Anomaly Detection in Payments

IE 406 : Machine Learning

Group no. 18

Assigned By

Prof. M.V. Joshi

Problem Statement

In training dataset, there are transactions made by credit cards in September 2013 by european cardholders. We want to to separate fraud and non-fraud transactions by obtaining a decision boundary in the feature space defined by input transactions using supervised binary classifiers. In addition to that because of Imbalanced dataset and concept drift, We are going to use unsupervised algorithms which detects outlier (Fraudulent transactions). Compare these models to know which model has best results for given dataset.
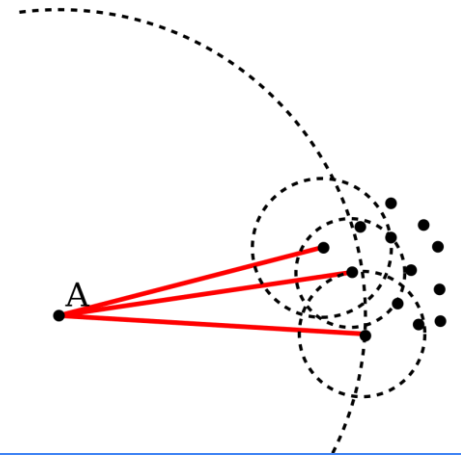
Models which we are going to use

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

Isolation Forest :-

Outliers are easier to isolate
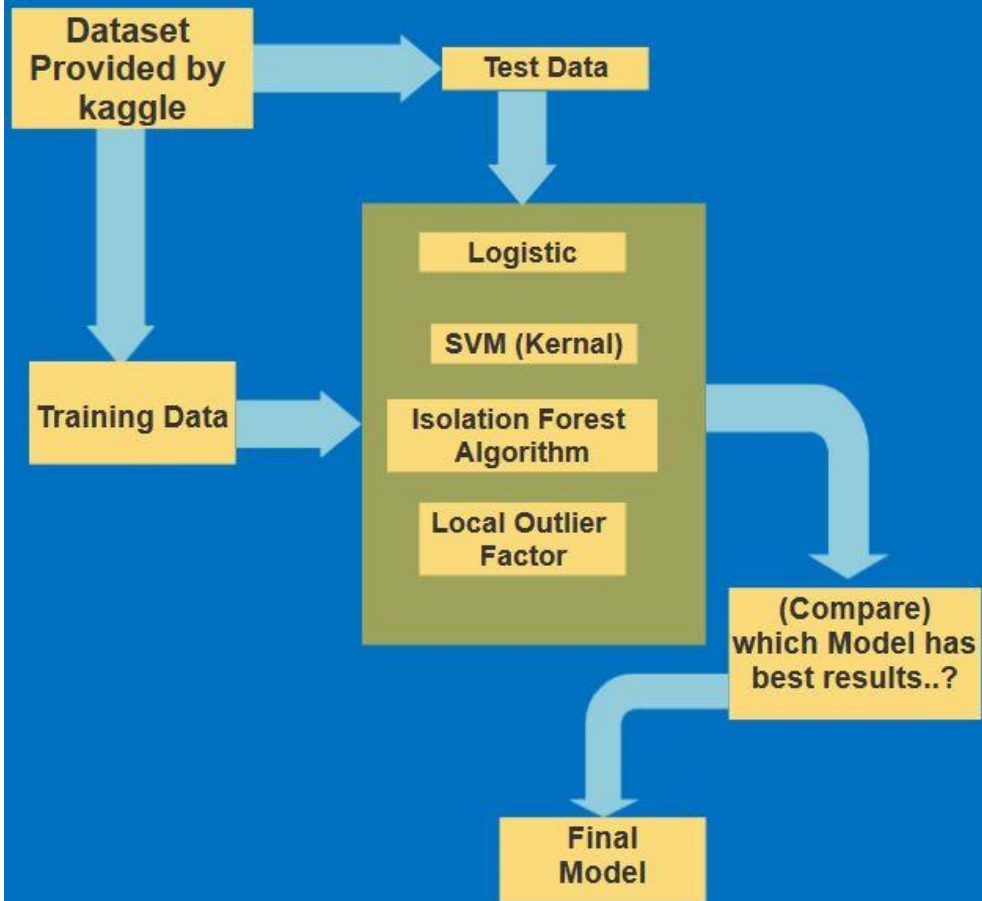
Inliers are harder to isolate
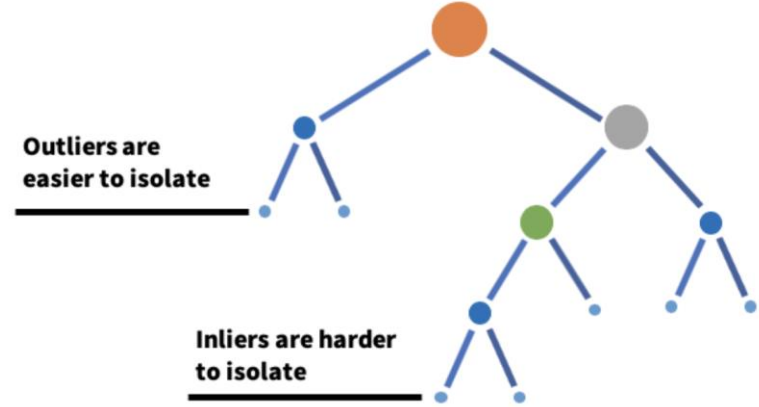
[5]

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]**

A

Block Diagram:-

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..!