Credit card fraud detection using Machine Learning techniques

Anzhel Abgaryan
Applied Statistics and Data Science, Yerevan State University

Abstract

Credit card fraud is an inclusive term for fraud committed using a payment card, such as a credit card or debit card. The purpose may be to obtain goods or services, or to make payment to another account which is controlled by a criminal. The Payment Card Industry Data Security Standard (PCI DSS) is the data security standard created to help businesses process card payments securely and reduce card fraud.

Introduction

Nowadays payment operators and financial institutions are mostly relying on different risk management tools to identify and detect fraudulent transactions. Those tools are fraud detection systems, which are working mostly based on using machine learning algorithms. In this project, I used some popular machine learning models, including logistic regression, decision trees, support vector classification and k-nearest neighbor classification, and a real-life dataset from Kaggle. The main challenge of this project is that in real life, the fraud transactions usually represent a very small percent of all the transactions. The data, which was used for this project is containing only 0.2 % of fraudulent transactions. This means that the dataset is heavily imbalanced, and standard approaches cannot be a solution for this case.

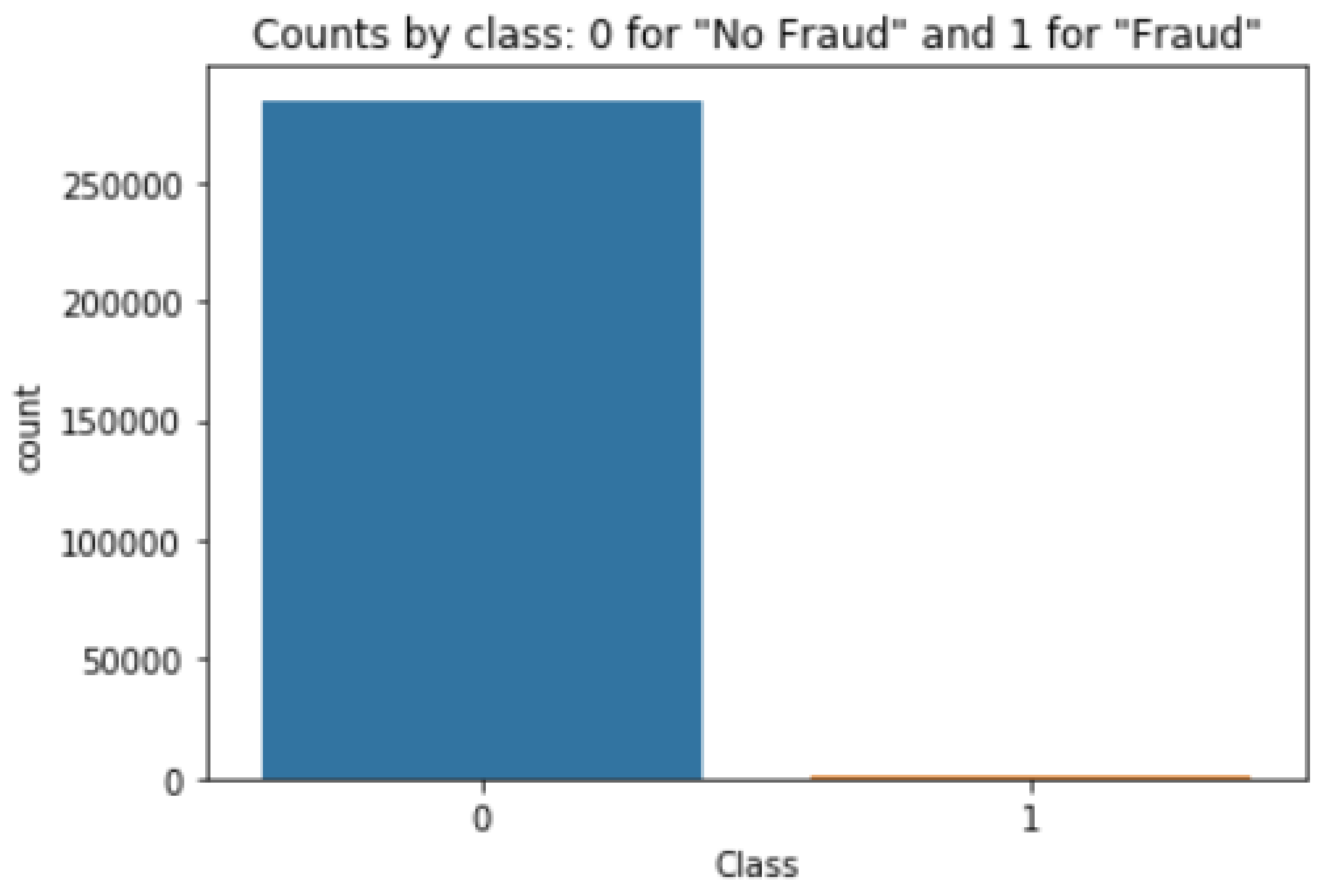
Dataset and features

The dataset contains 31 variables and about 300,000 rows of credit card transactions. Table 1 provides detailed description of the dataset and variables, which comes from Kaggle. As you can see another challenge of this dataset is that 28 out of the total 31 variables are the result of dimensionality reduction which was done to protect sensitive information.

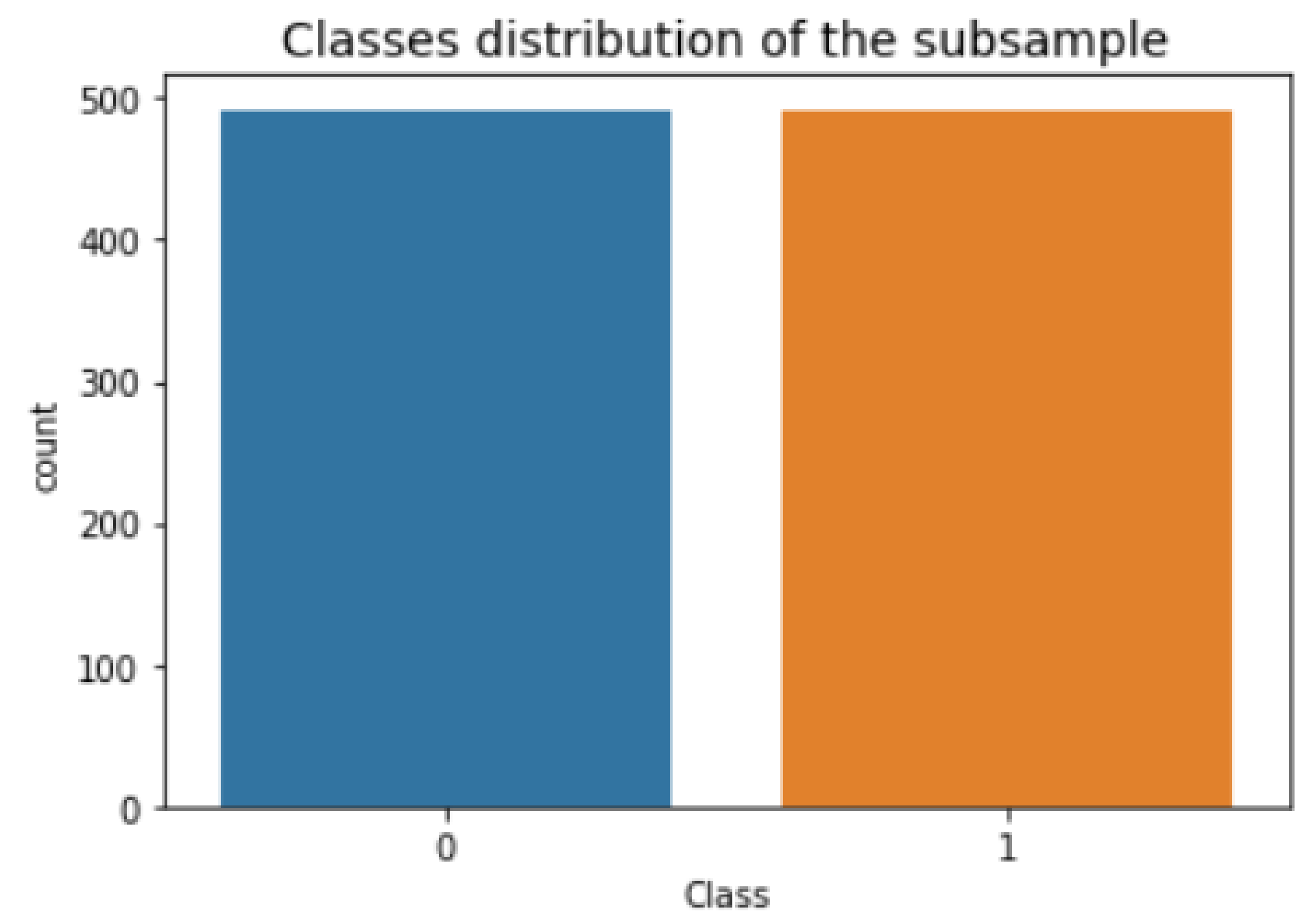
Table 1: Dataset description

Variables	Description
Time	Number of seconds elapsed between this transactions and the first transaction in the dataset
V1 ‖ V28	May be result of a PCA Dimensionality reduction to protect user identities and sensitive features
Amount	Transaction amount
Class	1 for fraudulent transactions, 0 otherwise

Dataset distribution before balancing by undersampling



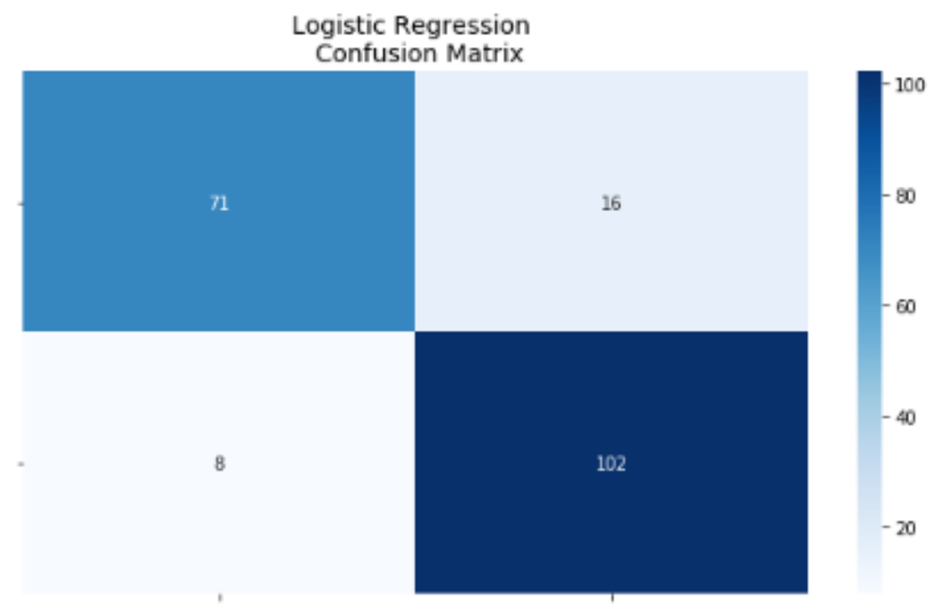
Dataset distribution after balancing by undersampling



Methods

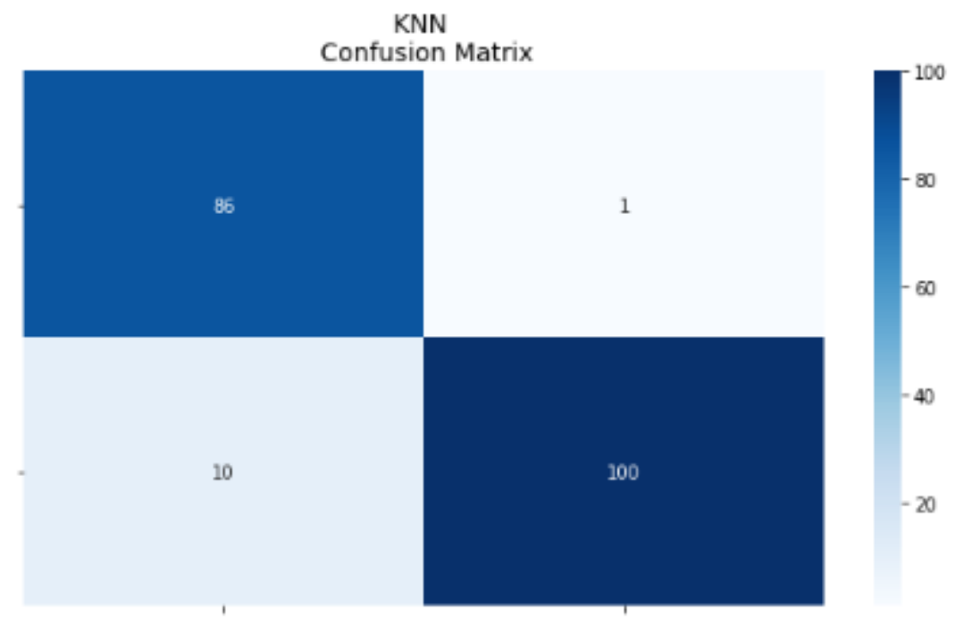
Logistic Regression

$$\ell(\theta)=\sum_{i=1}^{n} y^{(i)} \log h\left(x^{(i)}\right)+\left(1-y^{(i)}\right) \log \left(1-h\left(x^{(i)}\right)\right)$$



K-nearest neighbor

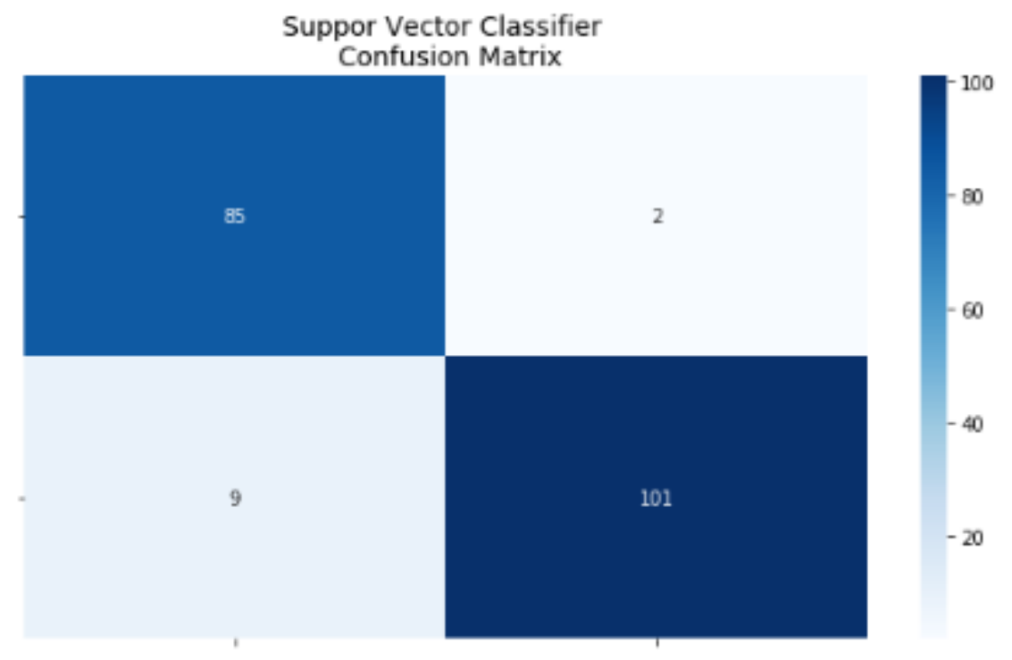
$$d\left(\mathbf{x}_i, \mathbf{x}_l\right)=\sqrt{\left(x_{i1}-x_{l1}\right)^2+\cdots+\left(x_{ip}-x_{lp}\right)^2}$$



Support Vector Classifier

$$h\left(x_i\right)=\begin{cases} +1 & if w \cdot x+b \geq 0 \\ -1 & if w \cdot x+b<0 \end{cases}$$

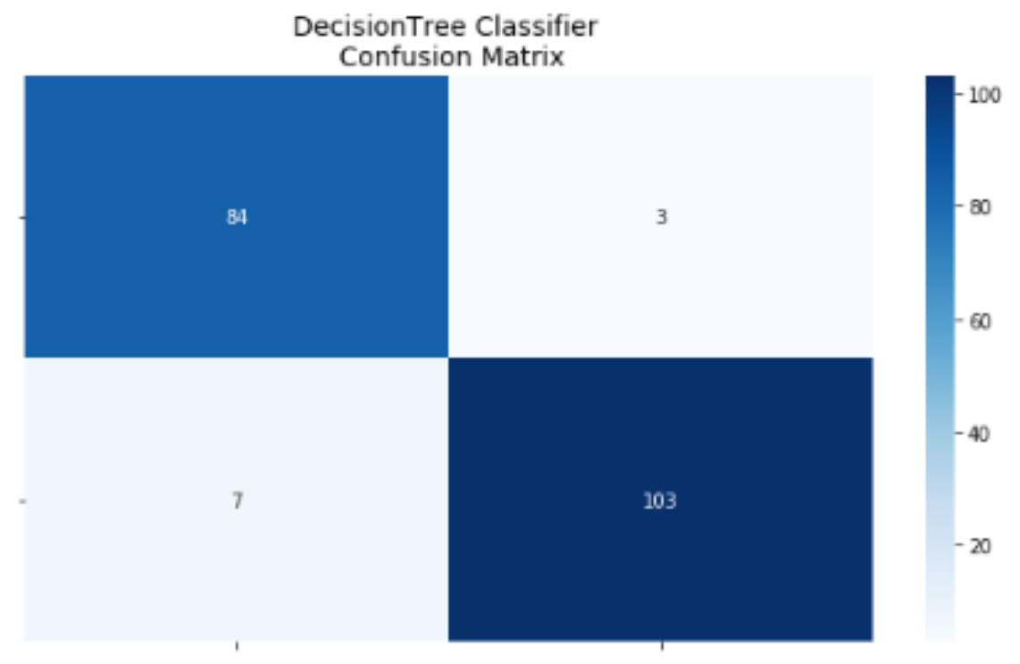
$$\left[\frac{1}{n} \sum_{i=1}^{n} \max \left(0,1-y_i\left(w \cdot x_i-b\right)\right)\right]+\lambda\|w\|^2$$



Decision Tree

$$I_H(t)=-\sum_{i=1}^{c} p(i \mid t) \log_2 p(i \mid t)$$

$$I_G(t)=\sum_{i=1}^{c} p(i \mid t)(1-p(i \mid t))=1-\sum_{i=1}^{c} p(i \mid t)^2$$



Conclusion

There are many metrics to evaluate machine learning algorithms and models. The most popular metric so far is the accuracy, but again, it is not suitable for this case, due to the class imbalance. For example, if we'd always stated that the transaction is not fraudulent, the accuracy would be 99,7\%, which is considered as excellent coefficient for any metric. Also, in this case, if a fraudulent transaction (Actual Positive) is predicted as non-fraudulent (Predicted Negative), the consequence can be very bad for the financial institution. So we mostly care about False Negative values. That's why we should use ROC AUC score as an evaluation metric for our models.

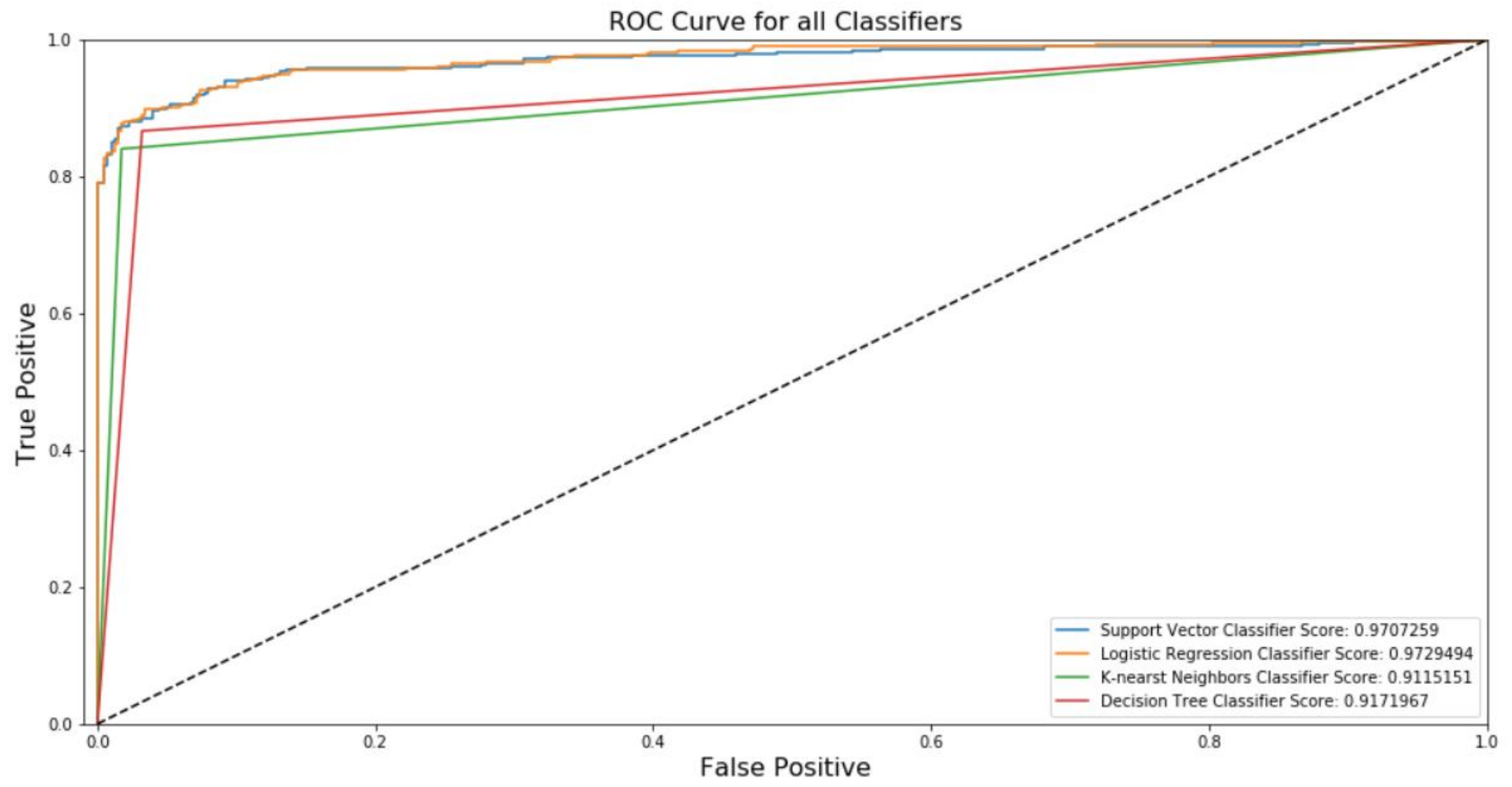


Table 3: ROC AUC scores for the models

Logistic Regression	0.9729493891797556
K Nearest Neighbors	0.9115150927541853
Support Vector Classifier	0.9707258742162757
Decision Tree Classifier	0.91719669058238

Future work

As technology changes, it becomes difficult to track the behavior and pattern of fraudulent transactions. Preventing known and unknown fraud in real-time is not easy but it is feasible. In future might be done oversampling and synthetic data generation methods, I think in case of properly tuned algorithms, they can lead to superior predictive performance in the face of class imbalance. Given that payments fraud is constantly evolving, future areas of work might include the application of reinforcement learning to a real-time data stream. Although fraudsters will never retire, machine learning algorithms can give them a run for their money.