

Student semester marks prediction using linear regression algorithms in machine learning

Abirami T * and R. Vadivel

Department of Information Technology, Bharathiar University, Coimbatore - 641046.

World Journal of Advanced Research and Reviews, 2023, 18(01), 469–475

Publication history: Received on 27 February 2023; revised on 09 April 2023; accepted on 11 April 2023

Article DOI: <https://doi.org/10.30574/wjarr.2023.18.1.0591>

Abstract

Student Mark prediction using the previous semester's internal marks and semester marks project using a multivariate linear regression algorithm. Using the real-time data sets gathered from department, and analyze the data. The project's goal is to forecast the grade for the upcoming semester so that, by taking the necessary steps, for forecast the grade and enhance the outcomes in the upcoming semester. Designer developed three machine learning models for this project using the multivariate linear regression algorithm. Among these 3 models, two models are used to predict semester marks and another model is to predict the final internal or model exam marks. With the treatment of the data points and its specifies, to increase the machine learning model's effectiveness in this project.

Keywords: Artificial Intelligence; Machine Learning; Linear Regression Algorithm; Multiple Marks; Prediction

1. Introduction

Creator are using a multivariate linear regression Machine learning algorithm to analyze and forecast the upcoming semester mark using an existing real-time data set of a particular set of students collected from a department. To determine which model provides the most accurate model prediction, designer developed two separate models to forecast the results of the semester using data points with various details. Also, included a model to this project research that would help students anticipate their final internal exam or model exam marks, allowing to concentrate more on the future semester. This will highly help in forecasting the result and taking the required actions.

2. Literature Survey

It is viable to extract hidden facts and relationships among records the use of plenty of supervised and unsupervised gadget getting to know algorithms, for then to help decision-makers with inside the destiny in making the proper interventions. The huge variety of strong algorithms which can be carried out in all spheres of each day life, will includes tutorial system. This painting introduces students' educational overall performance prediction version that makes use of supervised sort of gadget getting to know techniques like assist vector gadget and logistic regression. The sequential minimum optimization approach surpasses logistic regression with the aid of using reaching more accuracy, consistent with records supported with the aid of using numerous trials making use of various technologies. Also, the facts gleaned from this observe can help instructional establishments in forecasting students' behavior with inside the destiny. (E. S. Bhutto, I. F. Siddiqui, Q. A. Arain and M. Anwar,2020).

By this assignment, college students will actually be capable of estimate their very last results. The students may be capable of put together themselves for a tremendous final results as a result. Students occupy their time via way of means of taking element in a variety of activities. Because of the experiment, if a scholar spends an excessive amount of

* Corresponding author: Abirami T

time on any of these activities, their grades won't, college students can create a studies habitual or tenet that they could comply with for all in their different assignments. The movements of college students may even be expecting their results. Because of the experiment, if a scholar spends an excessive amount of time doing any of these tasks, Nonetheless, college students can create a studies project (M. Saifuzzaman, M. Parvin, I.Jahan, N. N. Moon, F. NarinNur and S. F. Shetu, 2021).

It can help instructors in figuring out college students require extra assist and stopping college students from quitting earlier than very last exams, the student's overall performance prediction is an crucial studies topic. A logistic regression turned into one of the gadget mastering methods. So, from the subsequent session, it's miles predicted that teachers can also additionally record progressed college students overall performance (Hussain, M.,Zhu,W., Zhang, W. et al, 2019).

Education establishments are seeking for to forecast pupil success that's an essential observe area. Predicting pupil overall performance can allow instructors to save college students from losing out earlier than very last tests, discover those want similarly assist and boom organization status and prestige. The most important fundamental elements can constitute the schooling dataset for supervised device studying algorithms. Evaluated the effectiveness of Logistic Regression and different supervised device studying techniques. Trained a version with the aid of using the datasets furnished with the aid of the college. Findings indicated that logistic regression classify the maximum correct in predicting the perfect very last grades of college students (68.7% for handed and 88.8% for failed) (Ali Salah Hashim, Wid Akeel Awadh and Alaa Khalaf Hamoud,2020).

3. Methodology

3.1. Linear Regression

Linear regression is a supervised learning algorithm. In this have single – variate and multi – variate algorithms.

3.2. Single Variate Linear Regression

Single variate linear regression is a statistical technique that is commonly used to model the relationship between a dependent variable (Y) and a single independent variable (X). The method assumes that the relationship between Y and X is linear and can be expressed as:

$$Y \approx \beta_0 + \beta_1 X \quad \text{---(1)}$$

Equ(1) Where Y is the dependent variable, X is the independent variable, β_0 is the intercept term, and β_1 is the slope coefficient that measures the change in Y for a unit change in X. This technique is useful for predicting a response based on a single predictor variable.

3.3. Multi Variate Linear Regression

Linear regression model takes the form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon \quad \text{---(2)}$$

Equ(2) Where, $\beta_0, \beta_1, \beta_2, \beta_p$ are the coefficients of the input variables. X_1, X_2, X_3, X_p . ϵ - is the intercept of the line. This linear regression is similar to the

$$y = mX + c \quad \text{---(3)}$$

Equ(3) Straight line equation.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \quad \text{---(4)}$$

Equ(4) Consider a standard bi - variate linear regression model.

3.4. Model 1

In the first model, users choosing factors such as test1, test2 and test3 as inputs and semester mark as the output. So the equation will be,

$$y=m_1X_1+m_2X_2+m_3X_3+b \quad \text{---(5)}$$

Where, Equ(5)

Y → Predicted marks

b → intercept of y

m₁, m₂, m₃ → coefficients of input factors X₁, X₂ and X₃

$$\text{Predicted mark} = m_1 \cdot \text{internal}_1 + m_2 \cdot \text{internal}_2 + m_3 \cdot \text{internal}_3 / \text{model} + b \quad \text{---(6)}$$

Equ(6) Now the algorithm is ready and user need to train the machine learning using the collected data set. For this need to use Pandas library and scikit learn library.

Let's create inputs X and output y as variables for training. Now data set is ready for the machine learning training.

Now let's predict with some sample data and find out the model score.

In the image, user can see for the shape of 330 datapoints with 4 columns the model score got is 0.6312. Now let's try to improve the efficiency of the model by increasing the data points.

3.5. Model2

In this model, user choosing factors such as test1, test2, test3, seminars, assignments and other marks. So equation will be,

$$Y = m_1X_1 + m_2X_2 + m_4X_4 + m_5X_5 + m_6X_6 + b \quad \text{---(7)}$$

Where, Equ(7)

Y → Predicted budget b → intercept of y

m₁, m₂, m₃, m₄, m₅, m₆ → coefficients of input factors X₁, X₂, X₃, X₄, X₅ and X₆ So the equation will be referred as

$$\text{Predicted mark} = m_1 \cdot \text{internal}_1 + m_2 \cdot \text{internal}_2 + m_3 \cdot \text{internal}_3 /$$

$$\text{Model} + m_4 \cdot \text{Assignments} + m_5 \cdot \text{Seminar} + m_6 \cdot \text{Others} + b \quad \text{---(8)}$$

Now the algorithm is ready and builders need to train the machine learning using the collected dataset.

Let's create inputs X and output y as variables for training. Designer dataset is ready for the machine learning training.

Let's see the score of the model.

In this model - model 2, can see the model score is improved to 0.9957 which is very efficient than model 1. Here creator have used 330 data points with 7 columns.

Table 1 Comparison between model 1 and model 2

	model 1	model 2
Data Points	330	330
columns	4	7
score	0.631246775	0.995737515

4. Project Flow Graph

Fig.1.The project flow graph explains the colab model file for all marks vs internal1, internal2, model are used to predict the semester marks.



Figure 1 The increasing of model marks

5. Workflow

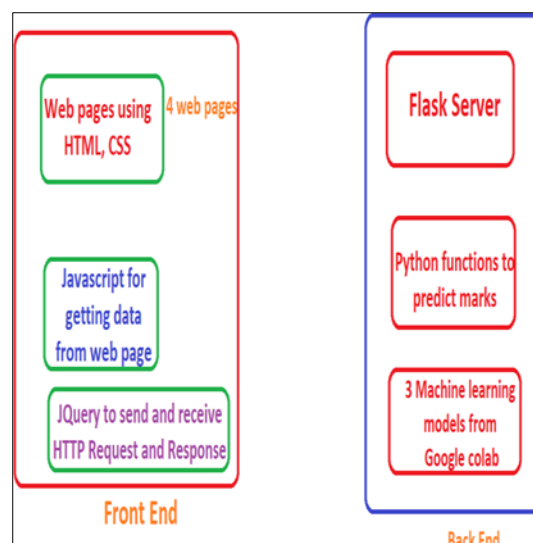


Figure 2 The work flow of machine learning front end back end training

Fig.2.The project workflow train the machine using web pages of HTML, CSS. Javascript is used to getting the data from web page. And then JQuery to send and receive the HTTP Request and response. Train the Flask Server, using python function to predict the marks, Create the three model colab file for training the machine learning algorithm.

6. Implementation

Fig.3.Describes the details of colab file have use x and y-axis for represents the dataset, Sklearn to predict the linear_model, Training the machine using linear regression algorithm to predict the model file.

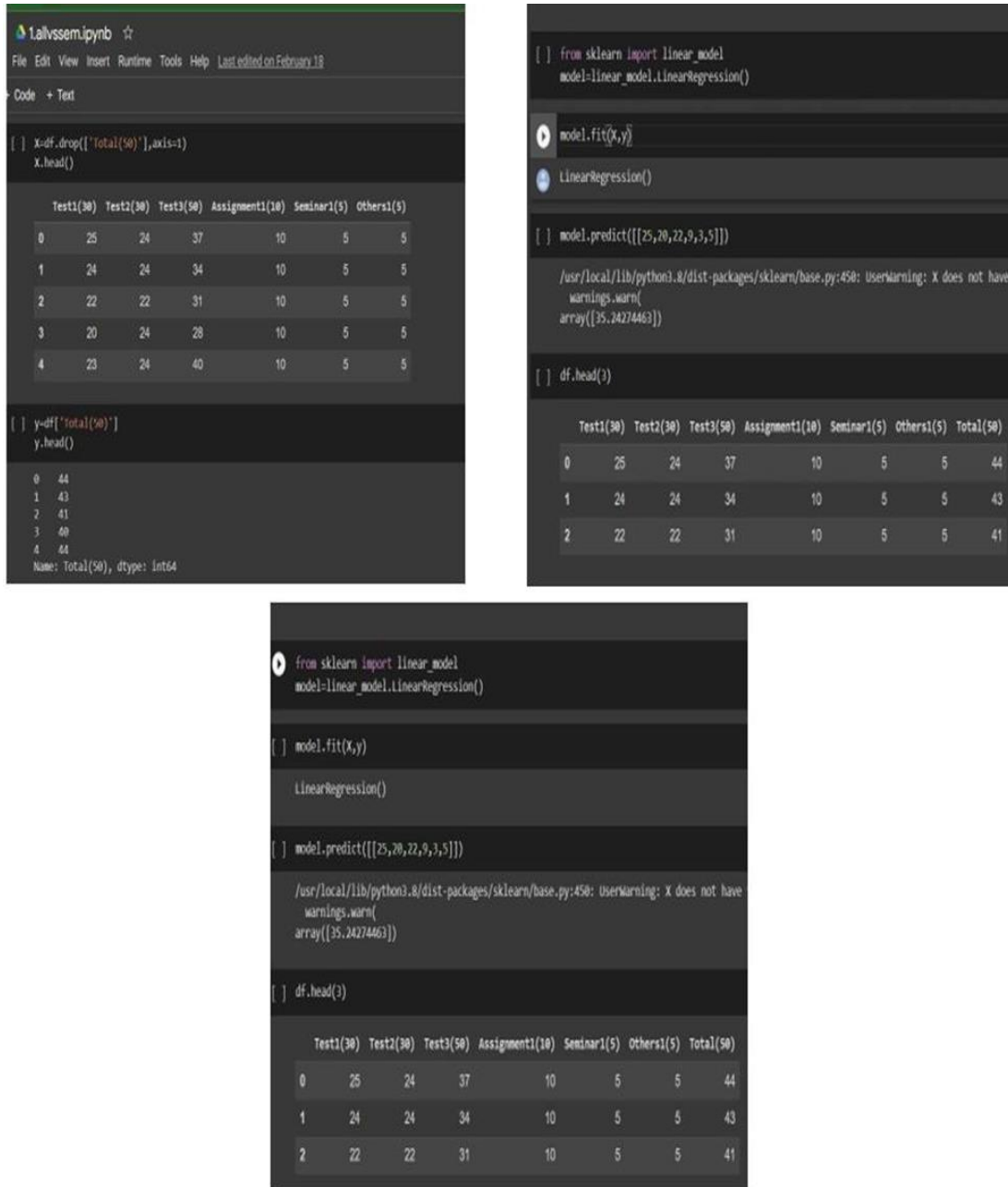


Figure 3 The figure represents the colab file

7. Execution

Fig.4. The below image represents the front end(html,css)design. Use the facts to predict the input. The first image is a representation of the internal model marks. Then overall scores are presented. The semester grades should be mentioned in the final graphic.

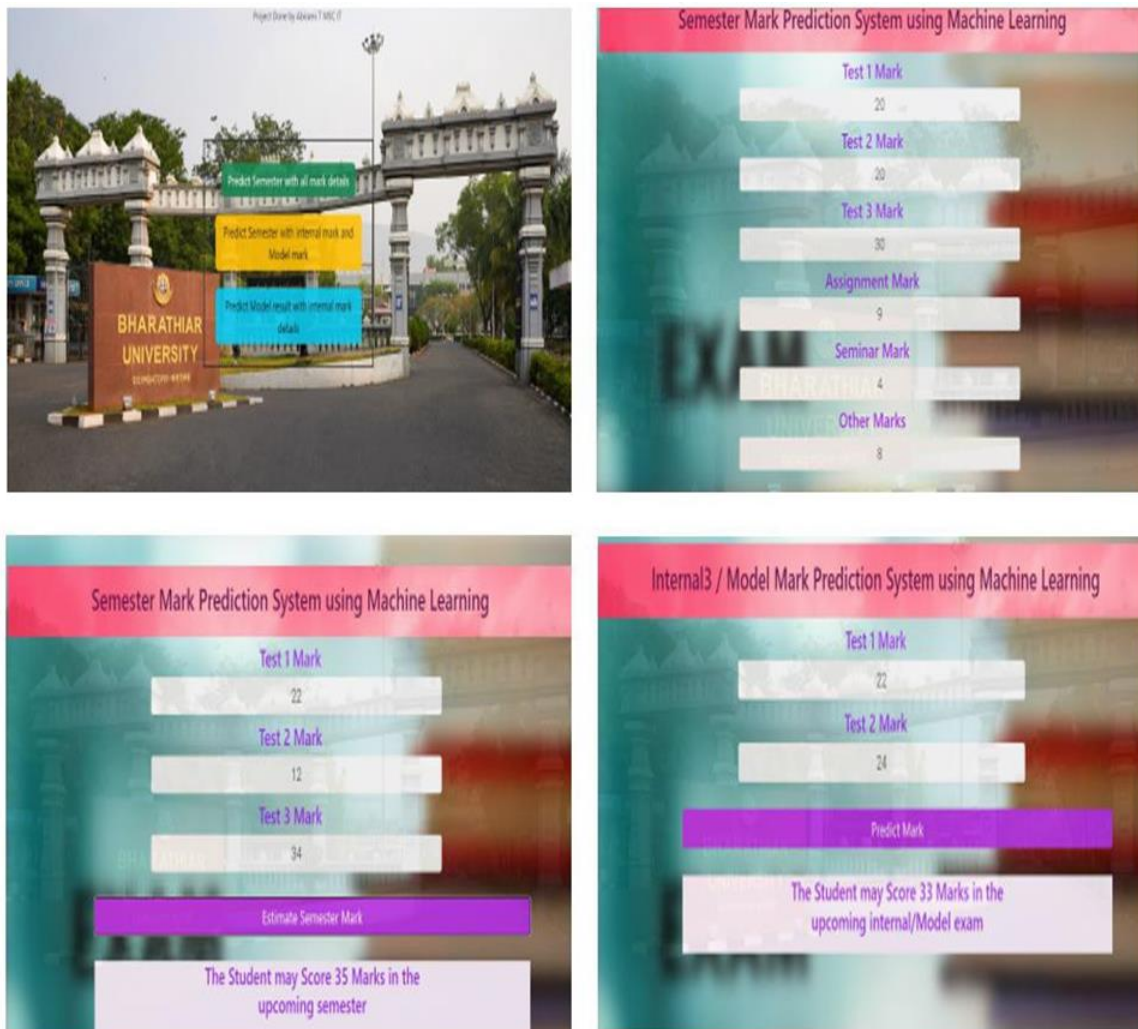


Figure 4 The output screenshot

8. Conclusion

The reason designer give more details about data points the efficiency of prediction is getting improved. Creator can shows the three types of output screen there are Semester Mark Prediction Machine Learning, Semester Mark Prediction System Using Machine Learning and Internal3/Model Mark Prediction System Using Machine Learning.

So in this scenario Model -1can train the dataset of test1, test2, test3 (model)is used to predict the Semester marks. Model – 2 can train the dataset of test1, test2, test3 (model), Assignment, Seminar, Others is used to predict the Semester marks. Model – 3 can train the dataset of test1, test2 are used to predict the Semester marks. Output was displayed in the html web page.

Compliance with ethical standards



Disclosure of conflict of interest

No conflict of interest.

References

- [1] H. M.R. Hasan, A. S.A. Rabby, M.T. Islamand S. A. Hossain, "Machine Learning Algorithm for Student's Performance Prediction,"2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-7,doi:10.1109/ICCCNT45670.2019.8944629
- [2] B. Sravani and M. M. Bala, "Prediction of Student Performance Using Linear Regression,"2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154067.
- [3] K.AlMayahi and M.Al-Bahri, "Machine Learning Based Predicting Student Academic Success," 2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), Brno, Czech Republic, 2020, pp. 264-268, doi:10.1109/ICUMT51630.2020.9222435.
- [4] Prediction of Students Performance using Machine learning J. Dhilipanet al2021 IOP Conf. Ser.: Mater. Sci. Eng. 1055012122,doi:10.1088/1757899X/1055/1/012122.
- [5] Tran, T., Dang, H., Dinh, V., Truong, T., Vuong, T. & Phan, X.(2017).Performance Prediction for Students: A Multi-Strategy Approach.Cybernetics and Information Technologies,17(2)164182. doi:10.1515/cait-2017-0024.
- [6] R. Alshabandar, A. Hussain, R. Keight and W. Khan, "Students Performance Prediction Using Machine Learning Algorithms," 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020, pp. 1-7, doi: 10.1109/IJCNN48605.2020.9207196.
- [7] E. S. Bhutto, I. F. Siddiqui, Q. A. Arain and M. Anwar, "Predicting Students' Academic Performance Through Supervised Machine Learning,"2020 International Conference on Information Science and Communication Technology (ICISCT), Karachi, Pakistan, 2020, pp.1-6, doi: 10.1109/ICISCT49550.2020.9080033.
- [8] M. Saifuzzaman, M. Parvin, I. Jahan, N. N. Moon, F. NarinNur and S. F. Shetu, "Machine Learning Approach to Predict SGPA and CGPA,"2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST), Yogyakarta, Indonesia,2021, pp. 211-216, doi: 10.1109/ICAICST53116.2021.9497847.
- [9] Hussain, M., Zhu, W., Zhang, W.etal.Using machine learning to predict student difficulties from learning session data. Art if Intell Rev52,381–407 (2019). <https://doi.org/10.1007/s10462-018-9620-8>.
- [10] Student Performance Prediction Model based on Supervised Machine Learning Algorithms Ali Salah Hashim,WidAkeel Awadh and Alaa Khalaf Hamoud, Volume 928,2nd International Scientific Conference of Al-Ayen University (ISCAU-2020) 15-16 July 2020, Thi-Qar,Iraq Ali Salah Hashim et al 2020 IOP Conf. Ser.: Mater. Sci. Eng. 928 032019 doi:10.1088/1757-899X/928/3/032019.

Author's short biography

	<p>T. Abirami received Bachelors Degree in Information technology in the year 2021 from Nirmala College Women Coimbatore, Tamil Nadu, affiliated to Bharathiar University. She is currently pursuing a Masters Degree in Information Technology from 2021 to 2023, at Bharathiar University, Coimbatore, Tamil Nadu. Her area of interest is Python in Machine Learning</p>
	<p>Dr.R.Vadivel is an Associate Professor, in the Department of Information Technology, Bharathiar University, Tamil Nadu, India. He received his Ph.D degree in Computer Science from Manonmaniam Sundaranar University in the year 2013. He obtained his Diploma in Electronics and Communication Engineering from State Board of Technical Education in the year 1999, B.E., Degree in Computer Science and Engineering from Periyar University in the year 2002, M.E., degree in Computer Science and Engineering from Annamalai University in the year 2007. He had published over 96 journals papers and over 40 conferences papers both at National and International level. His areas of interest include Information Security, Data mining,Digital Signal Processing</p>