Credit Card Fraud Detection

Abstract

The purpose of this project is to detect the fraudulent transactions made by credit cards by the use of machine learning techniques, to stop fraudsters from the unauthorized usage of customers' accounts. The increase of credit card fraud is growing rapidly worldwide, which is the reason actions should be taken to stop fraudsters. Putting a limitfor those actions would have a positive impact on the customers as their money would berecovered and retrieved back into their accounts and they won't be charged for items or services that were not purchased by them which is the main goal of the project. Detection of the fraudulent transactions will be made by using machine learning techniques Logistic Regression ,Decision tree classifier and Random Forest Classifier those models will be used on a credit card transaction dataset.

Problem Statement

The problem statement chosen for this project is to predict fraudulent credit card transactions with the help of machine learning models.

In this project, we will analyse customer-level data, which has been collected and analysed during a research

Dataset has a total of 2,84,807 transactions out of which 492 are fraudulent. Since the dataset is highly imbalanced, so it needs to be handled before model building.

Problem Overview

For many banks, retaining high profitable customers is the number one business goal. Banking fraud, however, poses a significant threat to this goal for different banks. In terms of substantial financial losses, trust and credibility, this is a concerning issue to both banks and customers alike.

With the rise in digital payment channels, the number of the fraudulent transactions is also increasing with new and different ways.

In the banking industry, credit card fraud detection using machine learning is not just a trend but also a necessity for them to put proactive monitoring and fraud prevention mechanisms in place. Machine learning is helping these institutions to reduce time-consuming manual reviews, costly chargebacks and fees, and denials of legitimate transactions.

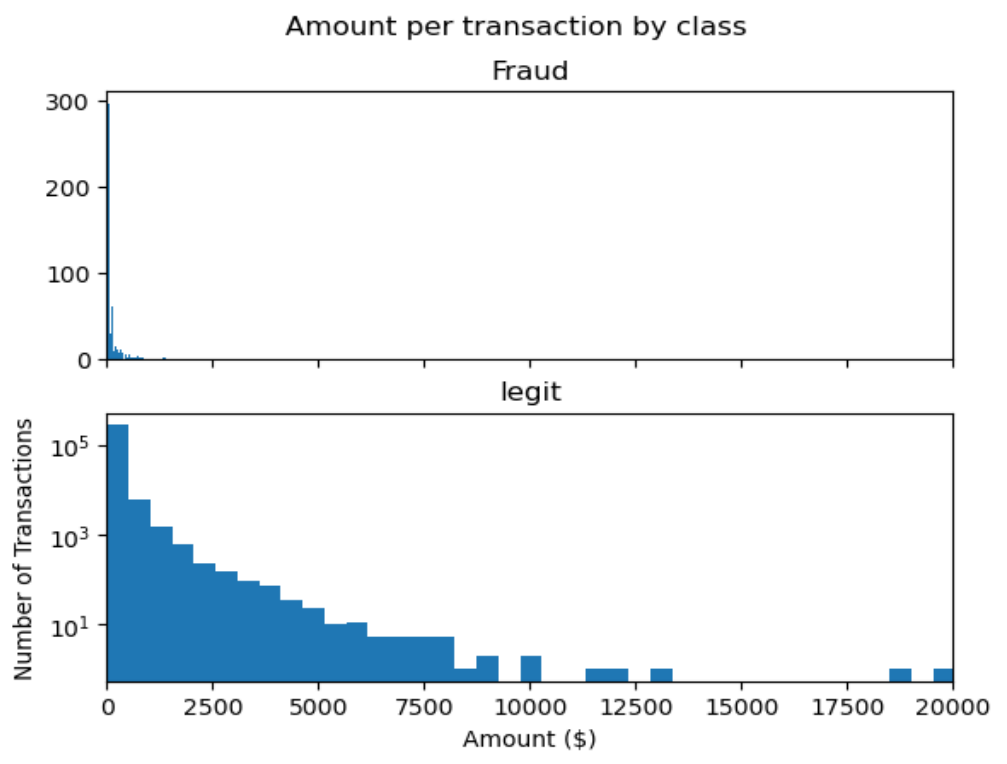
Understanding and Defining Fraud

Credit card fraud is any dishonest act and behaviour to obtain information without the proper authorization from the account holder for financial gain. Among different ways of frauds, Skimming is the most common one, which is the way of duplicating of information located on the magnetic strip of the card. Apart from this , the other ways are :

- Manipulation/alteration of genuine cards
- Creation of counterfeit cards
- Stolen/lost credit cards
- Fraudulent telemarketing

About Data

The dataset includes credit card transactions made by European cardholders over a period of two days in September 2013. Out of 2,84,807 transactions 492 were fraudulent. This dataset is highly unbalanced with the positive class (frauds) accounting for 0.172% of the total transactions. The dataset has also been modified with Principal Component Analysis (PCA) to maintain confidentiality. Apart from 'time' and 'amount' all the other features (V1,V2,V3,.... upto V28) are the principal components obtained using PCA. The feature 'time' contains the seconds elapsed between the first transaction in the data set and subsequent transactions. The feature 'amount' is the transaction amount. The feature 'class' represents class labelling, and it takes the value 1 in cases of fraud and 0 in others.



Project Pipeline

The project pipeline can be briefly summarized in the following four steps:

- **Data Understanding**

  Here we need to load the data and understand the features present in it. This would help us choose the features that we will need for your final model.

- **Exploratory data analytics (EDA)**

  Normally in this step, we need to perform univariete and bivariete analyses of the data followed by feature transformations, if necessary. For the current dataset, because Gaussian variables are used we do not need to perform Z scaling. However, you can check if there is any skewness in the data and try to mitigate it, as it might cause problems during the model-building phase

- **Model building / Hyperparameter tuning**

  This is the final step at which we can try different models and fine-tune their hyperparameters until we get the desired level of performance on the given dataset. We should try to see if we get a better model by the various techniques

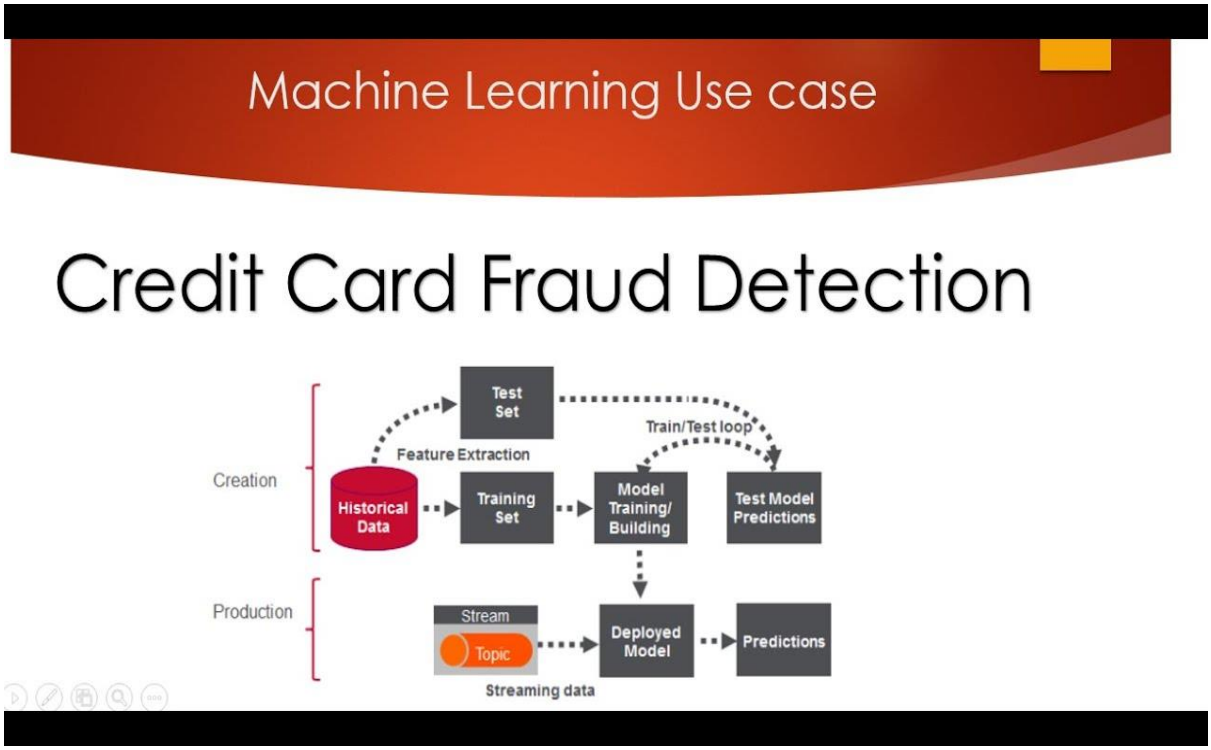
- **Model Evaluation**

  We need to evaluate the models using appropriate evaluation metrics. Note that since the data is imbalanced it is more important to identify which are fraudulent transactions accurately than the non-fraudulent transactions. We need to choose an appropriate evaluation metrics, which reflects the business goal

Main Steps

- Load the data.
- Exploratory data analysis.
- Clean the data.
- Deal with unbalanced data using undersampling techniques.
- Train models and find best-suited model for the data.
- Hyper parameter tuning using GridSearchCV
- Model Deployment using Streamlit

Network URL (http://192.168.29.147:8501)



Future work

This research project was more successful in completing the training of the credit card fraud-detection model, but there are many areas for improvement in future work.

- After completing the training of the optimal model, we can try to combine two or more classifiers with training and evaluating the detection performance. It can provide more possibilities.
- Use deep learning similar to neural networks. Deep learning is different from machine learning in that it is unsupervised learning. It uses unstructured or unlabeled data and does not require the developer to tell it what to look for in the data.
- In the data source, as we are using someone else's original dataset possibly. At a later stage, if we then extract more data from the network. The amount of data is gradually increasing which may be useful for training. The final predictive performance of the model is also improved. In other words, the detection accuracy is enhanced by a large data set.
- The classifier of machine learning is tested for different types of attacks. And analyse its performance under attack. And then use this make appropriate measures to improve its security.
- Using the existing mature and effective classification methods, we can enhance credit card detection—fraud detection performance. Then we use the current bank's credit card system to evaluate whether this model is accurate, as a way to test the real credit card fraud detection.

Conclusion

In conclusion, the main objective of this project was to find the most suited model in credit card fraud detection in terms of the machine learning techniques chosen for the project, and it was met by building the three models and finding the accuracies of them all, the bestmodel in terms of accuracies is Random Forest classifier which scored 91.37%. I believe that using the model will help in decreasing the amount of credit card fraud and increase the customers satisfaction as it will provide them with better experience in addition to feeling secure.