Student Performance Analysis

[1]Smt Sri Swathi, [2]Preshitha Puppala, [3]Divya Vasara, [4]Sowmya Chippa, [5]Prasanna Kumari Bhukya

[1]Professor, [2,3,4,5]UG Student
Department of CSE,
Gokaraju Rangaraju Institute of Engineering and Technology, India

Abstract: **Analyzing and predicting academic performance is essential for any educational institution. Predicting student performance can help teachers take action to create a strategy to improve performance early. With the development of machine learning supervised methods and supervised methods that develop these types of applications help teachers better analyze students compared to existing methods. In this case student mark prediction using back-to-back project efficiency is a hypothetical guess as previous students mark and predict marks in the next lesson and calculate model accuracy.**

Educational institutions are using new technologies to improve the quality of education but most of the applications used in colleges are related to services and development rather than web applications that help students to do online training and exams. There are many ways teachers can learn more about student performance. In view of this problem machine learning methods are used to predict students' marks based on past marks and to predict the outcome. Lower descent models are used to predict student performance and predict marks in the next lesson.

1. INTRODUCTION

Machine-based data mining methods are used to automate student performance prediction using the regression line method.
Design a machine learning model for predicting student marks and take steps to improve student performance.
Analyzing and predicting students' marks is done based on assumptions and personal mark data is not considered an academic test. Educational institutions use web applications to train students and evaluate performance based on marks but there are no specific steps to follow to predict student performance and take steps to improve efficiency.

The main purpose of higher education institutions is to provide quality education for their students. Another way to achieve a high level of quality in a higher education system is to obtain predictive information about student enrollment in a particular subject, to obtain unusual values in student results papers, to predict student performance and more.

2. METHODOLOGY

2.1 Linear Regression Algorithm

Predominant intention of our project is to become aware of the students who had been in at low performance. For this implementation we use linear regression algorithm. With the advancement of machine learning supervised and unsupervised techniques developing these kinds of applications are helping teachers to analyse students in better way compare to existing methods. In this student marks prediction using Linear regression project students' academic performance is prediction considering input as previous students marks and predict next subject marks and accuracy of the model is calculated.

Linear regression is one of the easiest and most famous machine getting to know algorithms. It is a statistical technique that is used for predictive analysis and shows the relationship between continuous variables. Linear regression makes predictions for non-stop/actual or numeric variables including sales, revenue, age, product fee, etc.

The linear regression model provides a sloped straight line representing the relationship between the variables. Consider the below image:
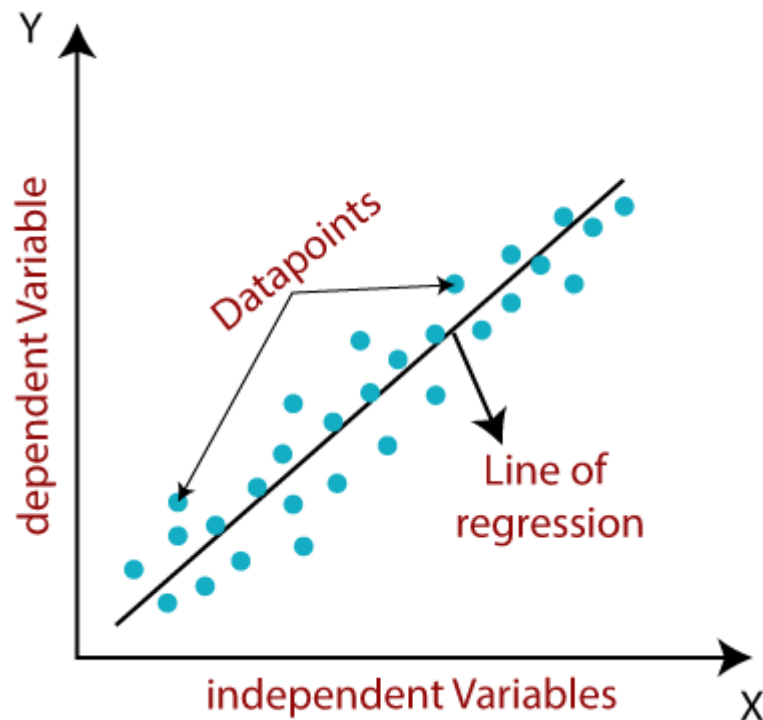
Mathematically, we can represent a linear regression as: $y = a_0 + a_1 x + \varepsilon$

Here,

Y = Dependent Variable(Target Variable)

X = Independent Variable(Predicted Varaiable)

a0 = Intercept of the line(Gives an additional degree of freedom)

a1 = Linear Regression Coefficient(Scale factor to each input value)

ε = Random Error

Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), consequently called linear regression. *If there is a single input variable (x), such linear regression is called **simple linear regression**. And if there is more than one input variable, such linear regression is called **multiple linear regression**.* The linear regression model gives a sloped straight line describing the relationship within the variables.

3. IMPLEMENTATION

3.1 Data Collection

Prepared our own datasets for the implementation of the model

3.2   Data Pre-processing

We have text data. That data contains lot of attributes and rows. Preprocessing the dataset is a major task. The text data has to be cleaned for good performance. This includes removing punctuations, converting upper case to lower case, removing numbers, removing words with length less than 1and removing special tokens. Firstly we filtered the null values present in data set. Preprocessing null values gives optimal predicted value while training and testing dataset. Then secondly changed attribute values to our understanding names. Then we removed the unnecessary data that are not required to our project. Then removed duplicate data(columns and rows).

3.3   Process/Workflow

The project makes use of linear regression algorithm to train the model using the dataset. Firstly, we divided dataset into two parts for training and testing purpose. The training dataset is used for training the model by linear regression algorithm. The testing dataset is used for testing the model after it is implemented completely to check whether it is predicting accurately or not. After that we implemented a web platform user friendly site. User can easily able to access the site and use the model that we implemented.

The website takes inputs as the previous marks of Database Management Systems, Operating Systems and Automation, Languages and Computation, Software Engineering, Managerial Economics and Accountancy. After entering inputs then user has to click submit button to predict student performance. The inputs are analyzed by the model that is built internally and calculates the performance.

Internally it calculates the values for student performance using the trained data. After performing the analysis in backend it returns the values that is predicted for student performance. During the testing the model and the pickle file (created by using predict () for each input) is loaded and predicted values is generated. The generated value is sent to user.


Flow chart
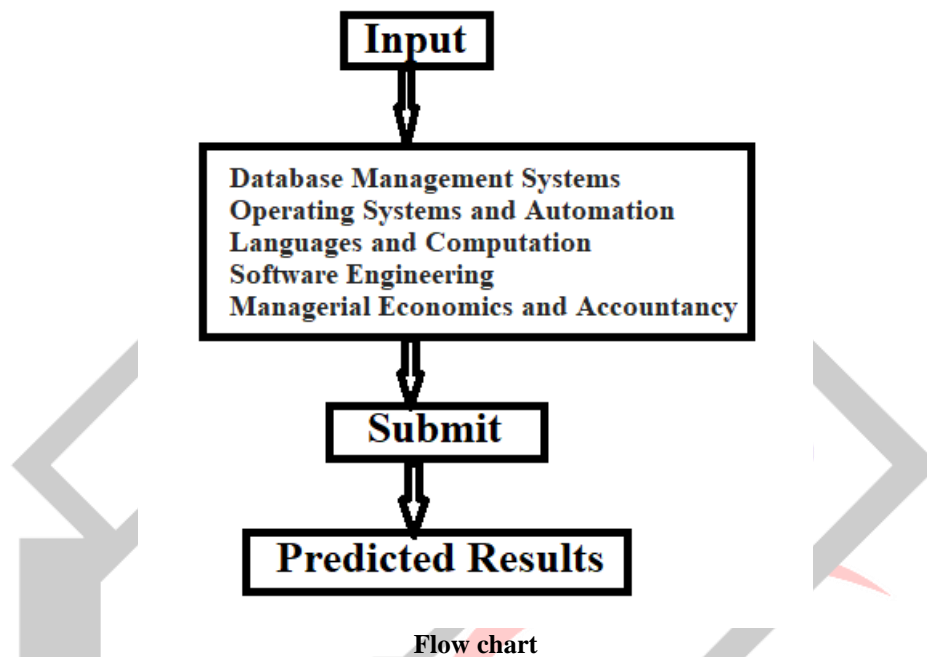
3.4   Evaluate the model
We used 5 datasets. Each consists of 637 rows and 2 columns. The evaluation of model is done by using the testing dataset. The model is get evaluated by entering inputs from the testing dataset on already trained model.

4. RESULTS
    We created a website platform in local host to make easy for users. The proposed model is implemented in Google Colab. The GPU provided by Colab helped a lot to train the model. Since there are more number of rows and columns in dataset, we trained the model by dividing the dataset to get more efficiency.
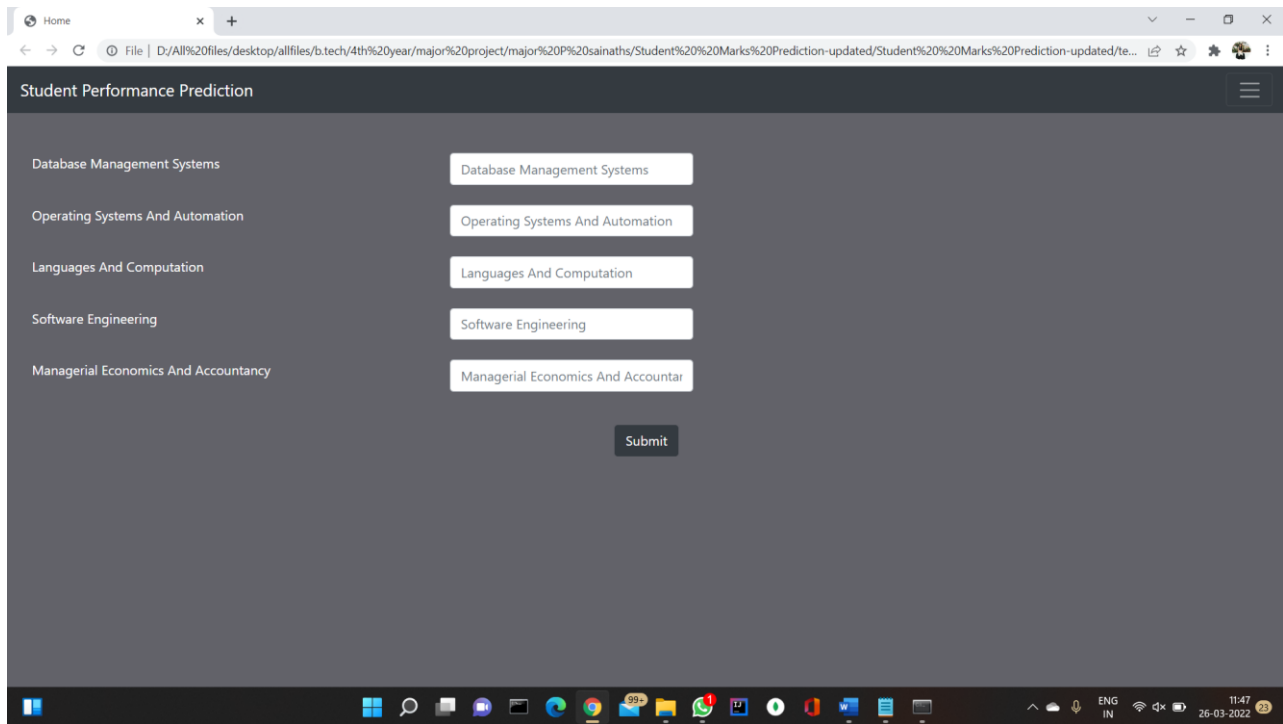
REFERENCES

[1] A. Sonali Agarwal, G.N.Pandey and M.D.Tiwari, "Data Mining in Education: Data Classification and Decision Tree Approach", International Journal of e-Education, e-Business, eManagementande-LearningVol.2, No.2, April2012.

[2] MounikaGoyal and RajanVokra, "Applications of Data MininginHigherEducation", International Journal on Computer Science Issues (IJCSI), ISSN (Online): 1694- 0814, Vol 9, Issue2, No 1, March 2012

[3] Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar, "Mining student data using decision trees", International Arab Conference on Information Technology (ACIT 2006), YarmoukUniversity, Jordan, 2006.

[4] KarishmaKhan, JanhaviPatil" Analysis of Intelligent System for Student's Performance using E-learning Approach", 2018.