

Credit Card Fraud Detection Using Machine Learning

Project Overview

- This project focuses on the creation of a machine learning model for credit card fraud detection for all stakeholders involved in the credit card company looking to improve its fraud detection system.

Business Understanding

- Identify fraud transactions and report them to an analyst while letting normal/non-fraud transactions be automatically processed.

Problem Statement

- High rate of false positives in the current system, which leads to unnecessary inconvenience for customers and additional workload for the fraud investigation team.

Objectives

- General Objective

Detection of credit card fraudulent transactions

- Specific Objective

Classify transaction as either fraud or normal transaction.

Data Understanding

- The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.
- 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. 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'.

Modelling

- We begin our modelling with a baseline logistic regression model followed by an intermediate Random forest model and cap it off with a hyper parameter tuned complex model. The models will be evaluated based on their precision, accuracy, recall and F1 score. The model with the best metrics will be chosen for use and deployment.

Model Evaluation

- Hyper parameter tuning was only performed on the Random Forest model and which had a higher F1 score compared to the logistics regression and it was the one chosen for the modelling.
- Indicating an improvement in the model's overall performance, particularly in terms of balancing precision and recall.
- This suggests that the tuning process and increased model complexity have resulted in a model that is better at distinguishing between legitimate and fraudulent transactions.

Challenges

- Imbalanced Data: Fraudulent transactions are typically rare compared to legitimate ones, leading to imbalanced datasets.
- Evolution of Fraud Patterns: Fraudulent activities are dynamic and can evolve over time. Fraudsters continually adapt their techniques to avoid detection.
- Feature Engineering: Identifying relevant features for fraud detection can be challenging.
- High-Dimensional Data:
- Credit card transaction datasets can be high-dimensional, with numerous features. Managing and processing large amounts of data efficiently can be computationally demanding.
- Cost of False Positives: False positives (incorrectly flagging a legitimate transaction as fraudulent) can inconvenience customers and result in financial losses for the business.

Conclusion

- In conclusion, developing a robust credit card fraud detection model involves navigating through various challenges inherent to the dynamic and complex nature of financial transactions. The imbalanced distribution of fraudulent and legitimate transactions, the evolving tactics of fraudsters, and the need for interpretability and compliance present significant hurdles.