Credit Card fraud Detection using SVM's(Machine Learning)

Asim Shaik, Mohammed K Nasirulla Duvvur

February 17, 2023

1 Introduction

Machine Learning (ML) is increasingly important in credit card fraud detection due to its ability to automatically learn from data and detect patterns that are difficult for human analysts to identify. Some of its abilities includes handling large and complex data-sets, detect fraud in real-time, adapt to new fraud patterns, increase accuracy, and reduce the cost of fraud detection. Due to increasing success of ML algorithms in fraud detection of credit cards transactions, we decided to use SVM's(support vector machines) for detecting the fraudlent transactions using the behaviour of spender.

1.1 Background

Preventing credit card fraud is a critical issue that impacts financial institutions, merchants, and consumers. Credit card fraud detection is a challenging and impactful initiative that uses advanced technologies, such as machine learning and big data analytics, to detect and prevent fraudulent transactions in real time. By working on this project, we will contribute to the overall safety and trust of financial systems, which is crucial for a stable and healthy economy.

Building a credit card fraud detection model is an essential step to protect customer's financial assets and ensure the integrity of financial transactions. Credit card fraud is a constantly evolving problem, and implementing sophisticated algorithms and data analysis techniques can help detect and prevent fraudulent activities in real time. By building a credit card fraud detection model, we will contribute to maintaining the safety and trust of financial systems, providing a sense of fulfillment and accomplishment.

Scikit-learn and TensorFlow are the trending frameworks used in credit card fraud detection models. Platforms like Apache Spark, and Hadoop are data processing platforms that are capable of handling large datasets, can be used and integrated with various machine learning frameworks. Algorithms such as Support Vector Machine(SVM), Logistic Regression, Random Forest, and Deep Neural Networks are being implemented for this model these days.

1.2 Problem Definition

This project is used to determine the fraud in the credit card transactions as fraud is increasing along with the development in technology in today's world, So as the increase in the use of credit card. The main objective of this approach is to detect frauds with very high accuracy as well as in number. Here we have chosen a classification approach using Support Vector Machine (SVM). The SVM has the ability to separate the data in the form of threshold and can handle huge number of transactions with decent result. This approach gives high accuracy to detect the frauds.

1.2.1 Importance of SVM in Credit Card fraud Detection

SVMs plays crucial role in identifying fradulent transactions for many reasons, some of them are listed below:

- **Non-linearity** Credit card fraud detection involves complex patterns in the data that may not be linearly separable. SVMs are effective in handling non-linear relationships between variables and can capture complex patterns in the data.

- **High-dimensional data** Credit card transaction data often contain a large number of variables (features). SVMs can handle high-dimensional data by selecting the most relevant features and creating a hyperplane to separate the fraudulent transactions from the legitimate ones.

- **Imbalanced Data** Credit card fraud is a relatively rare event, which means that the data is often imbalanced with a small number of fraudulent transactions compared to a large number of legitimate ones. SVMs can handle imbalanced data by assigning different weights to the classes and adjusting the classification threshold.

- **Generalization** SVMs have good generalization performance, which means that they can accurately classify new, unseen transactions that have not been seen during the model training process.

1.3 Challenges

The main and biggest challenge here is the process of detection of frauds as well as to determine which ones are non-frauds. Credit card fraud patterns can change over time, which can lead to concept drift, where the relationship between the features and the target variable changes. SVMs can be computationally intensive, especially when dealing with large data-sets.

1.3.1 Using SVMs to Overcome the above challenges

By calculating the accuracy of the classifier, correlation of coefficients and focusing on the most important dimensions, it is possible to explain the majority of the problem, gaining amount of time while avoiding a significant reduction in accuracy. SVMs can be updated with new data to adapt to concept drift

and improve fraud detection accuracy. Techniques such as online learning and kernel approximation can be used to reduce the computation complexity of SVMs and improve their scalability.

1.4 Plan

In this project, we propose a credit card fraud detection model using SVM that accurately identifies fraudulent transactions, minimizes false positives, and provides a scalable and efficient solution for detecting credit card frauds.

We have chosen a classification approach using SVM. The main task here is the process of detection of frauds as well as to determine which ones are non-frauds. By calculating the accuracy of the classifier, correlation of coefficients and focusing on the most important dimensions, it is possible to explain the majority of the problem, gaining amount of time while avoiding a significant reduction in accuracy. we are testing the model by collecting the confusion matrix, then minimizing the errors in the SVM prediction and then deriving the confusion matrix once again. Our criterion will consider precision on the true fraud 4 times more significant than general accuracy for the remainder of the investigation.