**Experiment No. 8**

**Title: Mini project on a topic related to Machine learning.**

**Batch: B3 Roll No. 1814050/1814052/1814053 Experiment No.:8**

**Title: Mini project on a topic related to Credit Card Fraud Detection.**

Resources needed: Jupyter Notebook, Kaggle Dataset

Problem Definition:

A credit card remains a very widespread compensation method that is accepted online & offline, and provides cashless transactions. With the increase of developments credit card frauds are also growing. Billion dollars are at loss due to these fraudulent acts. In directive to minimalize disorder and bring order in place, having a well-organized method of fraud detection has become a need for all banks. A machine learning based technique is proposed in this direction to detect the credit card frauds are also growing.

Billion dollars are at loss due to these fraudulent acts. In directive to minimalize disorder and bring order in place, having a well-organized method of fraud detection has become a need for all banks. A machine learning based technique is proposed in this direction to detect the credit card frauds.

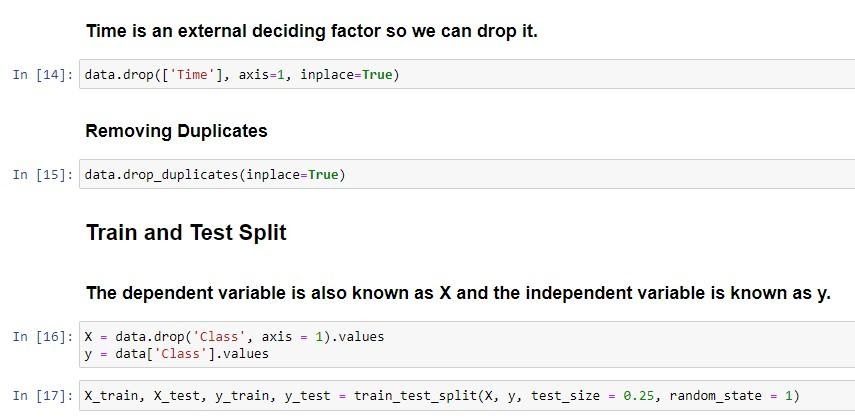
Dataset Description:

The dataset contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

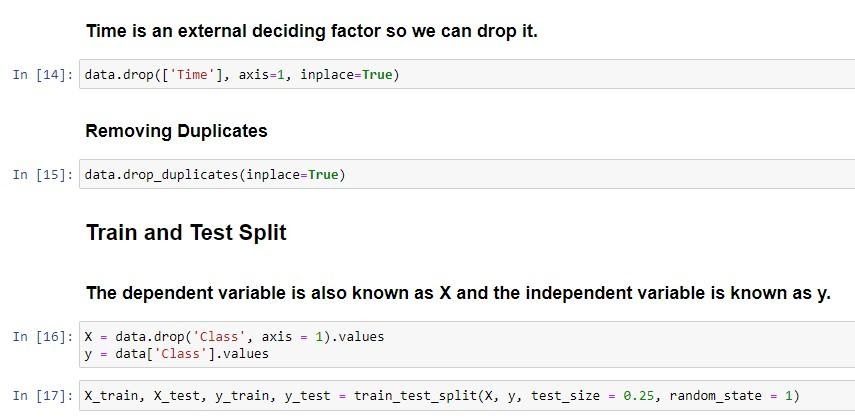
Data Pre-processing:

It contains only numerical input variables which are the result of a PCA transformation. Unfortunately, due to confidentiality issues, we cannot provide the original features and more background information about the data. We can find 29 feature columns and 1 final class column. Features V1, V2, … V28 are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'.



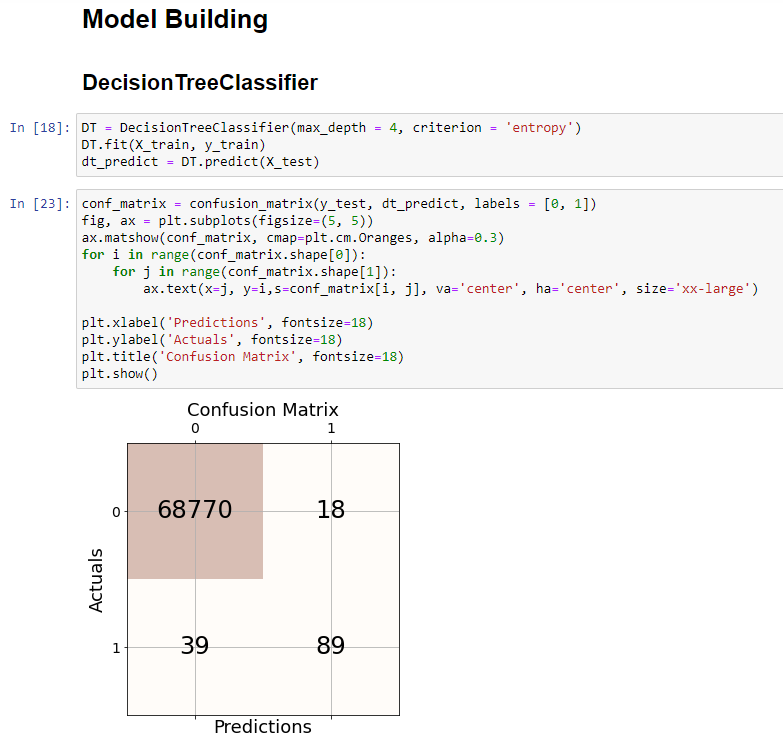


Splitting the model in training and testing:

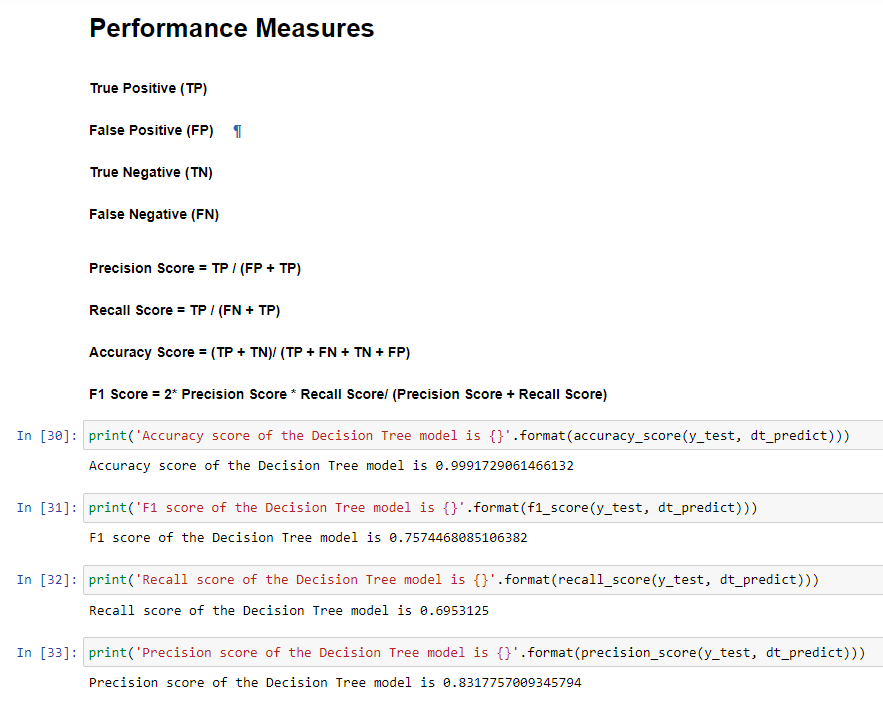


Model Building:

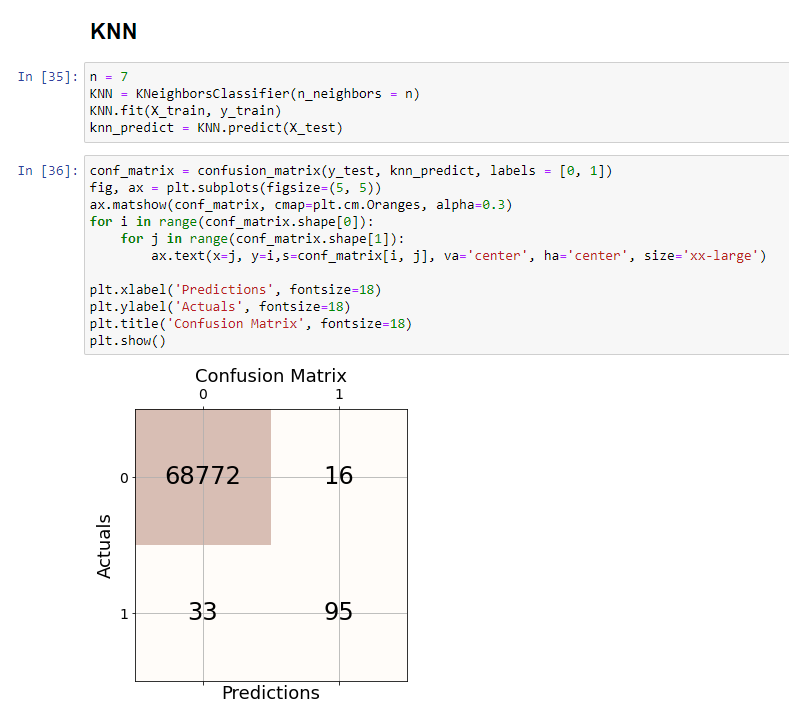
1. **Decision Tree Classifier:**



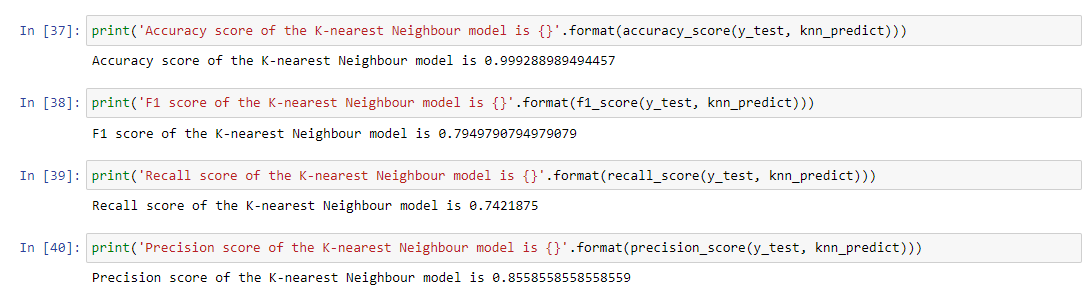
Performance measures:



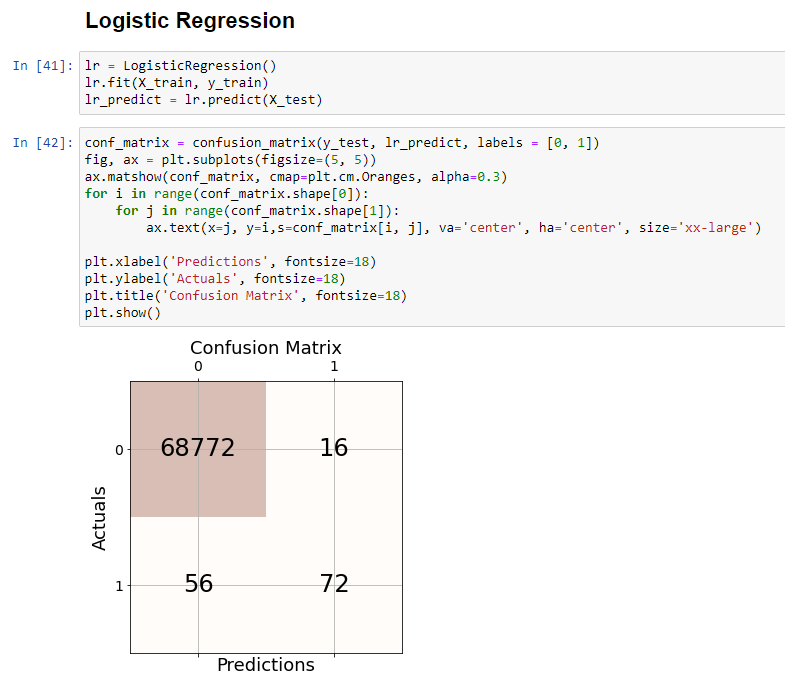
1. **K Nearest Neighbour (KNN):**



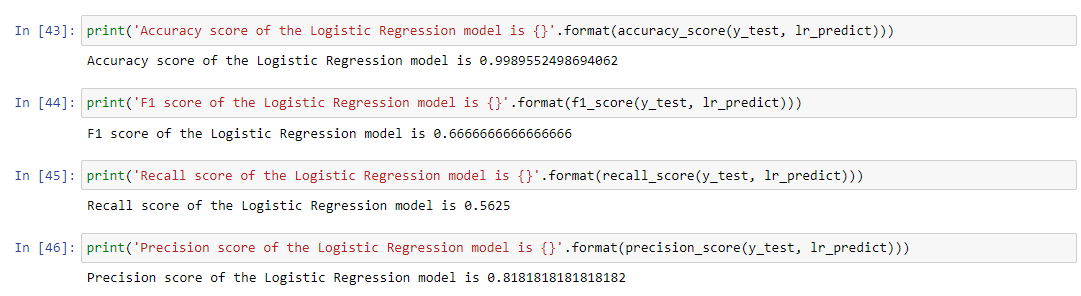
Performance measures:



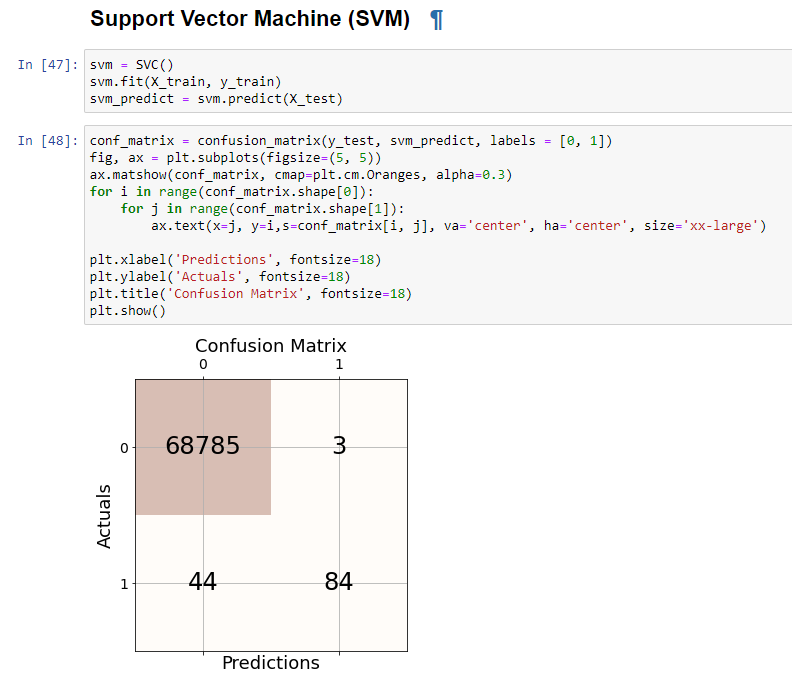
1. **Logistic Regression:**



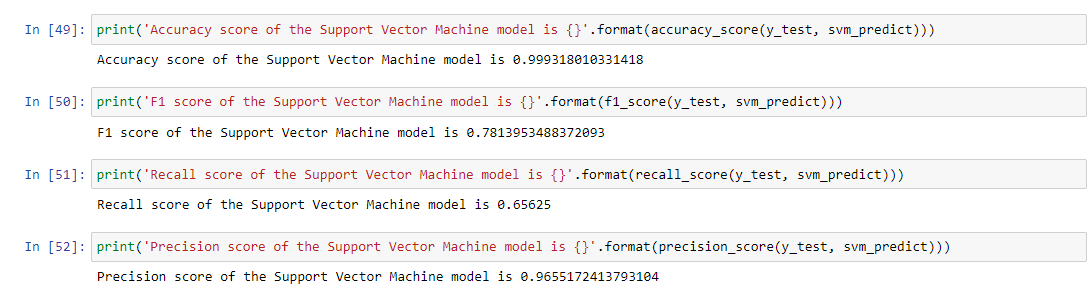
Performance measures:



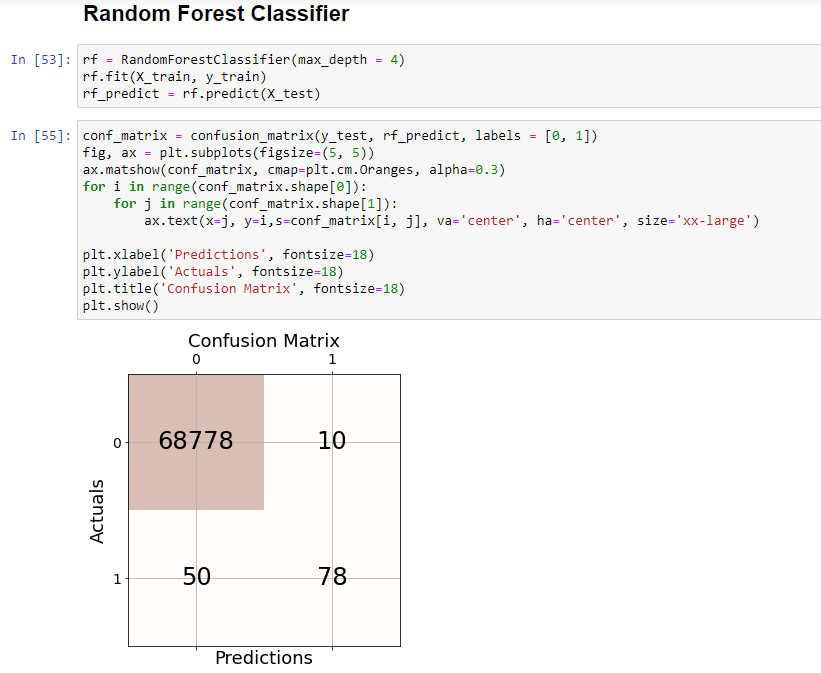
1. **Support Vector Machine (SVM):**



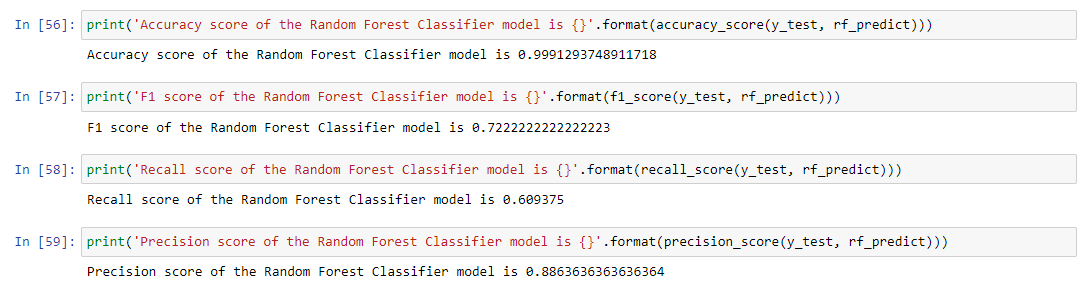
Performance measures:



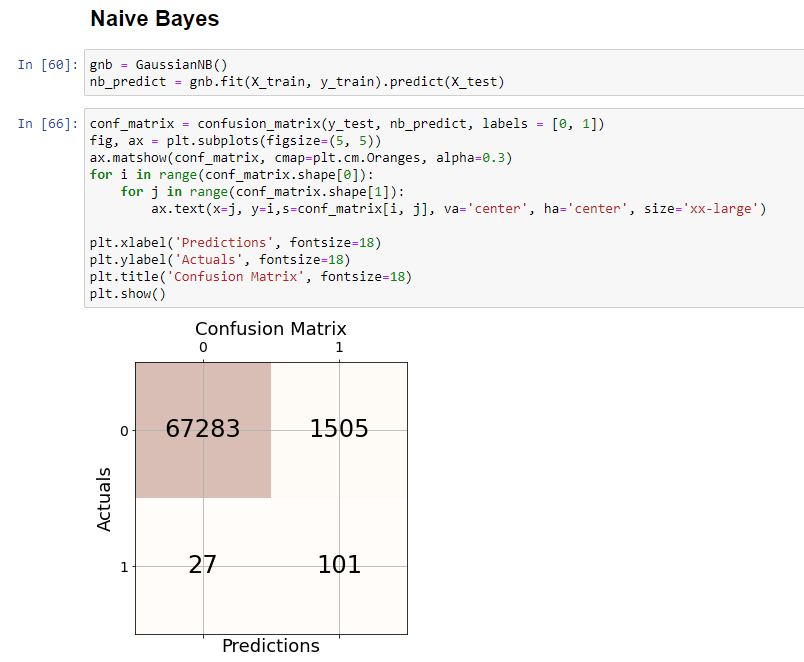
1. **Random Forest Classifier:**



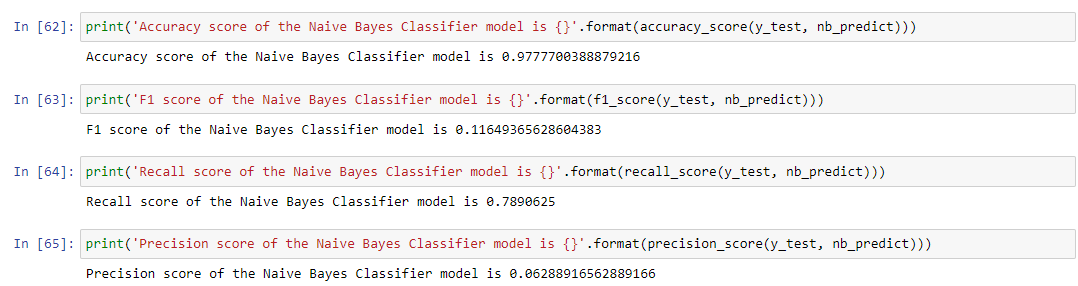
Performance measures:



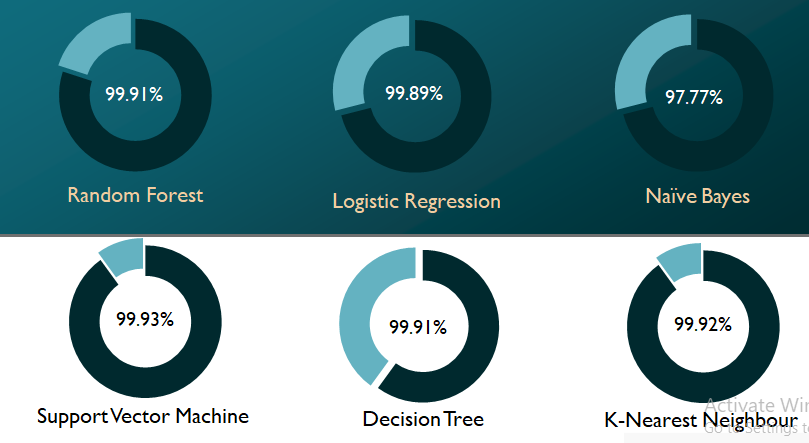
1. **Naïve Bayes:**



Performance measures:



Results:



Conclusion:

* The six algorithms are tested on the basis of their prediction accuracy and confusion matrix. These parameters are used for comparison of performance.
* A support vector matrix is better than any other technique as it provides an accuracy of 99.93%.
* All algorithms work almost the same, with small differences, but we can consider that if these algorithms are trained with more data from the real world, the efficiency and prediction will be improved.

Outcomes:

CO3: Comprehend radial-basis-function(RBF) networks and Kernel learning method Work

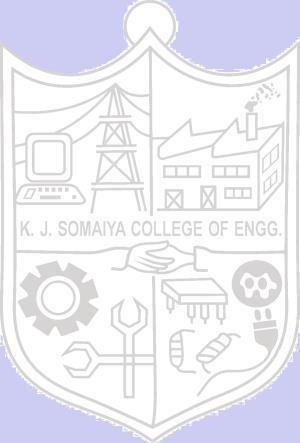
Github Link: https://github.com/dhairyashah10/Group12\_Credit-Card-Fraud-Detection.git

References:

https://www.kaggle.com/mlg-ulb/creditcardfraud

https://towardsdatascience.com/credit-card-fraud-detection-using-machine-learning-python-5b098d4a8edc

Grade: AA / AB / BB / BC / CC / CD /DD

Signature of faculty in-charge with date References:

Books/ Journals/ Websites:

KJSCE/IT/LY-B.Tech/SEM-VII/ML/2021-22