Anomaly Detection in Payments

IE 406 : Machine Learning

Group no. 18

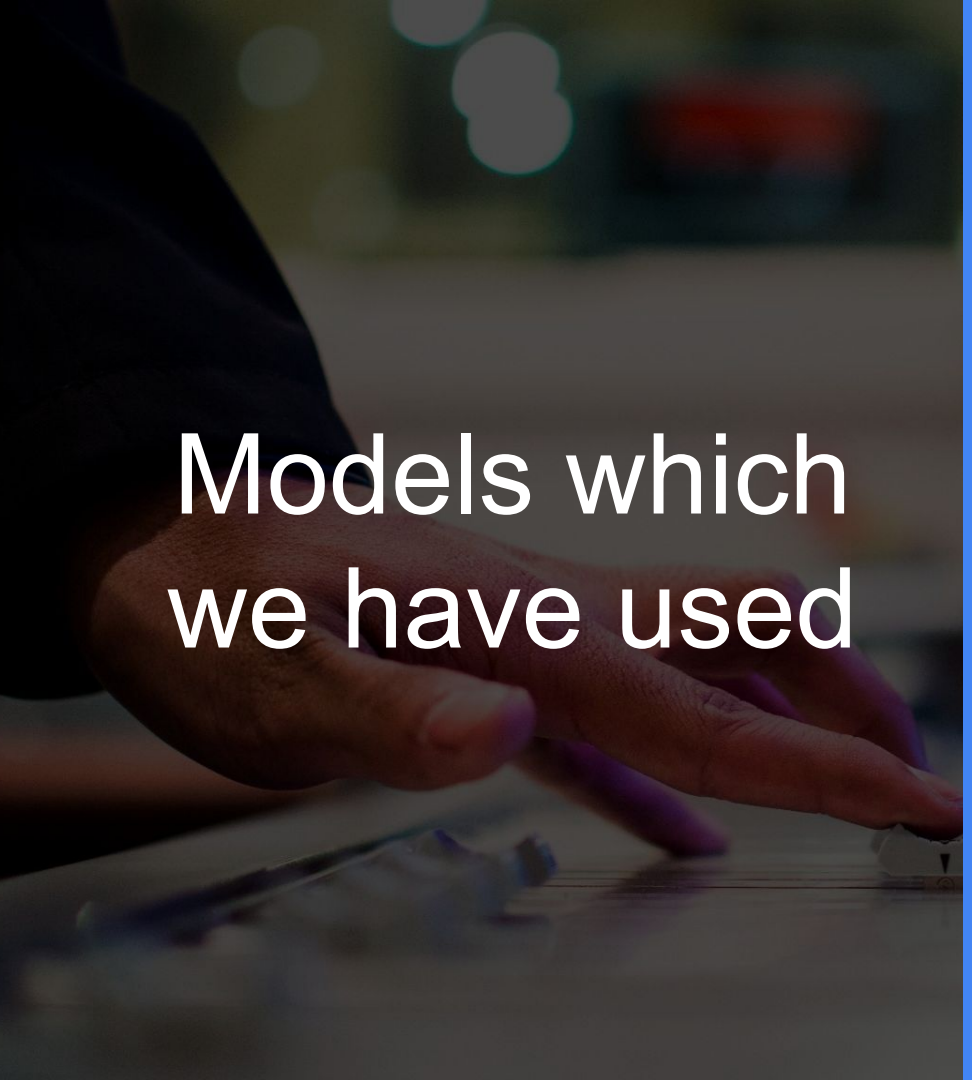
Assigned By

Prof. M.V. Joshi

Problem Statement

In dataset, there are transactions made by credit cards in September 2013 by european cardholders. As name of the project suggests, The goal is to separate fraudulent and normal transactions.

Github link :



Models which we have used

1. Logistic Regression
2. SVM (kernel)
3. Isolation Forest
4. Local Outlier Factor (LOF)

Logistic Regression:-

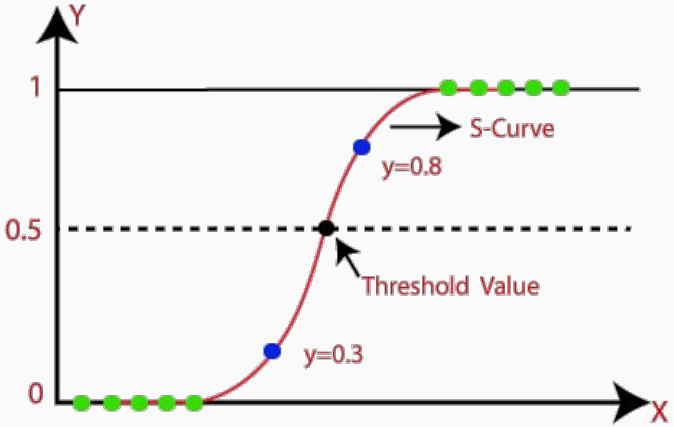
Number of Frauds :

Number of Frauds detected :

Number of Normal Transaction detected as Fraud :

Number of Frauds detected as Normal Transaction:

Fraud Detection Accuracy :



Support Vector Machine:-

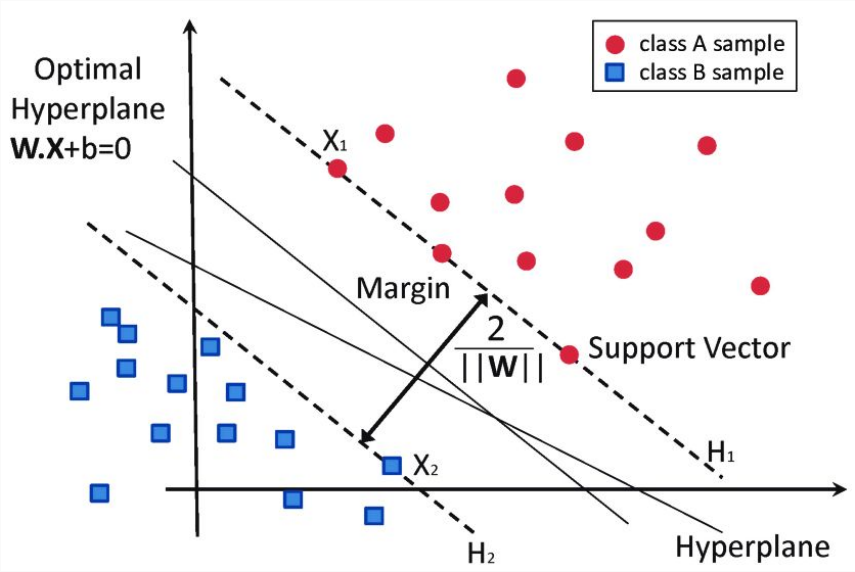
Number of Frauds :

Number of Frauds detected :

Number of Normal Transaction detected as Fraud :

Number of Frauds detected as Normal Transaction:

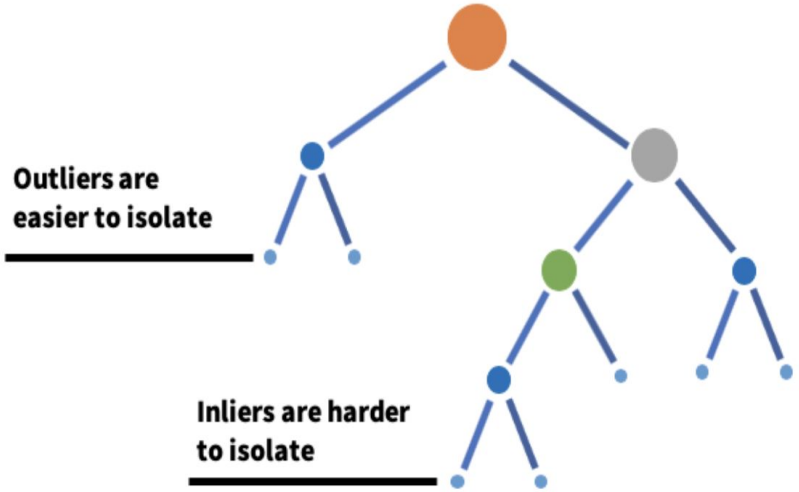
Fraud Detection Accuracy :



Why supervised algorithms aren’t good?

1. **Imbalanced dataset :** The dataset is highly unbalanced, the frauds account for 0.172% of all transactions. we have 492 frauds out of 284,807 transactions.
2. **Concept drift :** It means that the statistical properties of the target variable, which the model is trying to predict, change over time in unforeseen ways. This causes problems because the predictions become less accurate as time passes.

Isolation Forest



Number of Frauds :
Number of Frauds detected :

Number of Normal Transaction detected as Fraud :

Number of Frauds detected as Normal Transaction:

Fraud Detection Accuracy :

Local Outlier Factor(LOF)

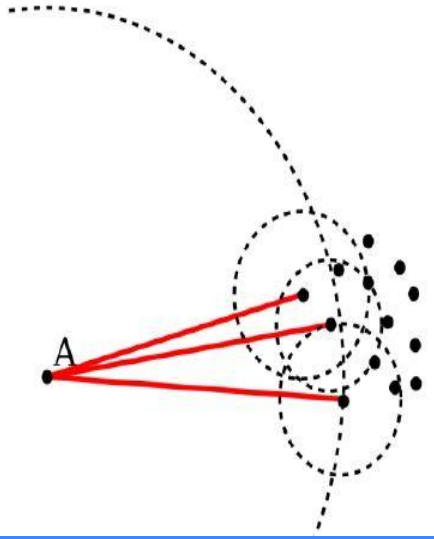
Number of Frauds :
Number of Frauds detected :

Number of Normal Transaction detected as Fraud :

Number of Frauds detected as Normal Transaction:

Fraud Detection Accuracy :

Basic idea of LOF :- comparing the local density of a point with the densities of its neighbors. A has a much lower density than its neighbors.**[4]**



Previous Works / Dataset / References

1. Dataset : https://www.kaggle.com/mlg-ulb/creditcardfraud
2. Local Outlier Factor, Isolation Forest : https://ieeexplore.ieee.org/document/8741421
3. Logistic regression, SVM, Random Forest : https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8757212
4. https://en.wikipedia.org/wiki/Local_outlier_factor
5. https://content.linkedin.com/content/dam/engineering/site-assets/images/blog/posts/2019/08/IsolationForest1.png


Thank You..!