**Department of Computer Engineering**

**Academic Year:** 2022-2023 **Semester:** VIII

**Subject:** Applied Data Science **Class / Division:** BE/CMPN/B

**Name:** Domingos J Pinto **Roll Number:** 007

**Experiment No.: 5**

**Implement Linear Regression and explore the performance evaluation metrics.**

**Aim :** Implement Linear Regression and explore the performance evaluation metrics.

**I OBJECTIVE**

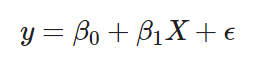
* To understand basic concepts of Linear regression
* To explore the performance evaluation metrics for Linear regression

**II THEORY**

**Linear Regression**

Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

The formula for a simple linear regression is:



y is the predicted value of the dependent variable (y) for any given value of the independent variable (x).

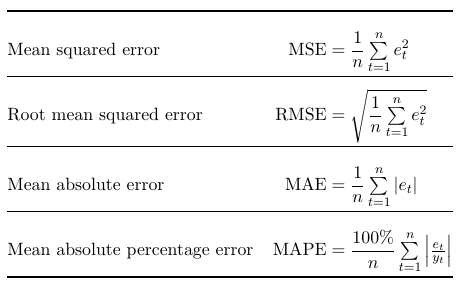
B0 is the intercept, the predicted value of y when the x is 0.

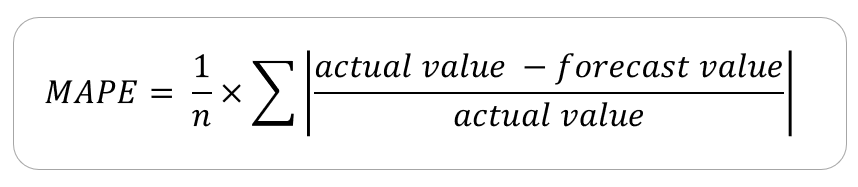
B1 is the regression coefficient – how much we expect y to change as x increases.

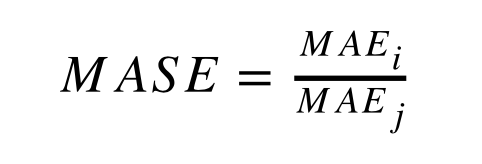
x is the independent variable ( the variable we expect is influencing y).

e is the error of the estimate, or how much variation there is in our estimate of the regression coefficient.

The metrics used to evaluate the performance of a linear regression model include:

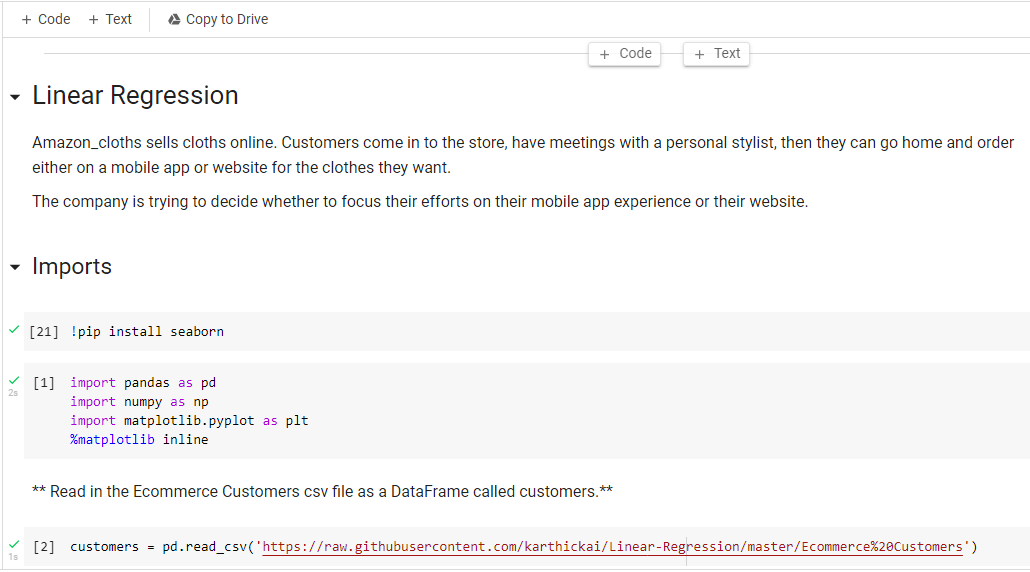


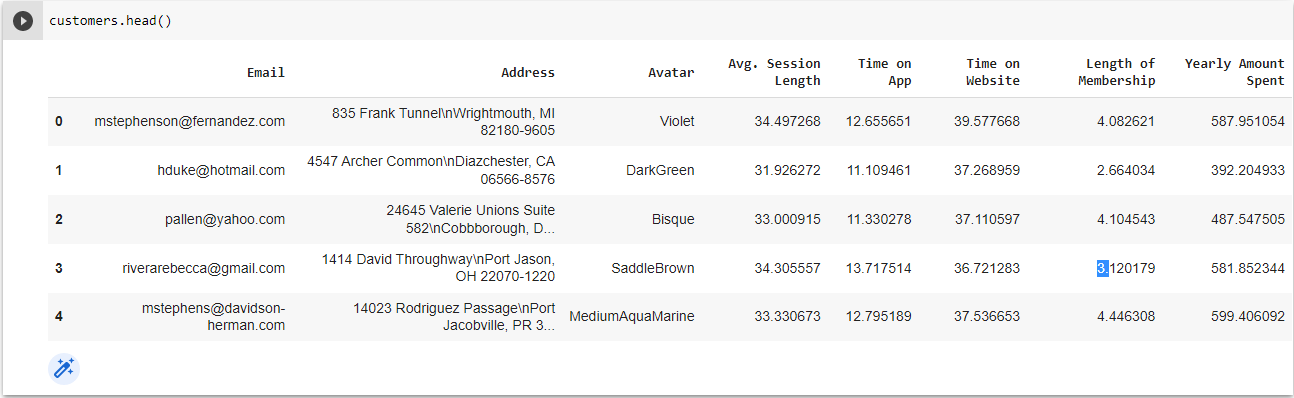


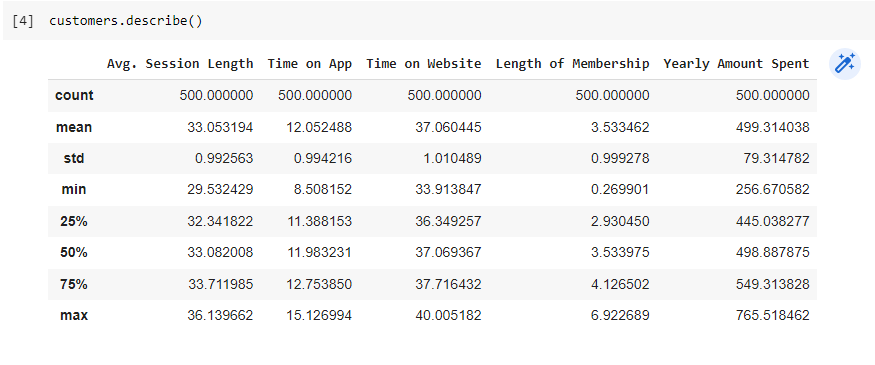


1. Mean Absolute Error (MAE): This measures the average absolute difference between the actual and predicted values.
2. Mean Squared Error (MSE): This measures the average of the squared differences between the actual and predicted values.
3. Root Mean Squared Error (RMSE): This is the square root of the MSE.
4. R-squared (R²): This is a statistical measure that represents the proportion of the variance in the dependent variable that is predictable from the independent variable(s).
5. Mean Absolute Percentage Error (MAPE): This measures the average percentage error between the actual and predicted values.
6. Mean Absolute Scaled Error (MASE): This measures the average error between the actual and predicted values relative to the mean absolute error of a naïve forecast.

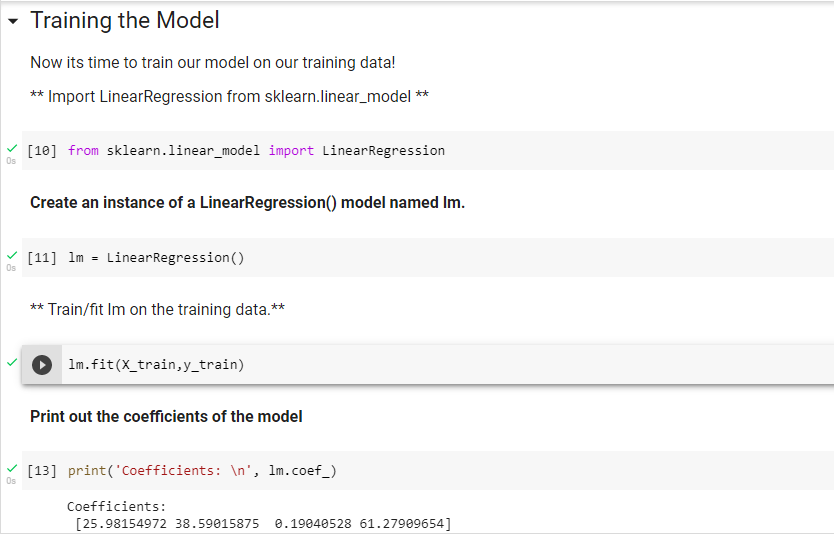
**III IMPLEMENTATION**

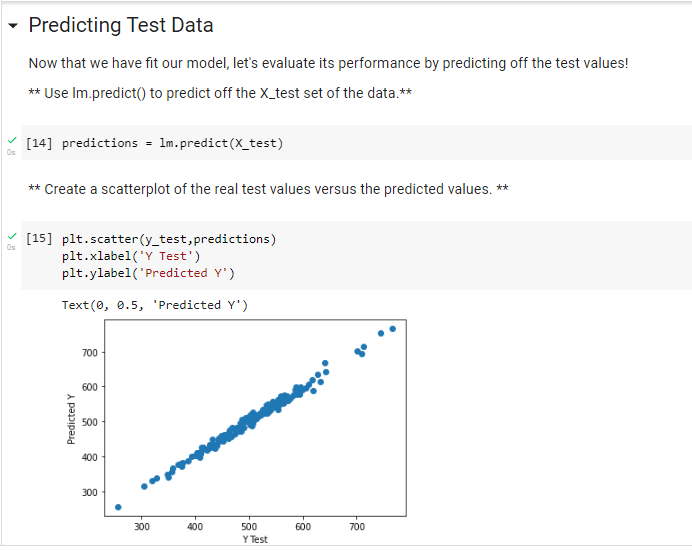
****

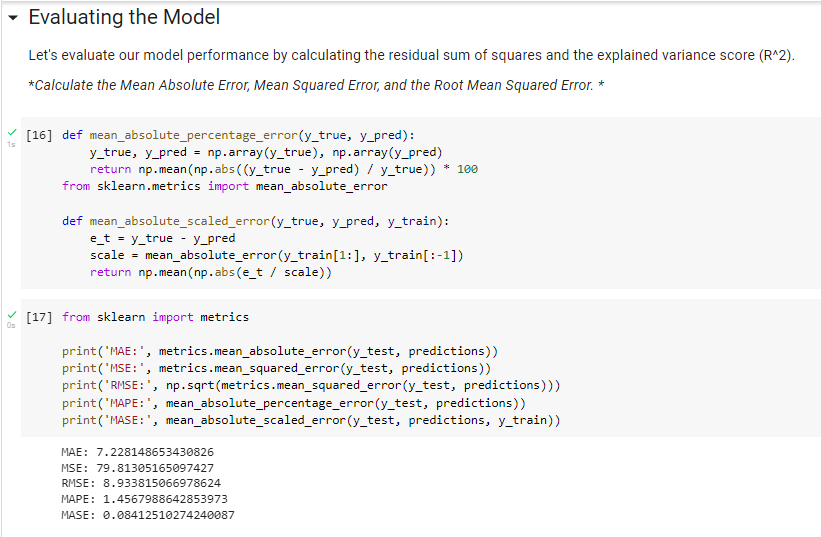
****

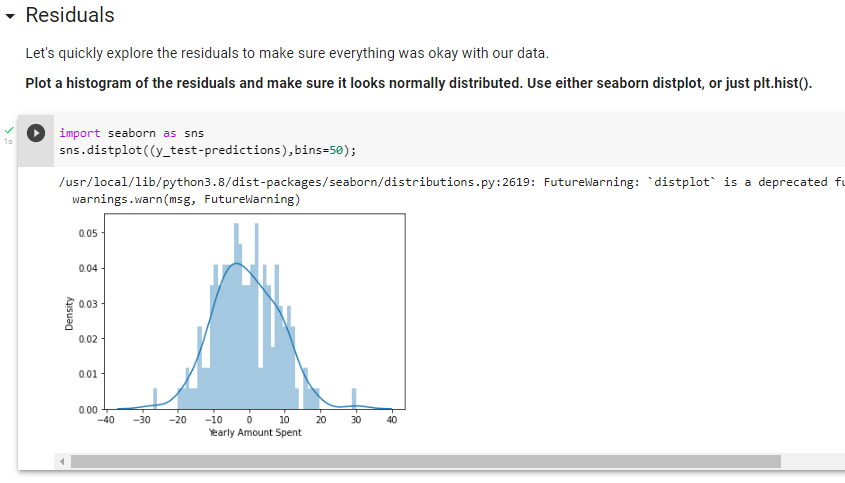
****

****

****

****

****

****

**IV CONCLUSION**

We have understood the basic concepts of Linear Regression and explored the performance evaluation metrics.

**V REFERENCES**

**<https://developers.google.com/machine-learning/crash-course/descending-into-ml/linear-regression>**

**<https://machinelearningmastery.com/regression-metrics-for-machine-learning/>**

**<https://www.scribbr.com/statistics/simple-linear-regression/>**

**VI POST LAB QUESTION/ANSWER**

1. **What do these metrics tell us about the model performance ?**

These metrics give us insight into the performance of a linear regression model and help us understand how well the model fits the data. Here's a brief explanation of what each metric tells us:

1. Mean Absolute Error (MAE): The MAE tells us the average magnitude of the errors in our predictions, with lower values indicating a better fit.
2. Mean Squared Error (MSE): The MSE is similar to the MAE, but it squares the errors before averaging, which gives more weight to larger errors.
3. Root Mean Squared Error (RMSE): The RMSE is the square root of the MSE, which gives the error in the same units as the dependent variable. The RMSE is a commonly used metric for evaluating the performance of regression models.
4. R-squared (R²): The R² is a statistical measure that represents the proportion of the variance in the dependent variable that is predictable from the independent variable(s). A value of 1 indicates that the model perfectly fits the data, while a value of 0 indicates that the model is not useful in explaining the variation in the dependent variable.
5. Mean Absolute Percentage Error (MAPE): The MAPE tells us the average percentage error between the actual and predicted values, which can be useful when the dependent variable has different units or scale.
6. Mean Absolute Scaled Error (MASE): The MASE measures the average error between the actual and predicted values relative to the mean absolute error of a naïve forecast. The MASE is useful for comparing the performance of different models, with lower values indicating a better fit.