Credit Card Fraud Detection

Credit card fraud detection is the collective term for the policies, tools, methodologies, and practices that credit card companies and financial institutions take to combat identity fraud and stop fraudulent transactions.

Building the credit card fraud detection project by performing:

Feature engineering

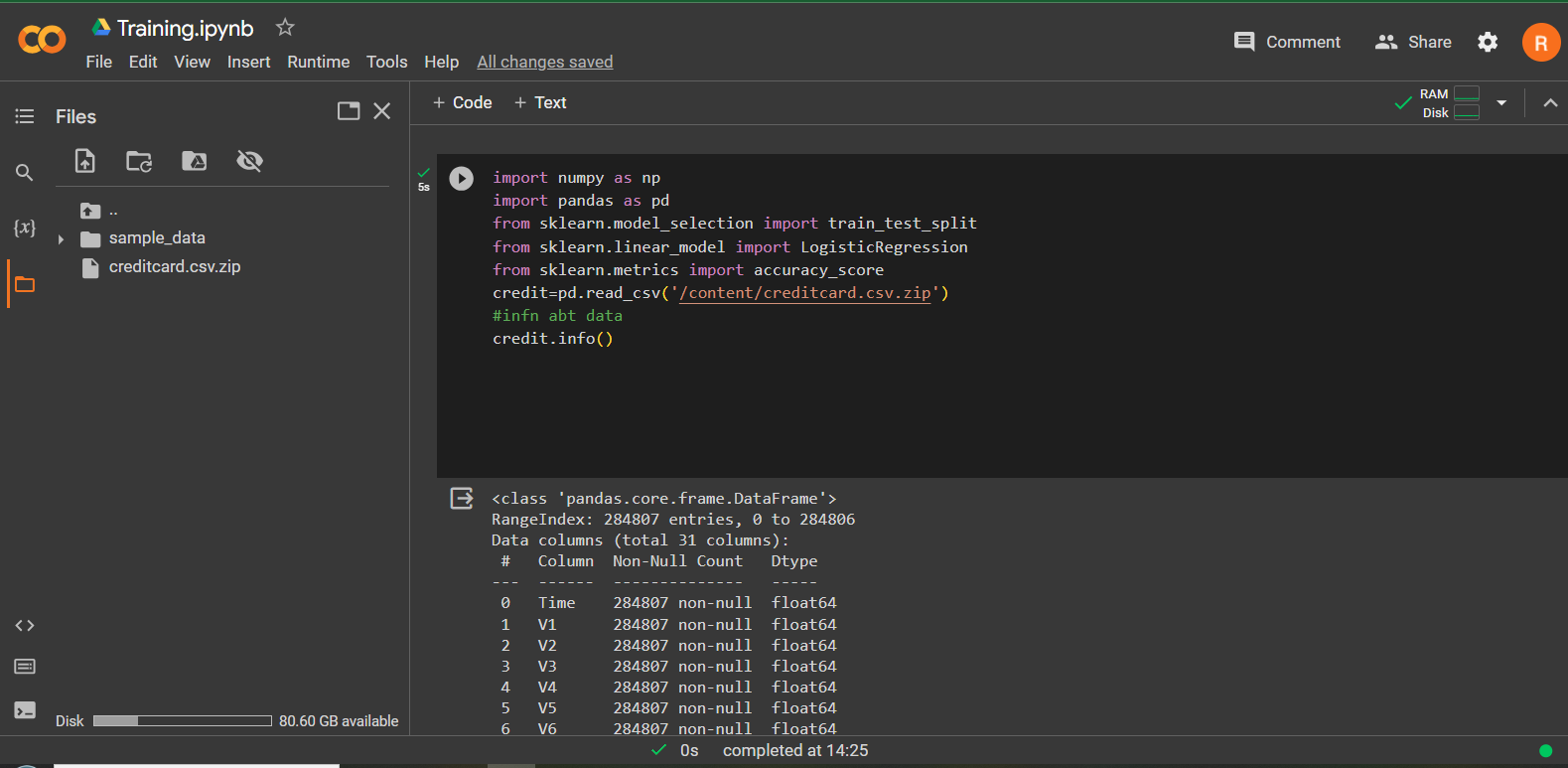
Model training

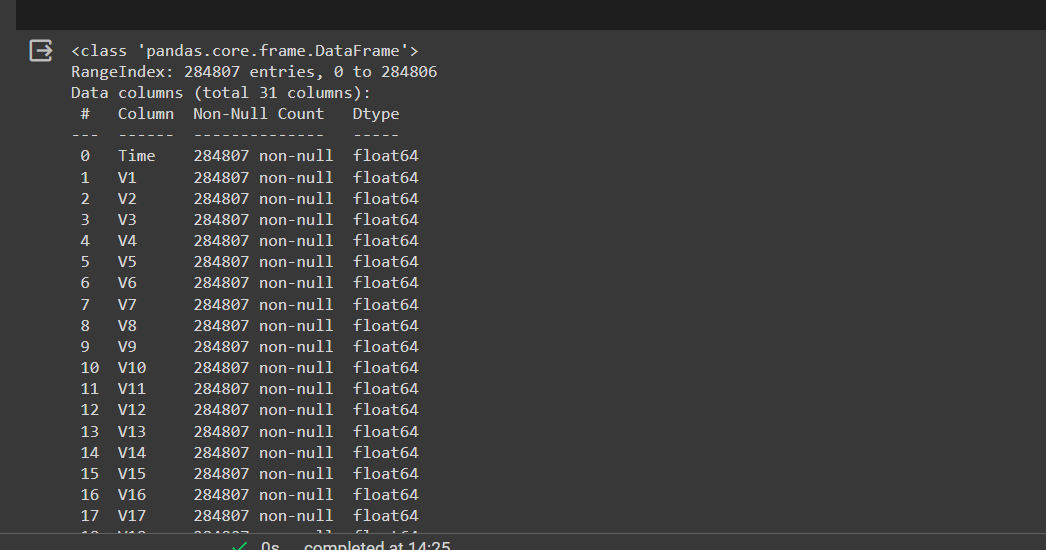
Evaluation.

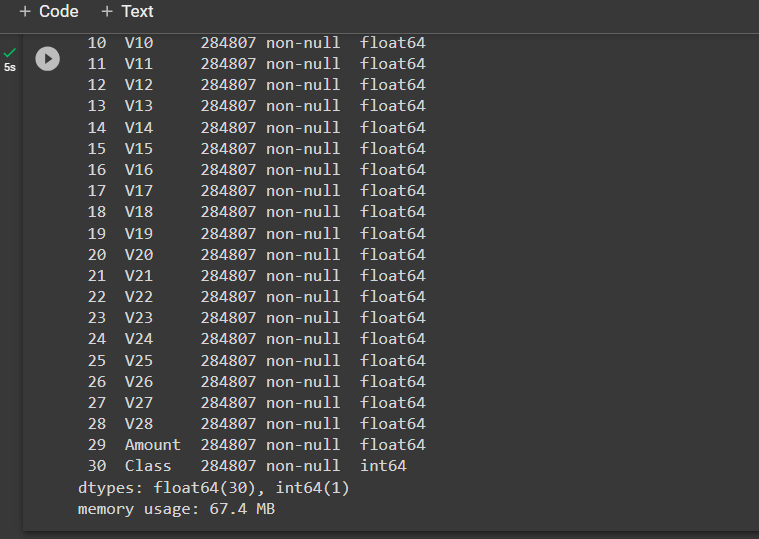
Feature Engineering:

Feature engineering is the process of transforming raw data into features that are suitable for machine learning models. In other words, it is the process of selecting, extracting, and transforming the most relevant features from the available data to build more accurate and efficient machine learning models.

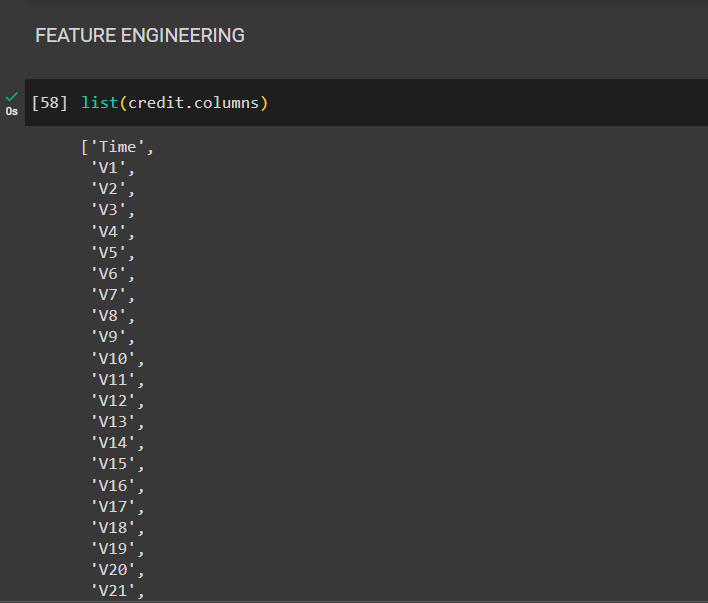
Importing the necessary Libraries

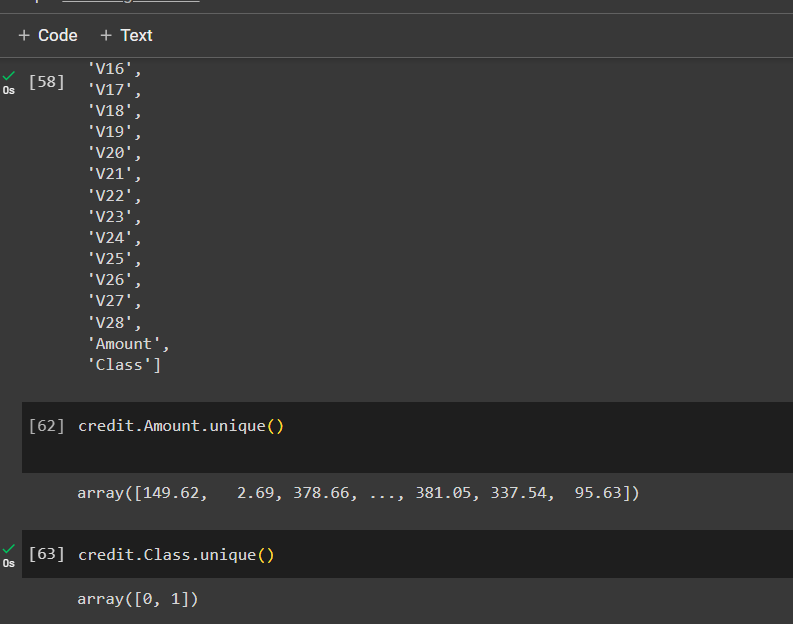




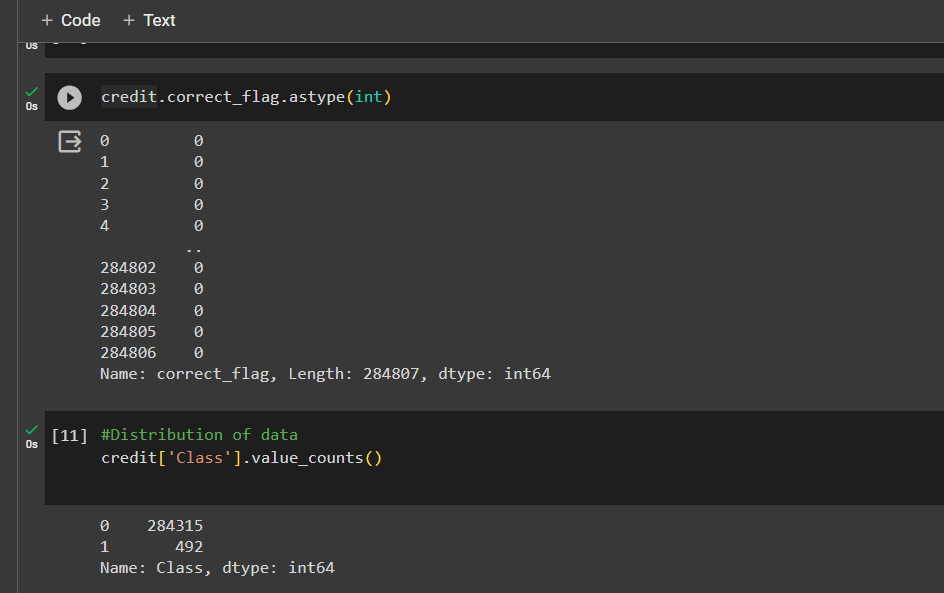


Feature Engineering

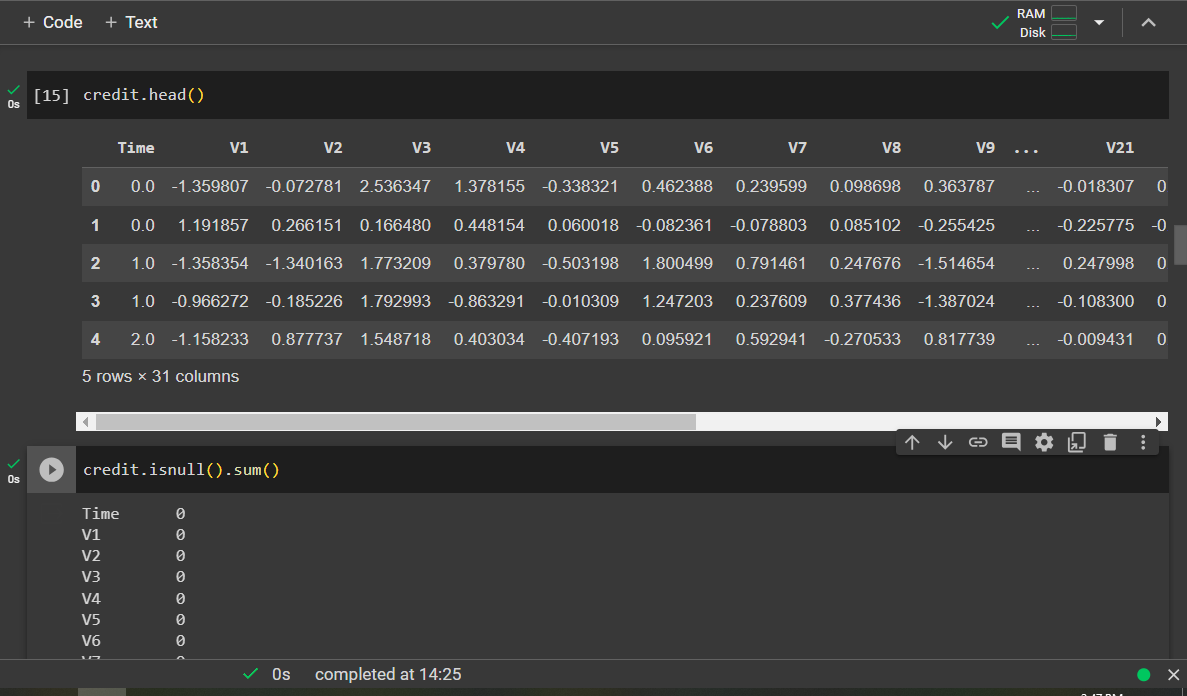


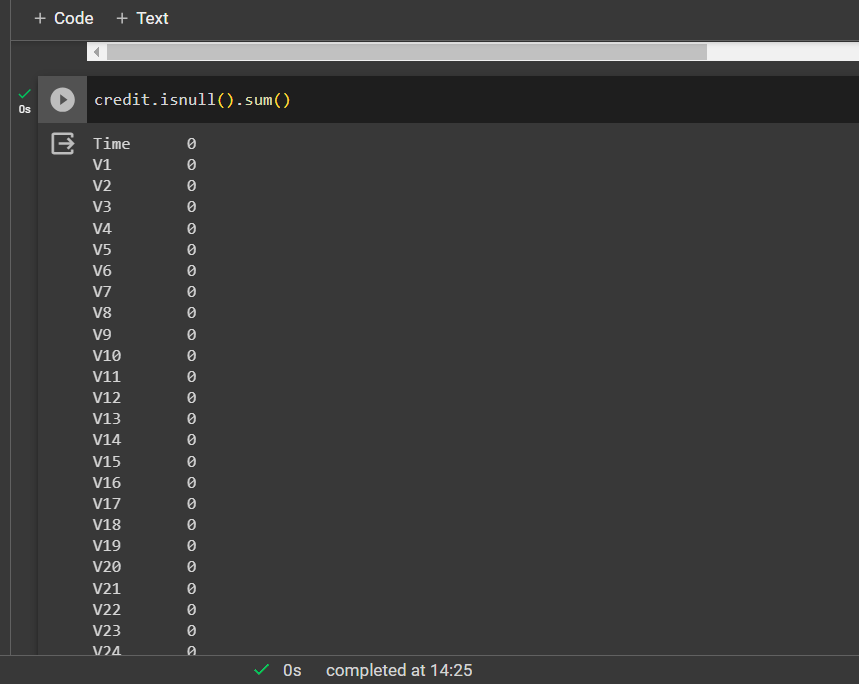


Using astype() function

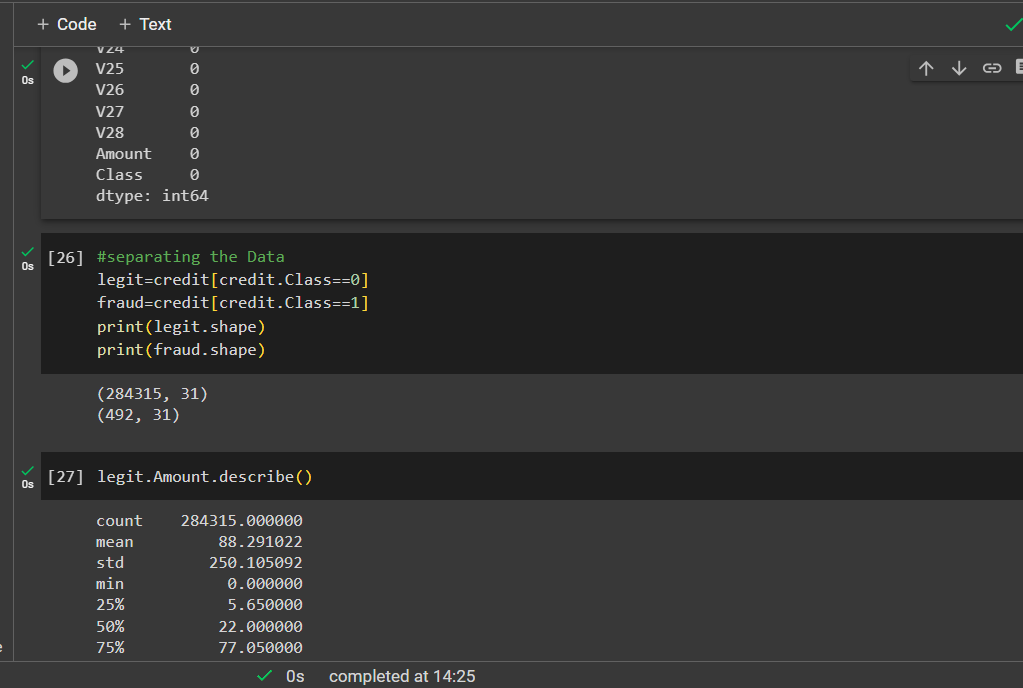


Using head(),isnull() and sum()

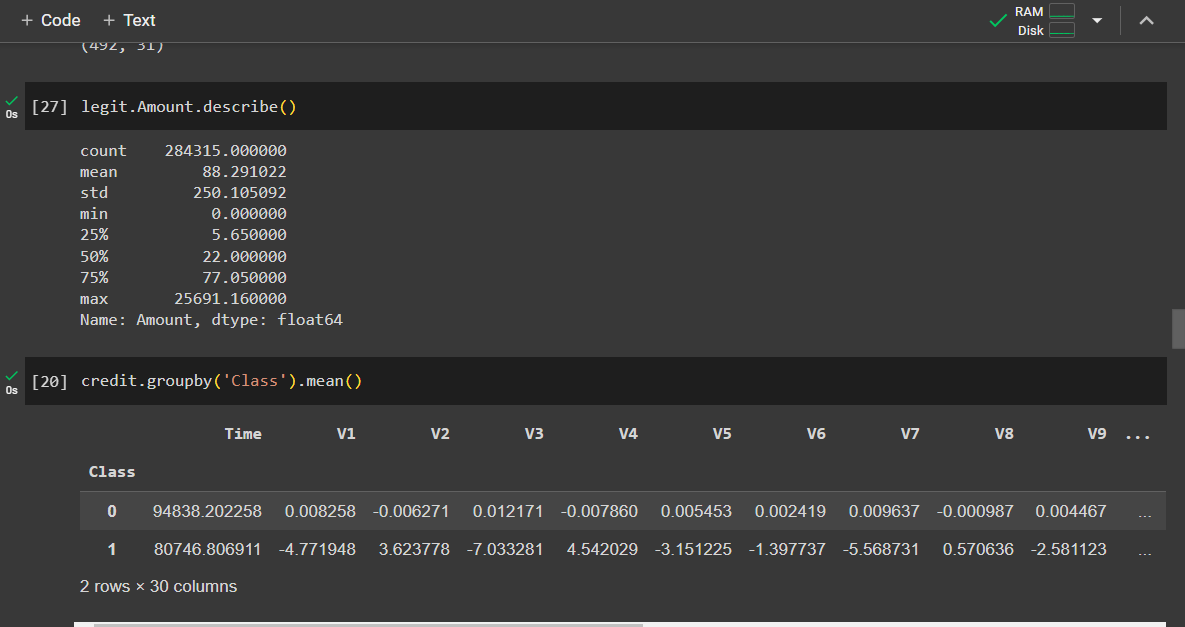




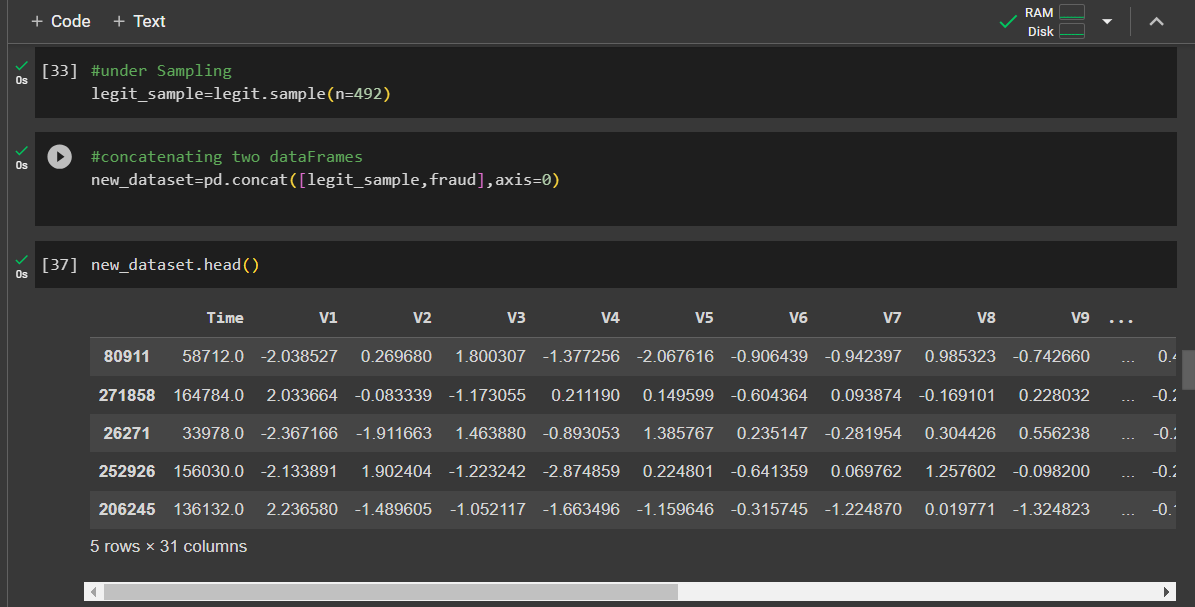
Separating the data



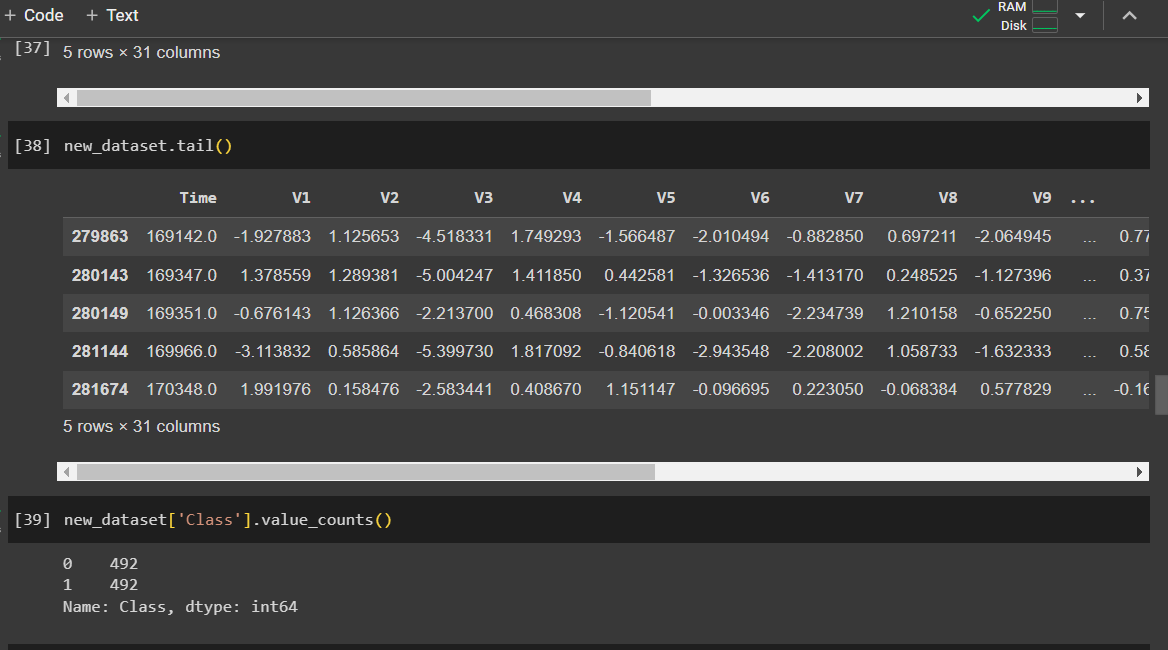
Using describe() and mean()



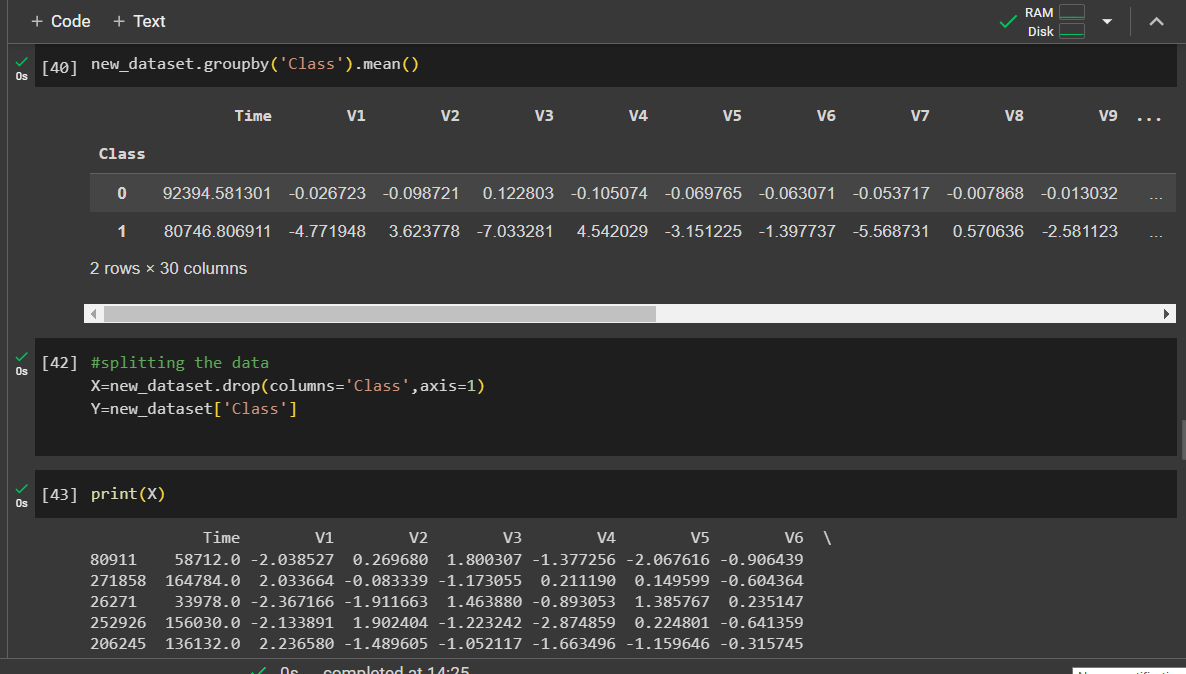
Under Sampling

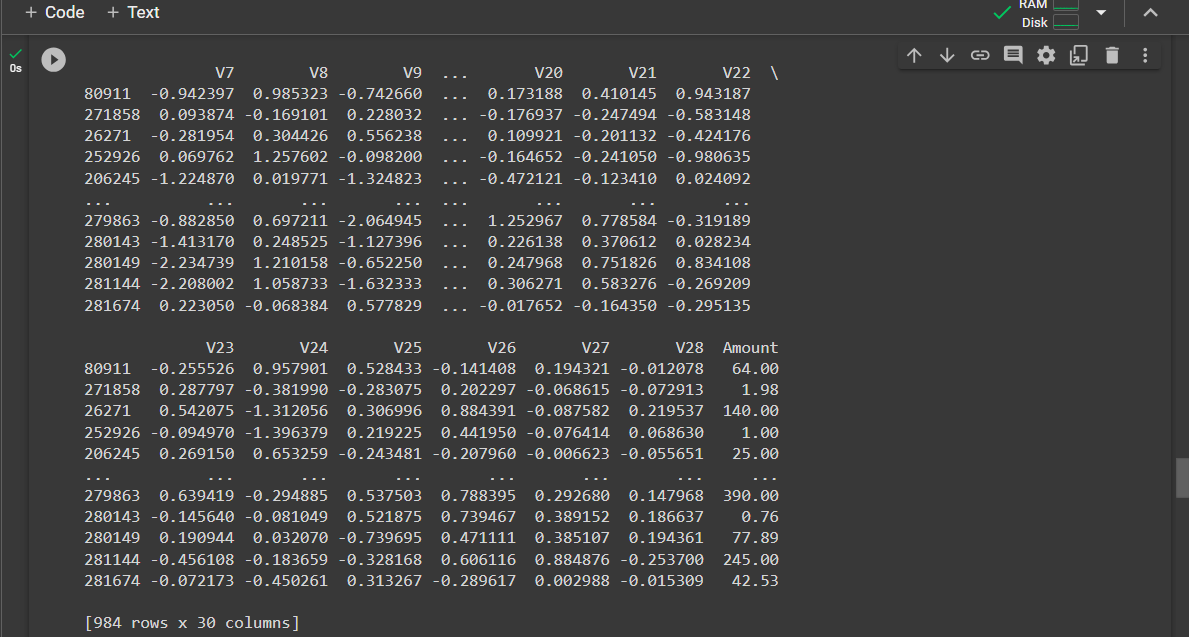


Using value\_counts()

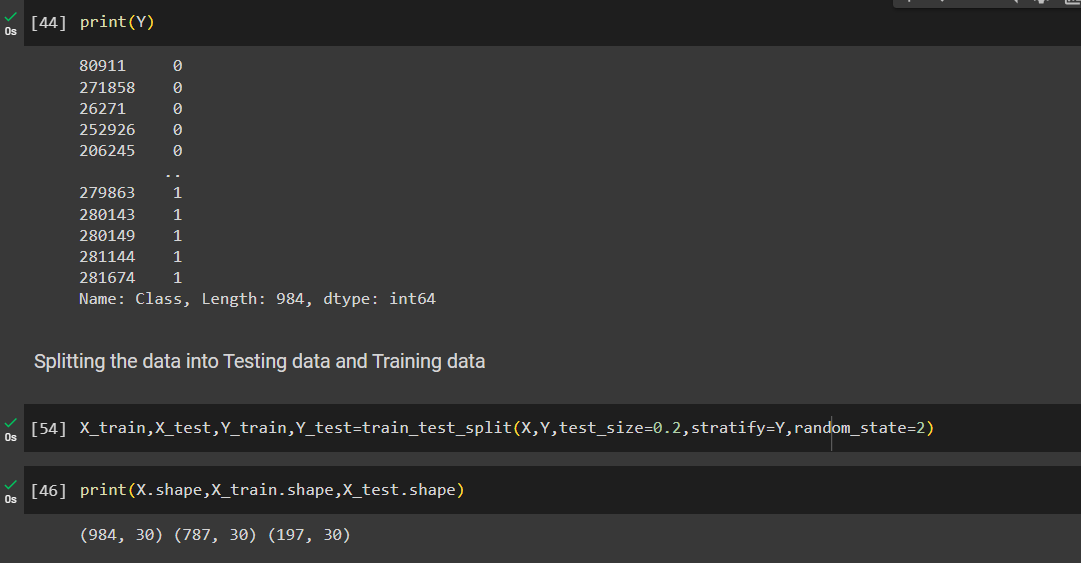


Using groupby()





Splitting the data into Testing data and Training data



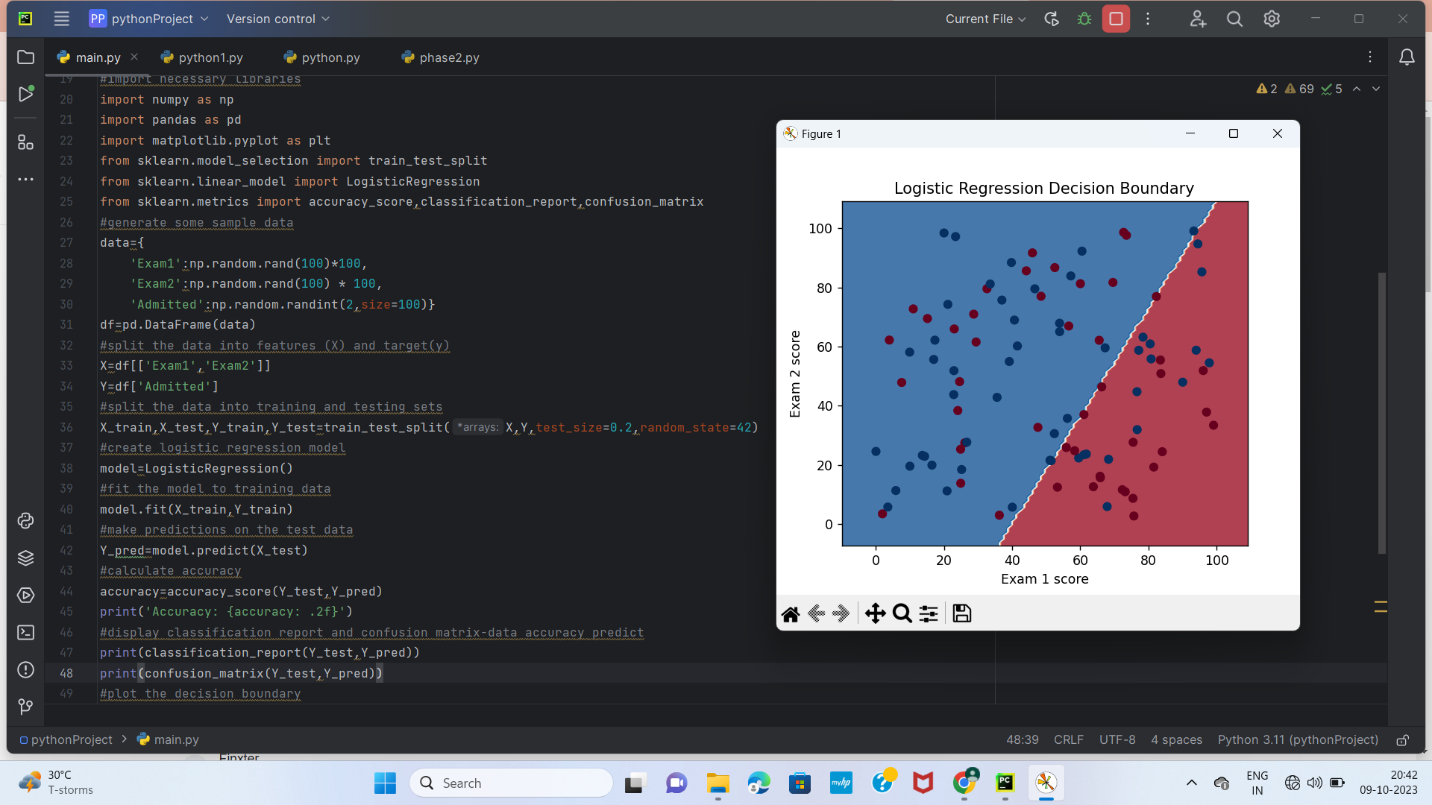
**Model training:**

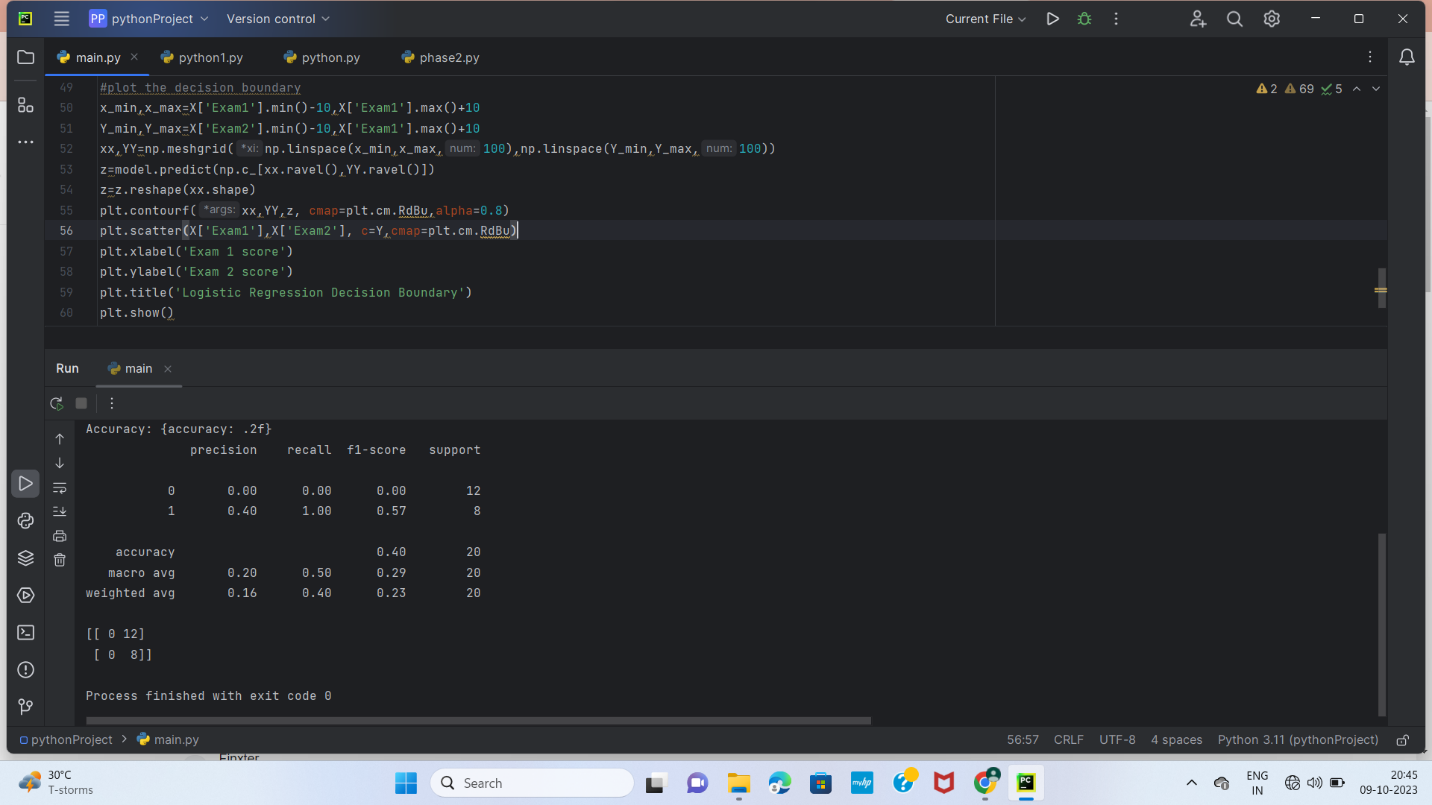
The machine learning models of logistic regression, random forest, and decision trees are evaluated for detecting fraudulent credit card transactions. • Random forest is the most suitable model for predicting fraudulent transactions.

machine learning models were created in the modeling phase, SVM, Logistic Regression . A comparison of the results will be presented later in the paper to know which technique is most suited in the credit card fraudulent transactions detection. The dataset is sectioned into a ratio of 70:30, the training set will be the 70% and remaining set will be the testing set which is the 30%.

Logistic Regression

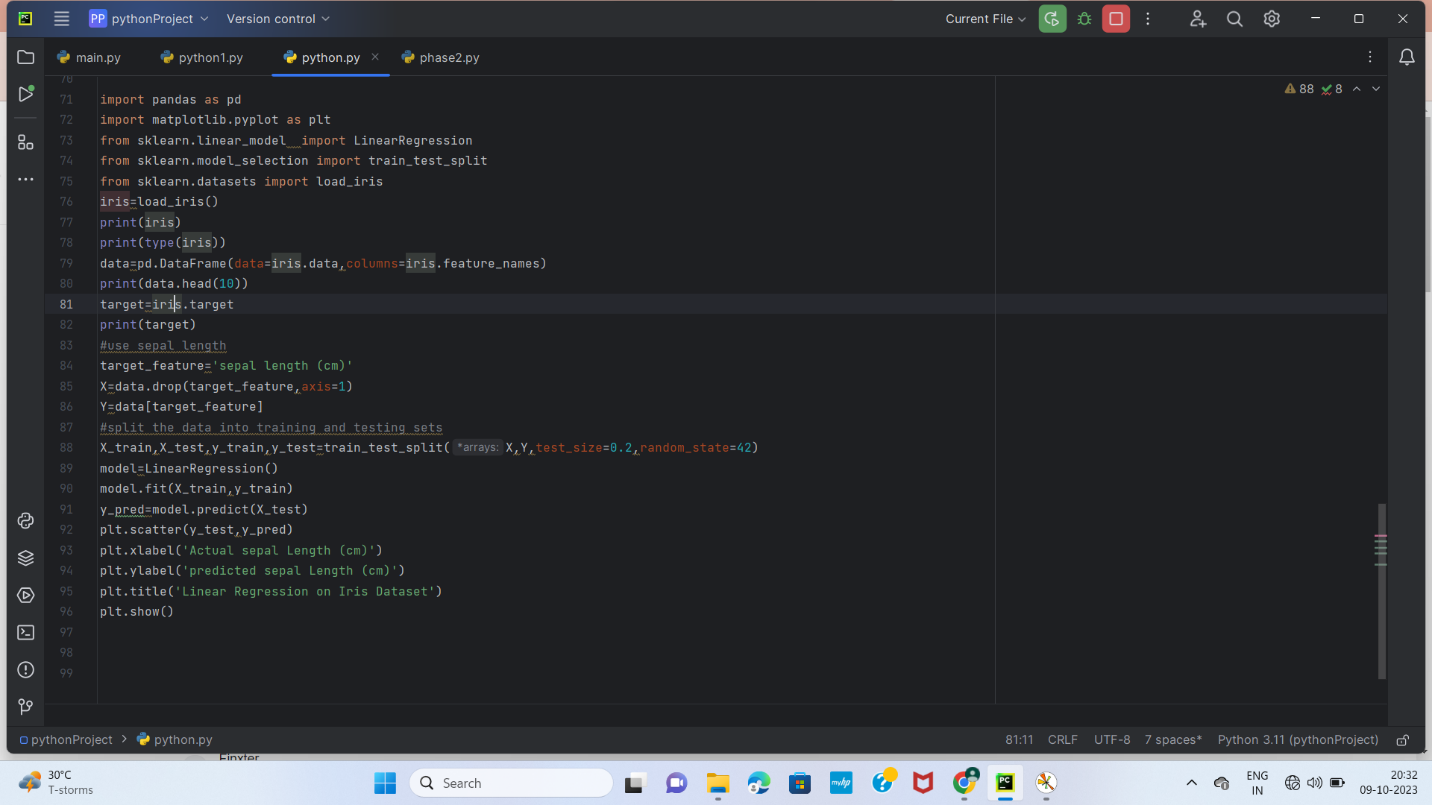
Logistic Regression model is statical model where evaluations are formed of the connection among dependent qualitative variable (binary or binomial logistic regression) or variable with three values or higher (multinomial logistic regression) and one independent explanatory variable or higher whether qualitative or quantitative

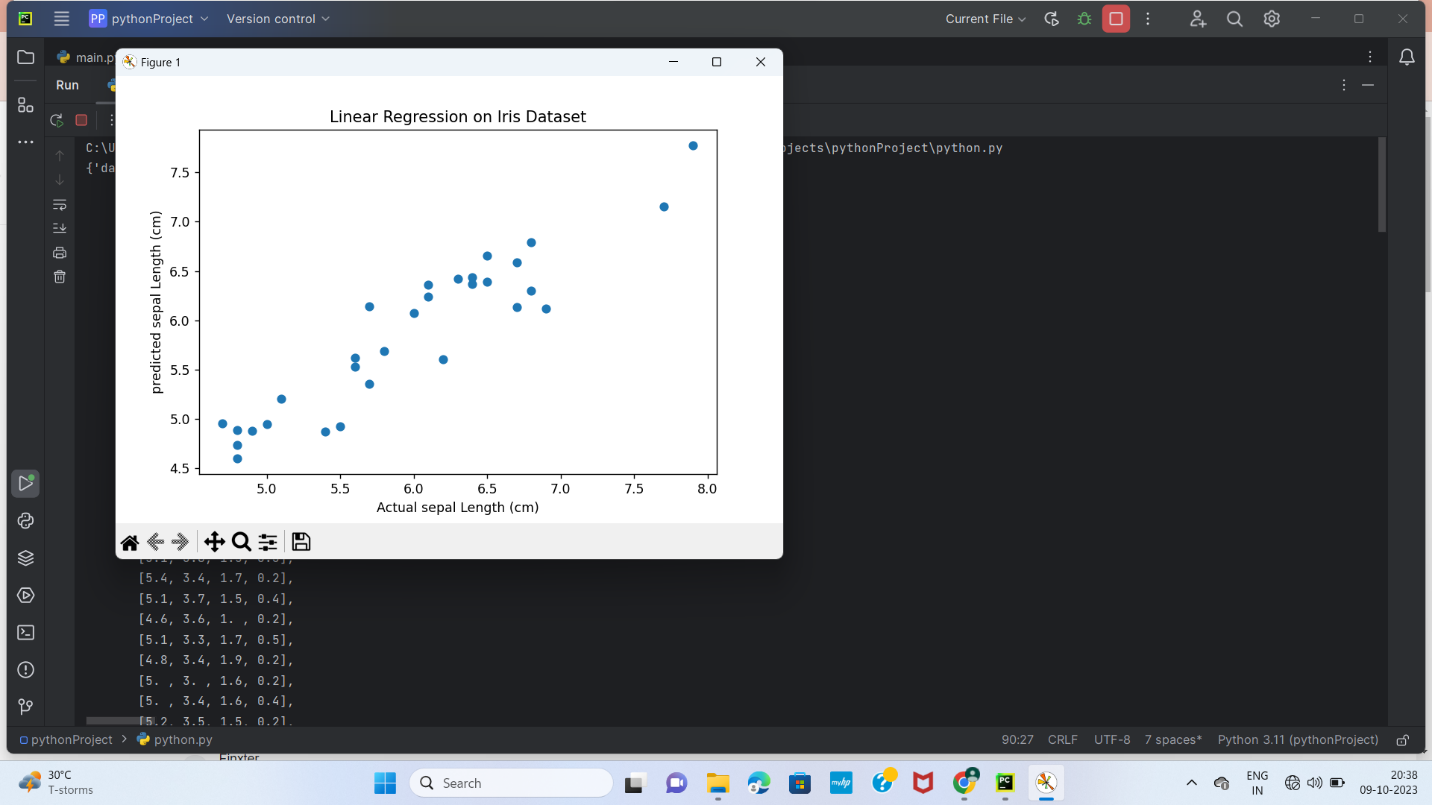


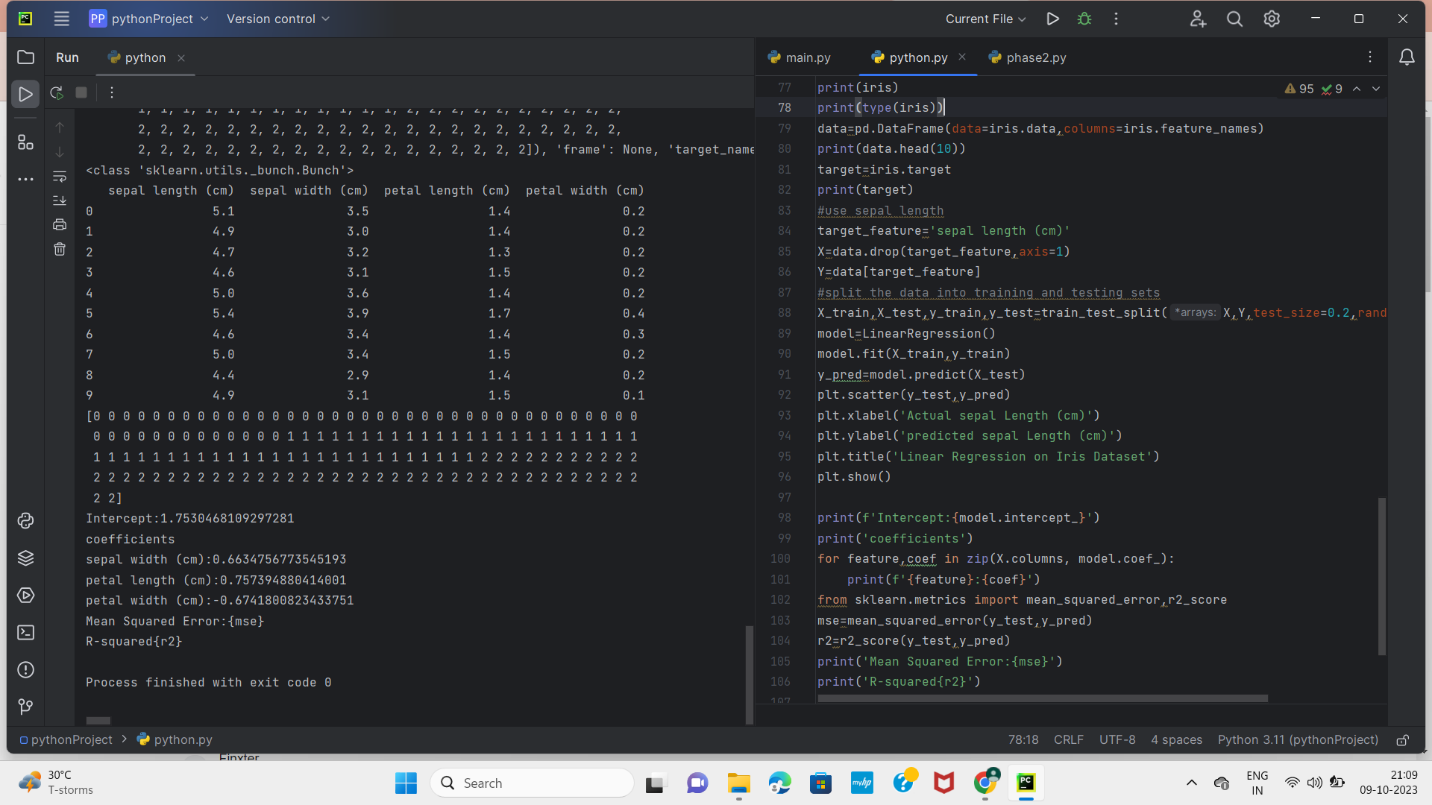


Linear Regression

  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.

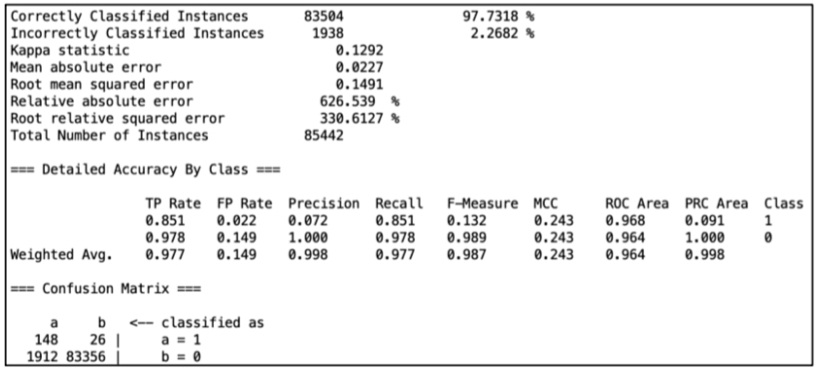






Support Vector Machine

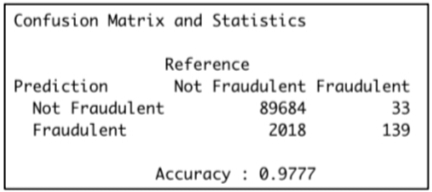
Support Vector machine is a supervised ML technique with connected learning algorithms which inspect data used for both classification and regression analyses, it also performs linear classification, additionally to non-linear classification by creating margins between the classes, which are created in such a fashion that the space between the margin and the classes is maximum which minimizes the error of the classification.



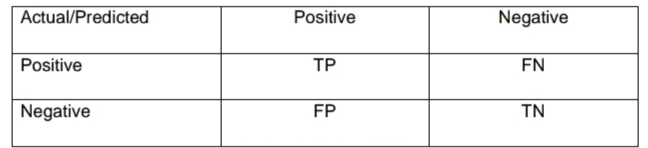
Support Vector Machine

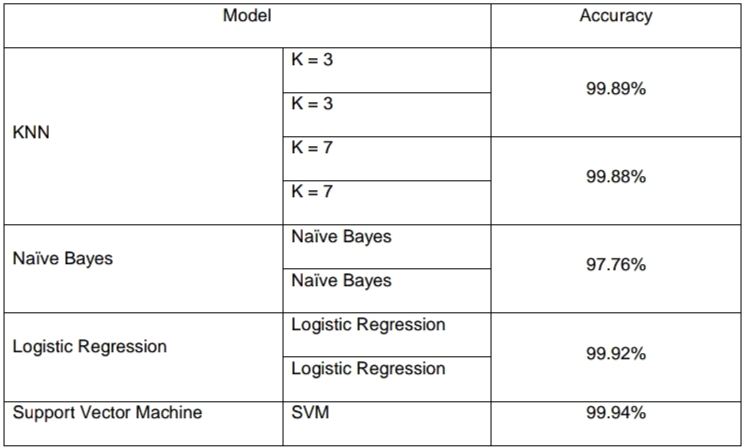
Evaluation

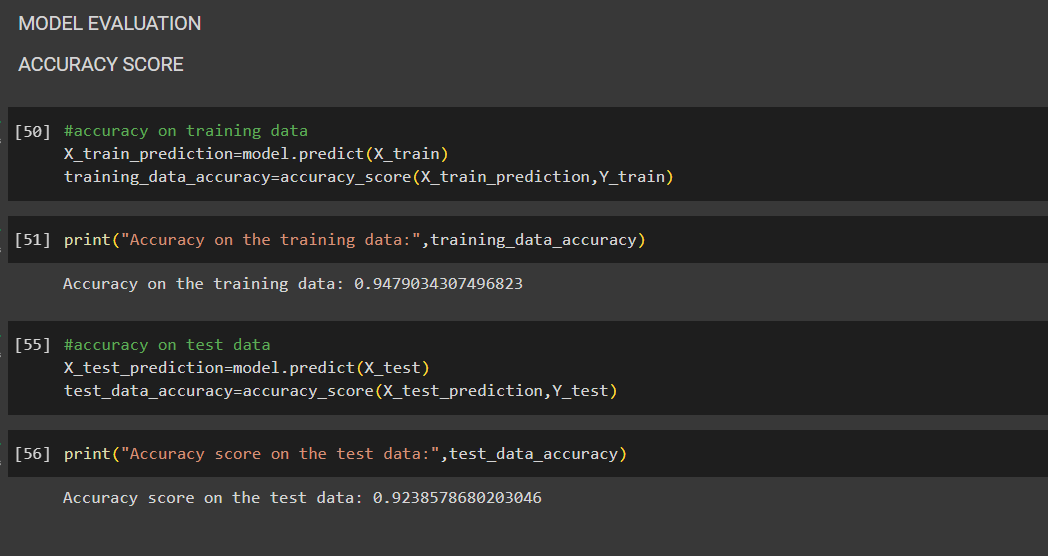
Accuracy is the overall number of instances that are predicted correctly, accuracies are represented by confusion matrix where it showed the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). True Positive represents the transactions that are fraudulent and was correctly classified by the model as fraudulent. True Negative represents the not fraudulent transactions that were correctly predicted by the model as Not fraudulent. The third rating is False positive which represents the transaction that are fraudulent but was misclassified as not fraudulent. And finally False Negative which are the not fraudulent transactions that were identified as fraudulent, table below shows the confusion matrix



Confusion Matrix







Conclusion:

In conclusion, the main objective of this project was to find the most suited model in credit card fraud detection in terms of the machine learning techniques chosen for the project, and it was met by building the four models and finding the accuracies of them all, the best model in terms of accuracies is Support Vector Machine which scored 99.94% with only 51 misclassified instances. I believe that using the model will help in decreasing the amount of credit card fraud and increase the customers satisfaction as it will provide them with better experience in addition to feeling secure.