So they have chosen internet where there is a lot of information provided about the area of their study. Nowadays internet plays a major role in and around the people, it forces them to rely on it for everything. This is mainly affected by student's community where they use internet for various purpose like getting notes for studies, The assignment and other related activity and also for communications. Almost every engineering student is obliged to take the equivalent engineering course that has a significant impact on the field of study. The Internet has opened a huge door to a new way of learning. There is more information available there than in any real library. Although the internet can be used for a variety of purpose, academic purpose have the highest priority among students. In general, college students are exposed to a wide range of subjects in the core area, which can be challenging for them to comprehend. As a result, students have chosen the internet, where they may find a wealth of material on the subject of their research. Nowadays, the internet plays a significant part in and around people's lives, causing them to rely on it for everything. This is mostly influenced by the student community, who use the internet for a variety of purposes, including getting notes for classes, completing assignments, and other related activities, as well as for communication. Almost every engineering student is obliged to take the equivalent engineering course that has a significant impact on the field of study.

II. RELATED WORK

The many strategies for prediction utilising various data mining techniques are explored in this section. There are numerous applications and areas where prediction can be used to forecast relevant data. one such application/ area is Healthcare prediction. In the healthcare environment, there are many diseases which can be predicted before the analysis.

2.1. Student Outcomes

Outcome-based education (OBE) has recently gained widespread acceptance and implementation as a new school of thinking in education. The focus of the teaching and learning process is shifted from traditional teacher objectives to so-called student outcomes in this educational paradigm. In simple terms, student outcomes refer to the knowledge, skills, and values to be attained by the students at the time of graduation or at end of a course. The results, which represent the desired competencies, can be specified and measured at the course level (course outcomes) or at the programme level (programme outcomes). Essentially, course outcomes enable the accomplishment of program outcomes, and their alignment (i.e., courses to program) is performed in a critical activity referred to as curriculum mapping. Computerized tools were developed to assist in realizing the OBE goals and effectively document the educational assessment activities. Their utility could be increased by introducing intelligent models that can predict learning results over the course of a semester.

2.2. Student Performance

Albeit, due to the substantial educational shift in teaching and learning, i.e., OBE, student performance remains a significant concern in higher education, especially given the low grades and increasing dropout rates even at world-class universities. The cumulative GPA and course assessments are the most commonly utilised determinants of student performance and achievement, according to previous studies. Indeed, several studies used next-term course grades as the main indicator of student performance. However, it is not uncommon to measure student performance in other forms, including dropout rate, student knowledge, post-course outcomes, among other indicators. In our view, student academic performance should not be assessed using assessment grades only. Instead, it should be studied within a broader context, particularly using the student outcomes, which are now guiding the learning process by looking at the cohort performance. More over, recent research recommends exploring the prospect of predicting the attainment of student outcomes to infer student performance.

2.3. Existing Student Performance Reviews and Literature Gaps

Our detailed analysis of past surveys found that, to our knowledge, no systematic literature study concentrating on the prediction of student academic performance from the standpoint of learning outcomes has ever been conducted. Table 1 outlines the most well-known surveys on student performance prediction and highlights their strengths and flaws. Our search yielded null results.

III. PROPOSED SYSTEM

In our proposed system we have the provision for adding the details of the students result by the admin. Then the application processes the result and send the email of the url to the students to check their result data online. Here our analysis focused on understanding the forms in which the learning outcomes were measured in the selected studies. The first thing we noticed was that the synthesised literature tossed around the terms "student outcomes" and "learning outcomes" without adopting or tying them to any systematic definition. The rather vague definition of the predicted variable (i.e., learning outcomes) by the predictive models was considered a major weakness that raises concerns about the usefulness and validity of the learning analytics result year of study in the institution. The data is analyzed to build a predictive model.

Gender	Male	50
	Female	100
CGPA wise	7.1-8.0 (Model 1)	70
split up	8.1-9.0 (Model 2)	56
	9.1-10 (Model 3)	24
Year wise	III	61
	IV	89
	Total number of students	150

Collecting the data from the undergraduate engineering students based on their performance in the academics up to the current semesters is named as S1, S2, S3, S4, S5, S6, and S7. Understanding the collected data and performing analysis over it. Select the data collected and split it as training and testing dataset. After splitting the dataset, the multivariate linear regression technique is applied on the training dataset and from the regression Equation got it is used for predicting by passing the test dataset. In training dataset, there will be three categories of students based on the CGPA criteria. Based on the categories of training variables is dependent variable which is dependent on another variable called as independent variable. One should make sure that there exists a relationship between the dependent and independent variables before modelling. Strength of the relationship between the variables can be known by using the scatterplot. Linear regression line is represented in the form of:

$$Y = a * X + b$$

Y-Dependent Variable x

a-slope x

X-Independent Variable

x b-Intercept

dataset, the models will be predicted. From the whole dataset, a sample category of students is considered for explanation. The values represent the regression coefficient of the variables (independent variables). The first value represents the intercept of the model.

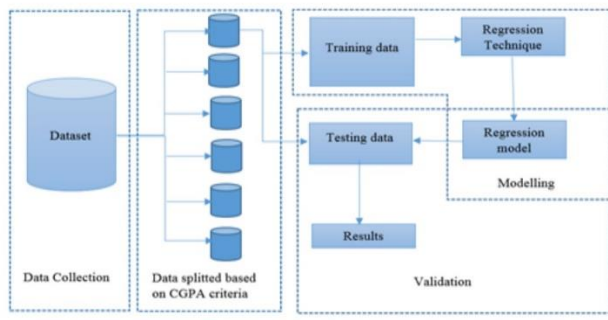


Fig-2 flow predictive model implementation

IV. ALGORITHM USED

There are many algorithms that are used to implement the thesis. In this thesis we use linear regression. Though they are all used to predict the dependent variable based on independent variables, they differ in implementation of the algorithm. A. Linear Regression Linear regression is one of the machine learning algorithms. It is based on supervised learning, which is a well-known method that can be easily comprehended even by those who are unfamiliar with machine learning algorithms. As the name suggests linear regression performs regression. It defines the relationship between the two variables by fitting regression line to the data. one of the two With the best fit regression line to the data the error It is possible to reduce the rate of difference between expected and true values. Linear Regression is classified into two types. One of it is Simple Linear Regression, in which only one independent variable is used and the second type of regression is Multiple Linear Regression. Multiple independent variables are used in this sort of regression, which we are now using for the thesis.

V. CONCLUSION

The system is intended for the student. And the privileges are provided to admin to upload student results and have a good sight on the student result. The whole result analyzer will be under the control of the administrator and The admin has complete access to the result, including the ability to read, write, and execute it. And admin gives the privileges to the Teacher and student will get result viewing url as an alert of results. Present studies show that academic performances of the students are also dependent on student's background and other attributes. Many studies show that, in addition to past academic performance, a student's background and other characteristics have a substantial impact on their performance. Machine learning has been increasingly important in recent years across many industries, and it may also be applied efficiently in academia. In the future, many applications with improved ability and efficiency may become an integrated part of every academic institutions.

REFERENCES

- [1] Amandepp Kaur, Nitin Umesh, Barjinder Singh” Machine Learning approach to predict Student Academic Performance, International Journal for Research in Applied Science Engineering Technology (IJRASET), Volume.6 Issue IV, April 2018.
- [2] Pedro Strecht, Luis Cruz, Carlos Soares, João Mendes-Moreira and Rui Abreu “A comparative study of classification and regression algorithms for Modelling student’s Academic performance”, Proceedings of the 8th International Conference on Educational Data Mining,2015.
- [3] G. Sujatha, S. Sindhu and P. Savaridassan “Predicting student’s performance using personalized analytics”, Volume.119 No. 12, 2018.
- [4] Ankitha A Nichat, Dr. Anjali B Raut “predicting and Analysis of student Performance Using Decision Tree Technique”, International Journal of Innovative Research in Computer and Communication Engineering.Vol.5, Issue 4, April 2017.
- [5] Z. Ibrahim, D. Rusli, Predictingstudents academic performance: comparing artificial neural network, decision tree and linear regression, in: 21st Annual SAS Malaysia Forum, 5th September, 2007.
- [6] C. Romero, S. Ventura, Educational data mining: A review of the state of the art, Trans. Sys. Man Cyber Part
- [7] D. M. D. Angeline, Association rule generation for student performance analysis using apriori algorithm, The SIJ Transactions on Computer Science Engineering & its Applications (CSEA) 1 (1) (2013) p12–16.
- [8] M. W. Sukumar Letchuman, Mac Roper, Pragmatic cost estimation for web applications.
- [9] Syed Immamul Ansarullah, Pradeep Kumar Sharma, Abdul Wahid, Mudasir M Kirmani, “Heart Disease Prediction System using Data Mining Techniques: A study”, International Journal of Engineering Sciences & Research Technology, 1375-1381,(August 2016).
- [10] Carlos Ordonez, Edward Omiecinski, Levien de Braal, Cesar A. Santana, Norbert Ezquerra, Jose A. Taboada, David Cooke, Elizabeth Krawczynska, Ernest V. Garcia, “Mining Constrained Association Rules to Predict Heart Disease”, IEEE Explorer, 433-440 (2001).

