



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 5, May 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

Student Mark Prediction Using Machine Learning

Achal Ramteke, Prof Vijay Rakhade, Prof Pushpa Tandekar

Department of Computer Science and Engineering, Shri Sai College of Engineering and Technology, Chandrapur, India

Department of Computer Science and Engineering, Shri Sai College of Engineering and Technology, Chandrapur, India

Department of Computer Science and Engineering, Shri Sai College of Engineering and Technology, Chandrapur, India

ABSTRACT: Education is very important issue regarding development of a country. The main objective of educational institution is to provide high quality education to its student. One way to accomplish this is by predicting student's academic performance and thereby taking early steps to improve student's academic performance and thereby taking early steps to improve student's performance and teaching quality. This system aims to predict student's mark using linear regression. The idea behind this analysis is to predict the mark of student by their studying hour. Analyzing and prediction of academic performance is important for any education institutions. Predicting student performance can help teachers to take steps in developing strategy for improving performance at early stages. With the advancement of machine learning supervised and unsupervised techniques developing these kinds of applications are helping teachers to analyze 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.

I. INTRODUCTION

In this project, the data collected from an students while they are studying for their exam that is how much hour a student can study based on which the student mark is predicted. This overview of the data collected helps in determining How much hour need to do the study to get 99% mark and if a student study $x()$ hours per day so how much mark he will get through these points and the school or college can determine the performance of the student. Data collected includes student study hours and student mark, which will help the prediction process. There are two columns in the dataset which includes numerical columns. Different machine learning methods are used to analyze the dataset and to predict the output. This is a time saving process for banks. After performing the evaluation with the three different machine learning models, the model with more accuracy is considered for further prediction

II. OBJECTIVES

Machine learning based data mining techniques are used to automate process of student performance prediction using linear regression technique. Education is very important issue regarding development of a country. The main objective of educational institutions is to provide high quality education to its students. One way to accomplish this is by predicting student's academic performance and thereby taking early steps to improve student's performance and teaching quality. Students marks of other subjects are taken as input for evaluation students' performance. Data set is pre-processed and features and labels are extracted from dataset then dataset is split in to test and train sets then linear regression is applied to dataset for prediction.

III. LIBRARIES USED

Pandas ,Numpy ,Matplotlib.pyplot and sklearn are the major libraries used in the project. Pandas is mainly used for data analysis. Pandas allows importing data from various file formats such as comma-separated values, JSON, SQL, Microsoft Excel. Pandas allows various data manipulation operations such as merging, reshaping, selecting, as well as data cleaning, and data wrangling features. Numpy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. Moreover Numpy forms the foundation of the Machine Learning stack. Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. Scikit-learn is a free machine learning library for Python. It features various algorithms like support vector machine, random forests, and k-neighbours, and it also supports Python numerical and scientific libraries like NumPy and SciPy.

Models Used:

Splitting The Data Into Training And Testing

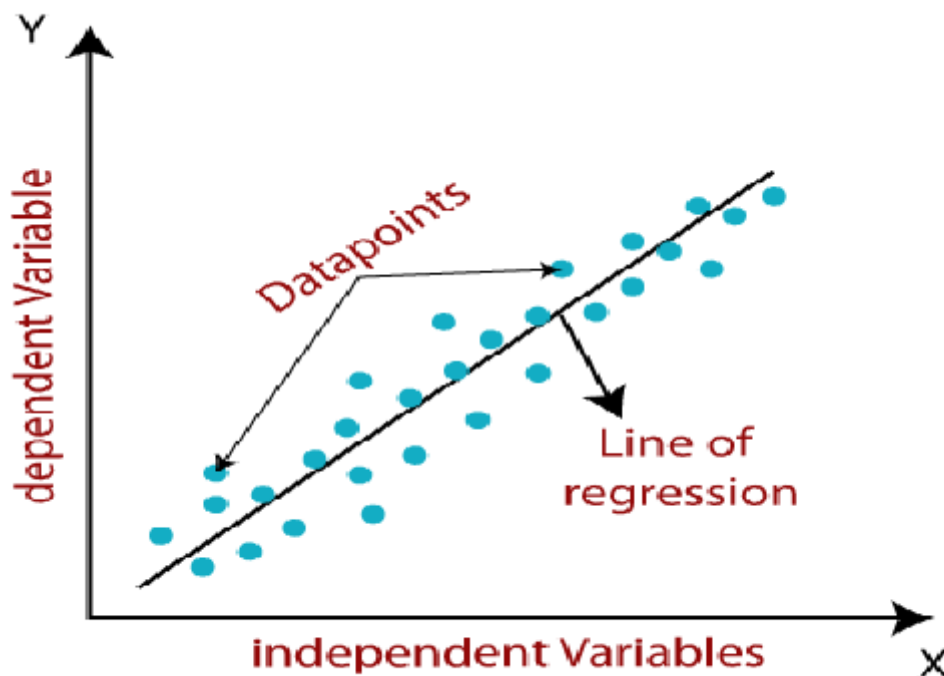
Train/Test is a method to measure the accuracy of your model. It is called Train/Test because you split the the data set into two sets: a training set and a testing set. ... You test the model using the testing set. Train the model means create the model. By using similar data for training and testing, you can minimize the effects of data discrepancies and better understand the characteristics of the model. After a model has been processed by using the training set, you test the model by making predictions against the test set.

We have a class called `train_test_split` which we import sklearn model selection which is used for testing our machine learning module. In this class we have to give two variable called the independent variable and dependent variable. In our project the student hour is the independent variable and student mark is the dependent variable which are X,y respectively. Here we are giving test size equal to 0.2 which means we are giving 20 % of our data for testing. The `train_test_split` gives us 4 dataframe called `x_train`, `x_test`, `y_train` and `y_test`

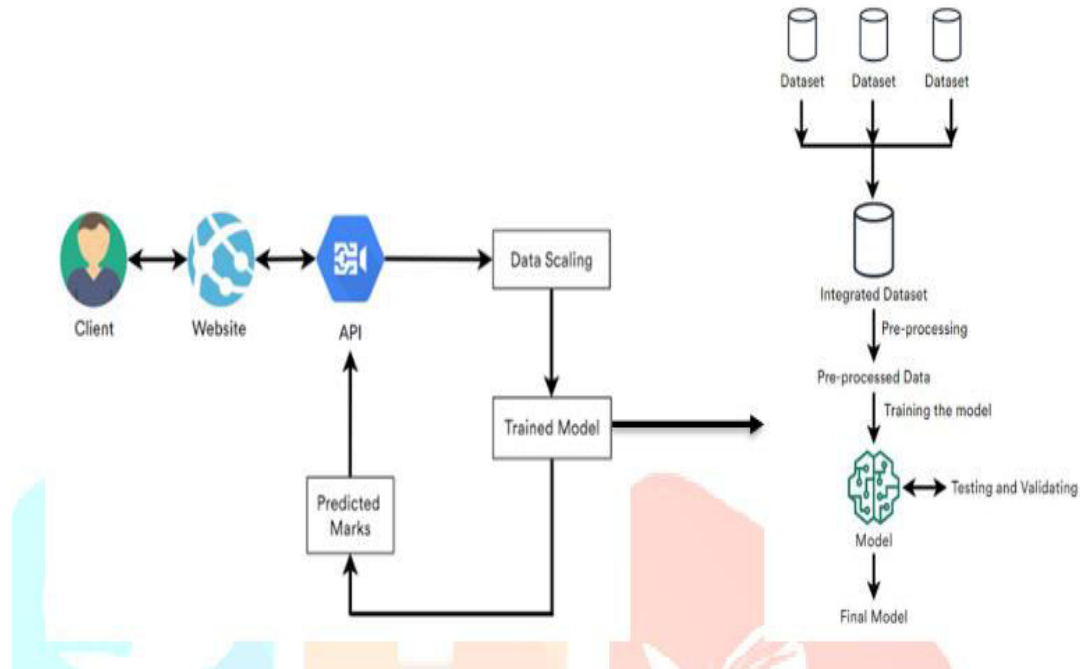
Selecting a model and training it(LINEAR REGRESSION)

LINEAR REGRESSION is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (x) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable. The linear regression model provides a sloped straight line representing the relationship between the variables



IV. METHODOLOGY



Initially the client/student visits the web application where the home page of the website is displayed which contains navigation buttons to various other pages that includes ‘about us’, ‘contact us’, ‘student login’ and ‘admin login’. If the client is a teacher, then they can navigate to the admin login page and enter the admin credentials.

Once the login button is clicked it takes the client to the admin dashboard wherein the client i.e., the teacher can enter student details into the database and can also edit the details of the students whose data is already registered in the database. If the client who visited the website is a student, then they can navigate to the student login page and enter the credentials given to that particular student. After the student is login with the credentials, a page containing the student’s details (USN of the student, percentage scored by the student in all the seven semesters) is displayed. When the student clicks on the predict button present in the page, the particular student’s form details are sent to the trained machine learning model trough an API. Trained machine learning model is created by first collecting the data from K S School of engineering and management.

All the small data is combined into a large dataset called the integrated dataset. This dataset is pre-processed by removing the errors and noises which can make the model less effective. Also, the missing data is handled by computing mean method and the outliers are handled by using inter quartile method based on the distribution of the columns. This pre-processed dataset is divided in a ratio of 80:20 where 80 percentage of the dataset is separated for training and 20 percent of the dataset is assigned for testing the model. The model is implemented support vector machine, random forest, decision tree, and K-Nearest Neighbor (KNN) algorithms. The model is trained using the training dataset. The model that is trained on various algorithms is tested using the testing dataset. The accuracy obtained from different machine learning algorithms is noted and the best algorithm amongst them which gives high accuracy and less error rate is selected as a final model. The final machine learning model takes the details of the student from the form as an input and predicts the grade that particular student is likely to score in the eighth semester. The grade predicted by the machine learning model is sent back to API and in return these predicted grades is sent to the website. Finally, these predicted grades are displayed on the website for the student.

Data Analysis Structure Of Dataset

There are two different columns including the target variable. These include Numerical data

Numerical_columns = [` study_hours ` , ` student_marks `]

Loading the dataset

In the second step we are loading the dataset. I have taken a csv file and we are loading it with the help of pandas which helps us to read the csv files, then I have given the path path of the csv file and assigned it a variable called df.

Data Visualisation

In the third step we want to visualize our data. We use seaborn and matplotlib.pyplot libraries to visualize and thereby analyze the dataset. Here are the visualizations of the data in the dataset.

Prepare the data for Machine Learning algorithms

First step in preparing the data for machine learning is data cleaning .In our dataset there can be some point where we can find null values. When we are training our machine learning model we always have to give clean data.

Splitting the data

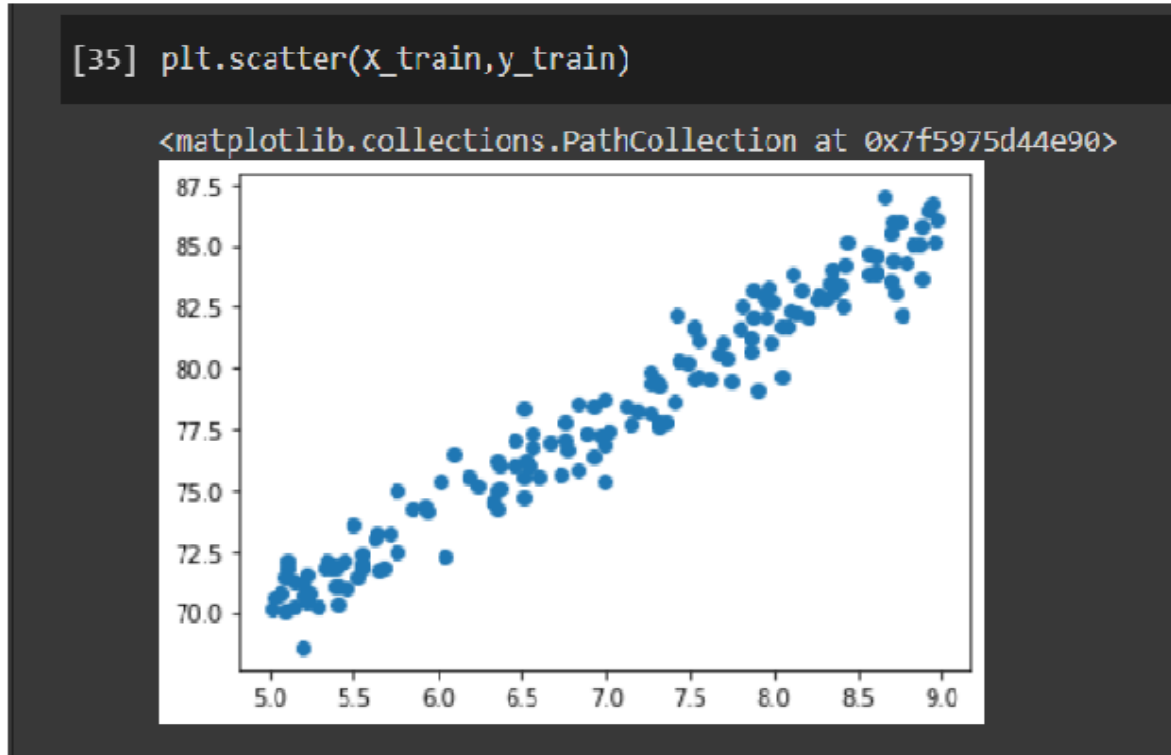
Here we are taking to variables X and y. We know that X variable is for the student hour and y is for the student mark so in the above diagram we are dropping student mark from from the X variable and giving axis is equal to column so X equal to student hour similarly for the y variable will be equal to the student mark. Finally after executing we get the shape of X and y

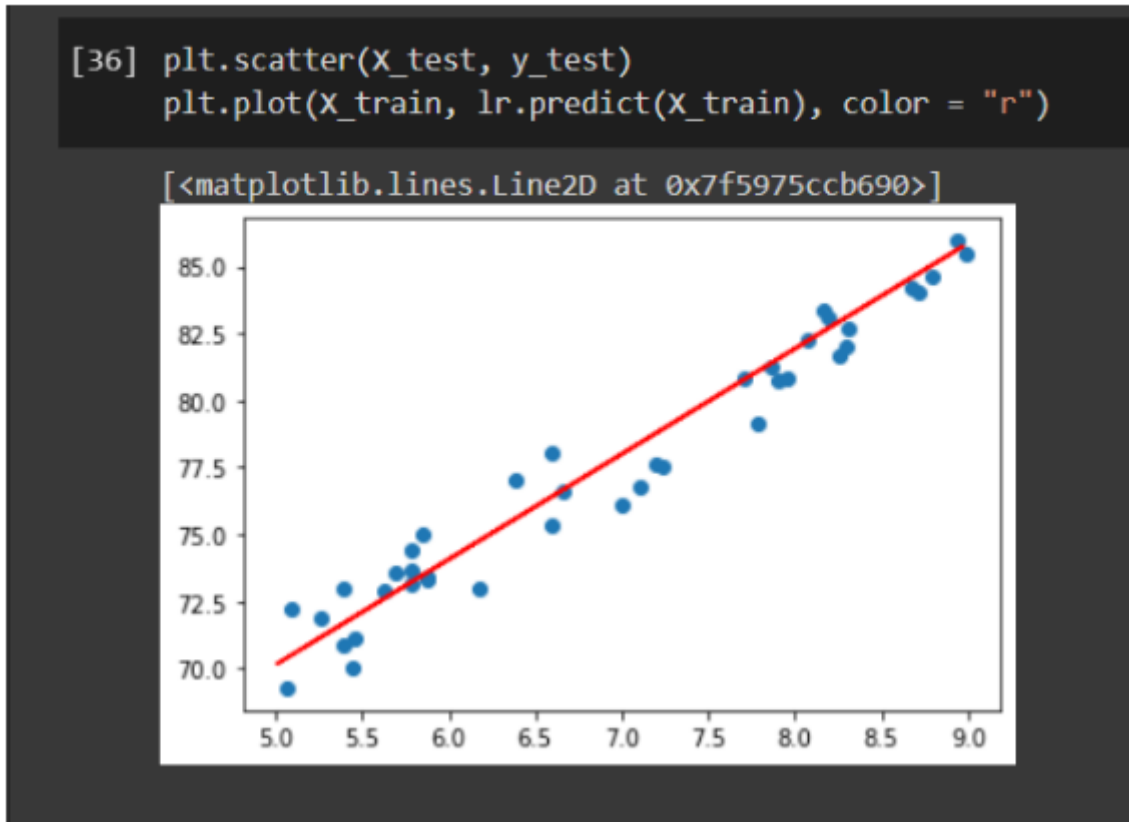
Test Results

Accuracy check

We are using a method called lr.score for which we will check for x_test and y_test. Based on this accuracy how our machine learning model predicted the line we have to use the below

Command





The dots in the the graph on the red line and our model has predicted 95 % accuracy

Saving our model

For saving our model we use a library called joblib, with the help of dump method we can save the linear regression model and also give a name for the model as student mark predictor.pkl Now we have to import the model for which we use joblib.load() and assign it to a variable and then predict the mode

```
[40] import joblib
      joblib.dump(lr, "student_mark_predictor.pkl")

      ['student_mark_predictor.pkl']

[41] model = joblib.load("student_mark_predictor.pkl")
```

Final result student mark prediction

If studying hour =10

```

▶ model.predict([[10]])[0][0]
89.80453520342567

```

If studying hour =2

```

▶ model.predict([[2]])[0][0]
58.31879107023909

```

V. CONCLUSION AND FURTHER SCOPE

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used. In this analysis, we did extensive analysis of input data and were able to achieve Test Accuracy of 95 % . From a proper analysis of the positive points and obstacles in the sections, it can be safely concluded that the linear regression is the mostly efficient. This prediction method works well and integrates with all type of student mark prediction process. Before final marks of all subjects are evaluated prediction can be performed. This section can be easily connected to many other programs. Soon this prediction module can soon be integrated with the automatic processing program module.

REFERENCES

- [1] Kaur, P. and Singh, W., 2016, August. Implementation of student SGPA Prediction System (SSPS) using optimal selection of classification algorithm. In 2016 International Conference on Inventive Computation Technologies (ICICT) (Vol. 2, pp. 1-8). IEEE.
- [2] Hasan, H.R., Rabby, A.S.A., Islam, M.T. and Hossain, S.A., 2019, July. Machine learning algorithm for student's performance prediction. In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.
- [3] Zohair, A. and Mahmoud, L., 2019. Prediction of Student's performance by modelling small dataset size. International Journal of Educational Technology in Higher Education, 16(1), pp.1-18.
- [4] Obsie, E.Y. and Adem, S.A., 2018. Prediction of student academic performance using neural network, linear regression and support vector regression: a case study. International Journal of Computer Applications, 180(40), pp.39-47.

- [5] Xu, J., Moon, K.H. and Van Der Schaar, M., 2017. A machine learning approach for tracking and predicting student performance in degree programs. *IEEE Journal of Selected Topics in Signal Processing*, 11(5), pp.742-753.
- [6] Bujang, S.D.A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H. and Ghani, N.A.M., 2021. Multiclass prediction model for student grade prediction using machine learning. *IEEE Access*, 9, pp.95608-95621.
- [7] Gull, H., Saqib, M., Iqbal, S.Z. and Saeed, S., 2020, November. Improving learning experience of students by early prediction of student performance using machine learning. In *2020 IEEE International Conference for Innovation in Technology (INOCON)* (pp. 1-4). IEEE.
- [8] Shah, M.B., Kaistha, M. and Gupta, Y., 2019, November. Student performance assessment and prediction system using machine learning. In *2019 4th International Conference on Information Systems and Computer Networks (ISCON)* (pp. 386-390). IEEE.
- [9] Turabieh, H., 2019, October. Hybrid machine learning classifiers to predict student performance. In *2019 2nd international conference on new trends in computing sciences (ICTCS)* (pp. 1-6). IEEE.
- [10]. Lowlesh Yadav and Asha Ambhaikar, "IOHT based Tele-Healthcare Support System for Feasibility and performance analysis," *Journal of Electrical Systems*, vol. 20, no. 3s, pp. 844–850, Apr. 2024, doi: 10.52783/jes.1382.
- [11] L. Yadav and A. Ambhaikar, "Feasibility and Deployment Challenges of Data Analysis in Tele-Healthcare System," *2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI)*, Raipur, India, 2023, pp. 1-5, doi: 10.1109/ICAIIHI57871.2023.10489389.
- [12] L. Yadav and A. Ambhaikar, "Approach Towards Development of Portable Multi-Model Tele-Healthcare System," *2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI)*, Raipur, India, 2023, pp. 1-6, doi: 10.1109/ICAIIHI57871.2023.10489468.
- [13] Lowlesh Yadav and Asha Ambhaikar, Exploring Portable Multi-Modal Telehealth Solutions: A Development Approach. *International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC)*, vol. 11, no. 10, pp. 873–879, Mar. 2024.11(10), 873–879, DOI: 10.13140/RG.2.2.15400.99846.
- [14] Lowlesh Yadav, Predictive Acknowledgement using TRE System to reduce cost and Bandwidth, March 2019. *International Journal of Research in Electronics and Computer Engineering (IJRECE)*, VOL. 7 ISSUE 1 (JANUARY-MARCH 2019) ISSN: 2393-9028 (PRINT) | ISSN: 2348-2281 (ONLINE).



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details