Fraud Detection of Transactions

Submitted by Thanima Firoz

Background information on the dataset and its relevance

Credit card fraud is a wide-ranging term for theft and fraud committed using or involving a payment card, such as a credit card or debit card, as a fraudulent source of funds in a transaction. The purpose may be to obtain goods without paying or to obtain unauthorized funds from an account. Credit card fraud is also an adjunct to identity theft. Although incidences of credit card fraud are limited to about 0.1% of all card transactions, they have resulted in huge financial losses as the fraudulent transactions have been large value transactions. It is important that credit card companies are able to recognize fraudulent credit card transactions so that customers are not charged for items that they did not purchase.

Overview of the Dataset

The dataset contains transactions made by credit cards in September 2013 by European cardholders, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, in which the proportion of the fraud transactions is only 0.172%. It contains 30 independent variables including 2 numerical features, ‘Time’ and ‘Amount’, and 28 principal components obtained with PCA. Feature 'Time' is a timestamp, which means the seconds elapsed between each transaction and the first transaction in the dataset. The feature 'Amount' is the transaction Amount, and this feature can be used for example-dependent cost-sensitive learning. Feature class is the response variable and it takes value 1 for fraud transactions and 0 for normal ones. Finally, 29 independent variables and one dependent variable are included in our research.

Data: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

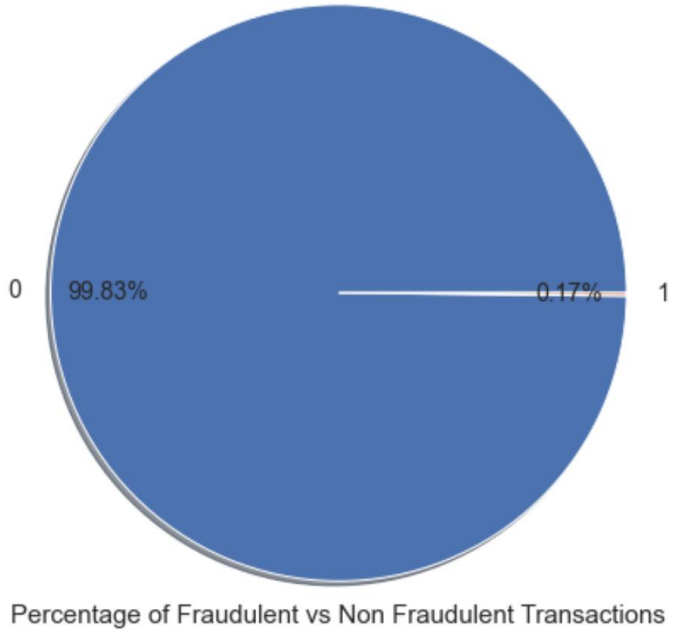
Actionable Insights derived from the dataset

Insights #1

Imbalanced Class Problem.

Fraudulent transactions account for a small percentage of total transactions.

Only 0.17% of the total transcations consist of fraudulent transactions.This suggests that fraudulent transactions are not frequently happening and that most credit card transactions are non-fraudulent transactions.

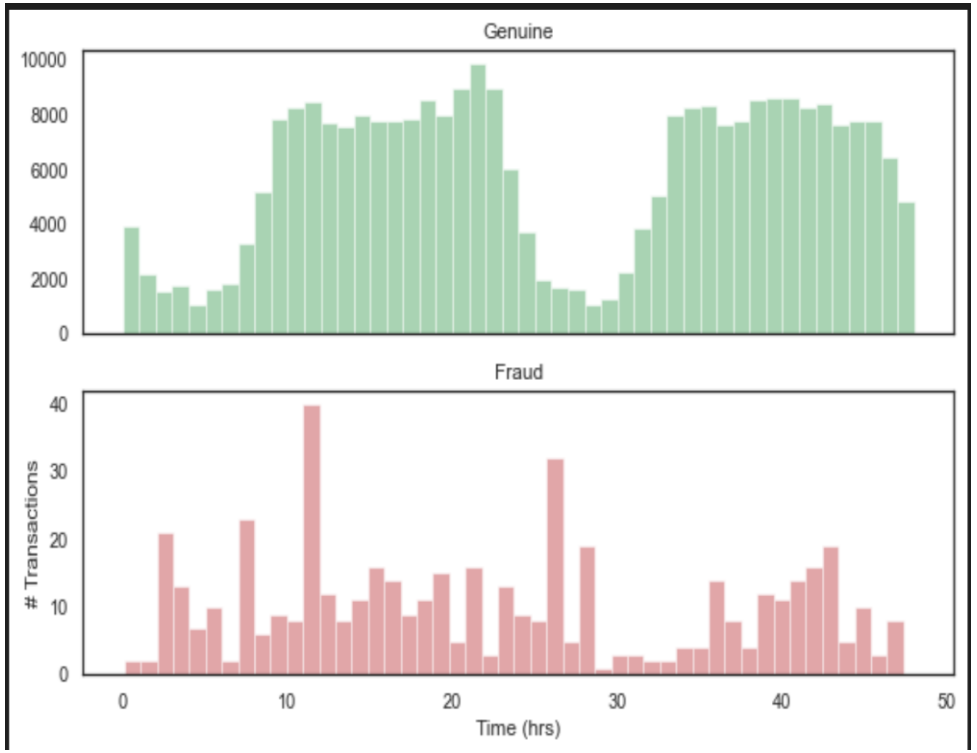


Percentage of Fraudulent vs Non Fraudulent Transactions

Insights # 2

Class Distribution by Time

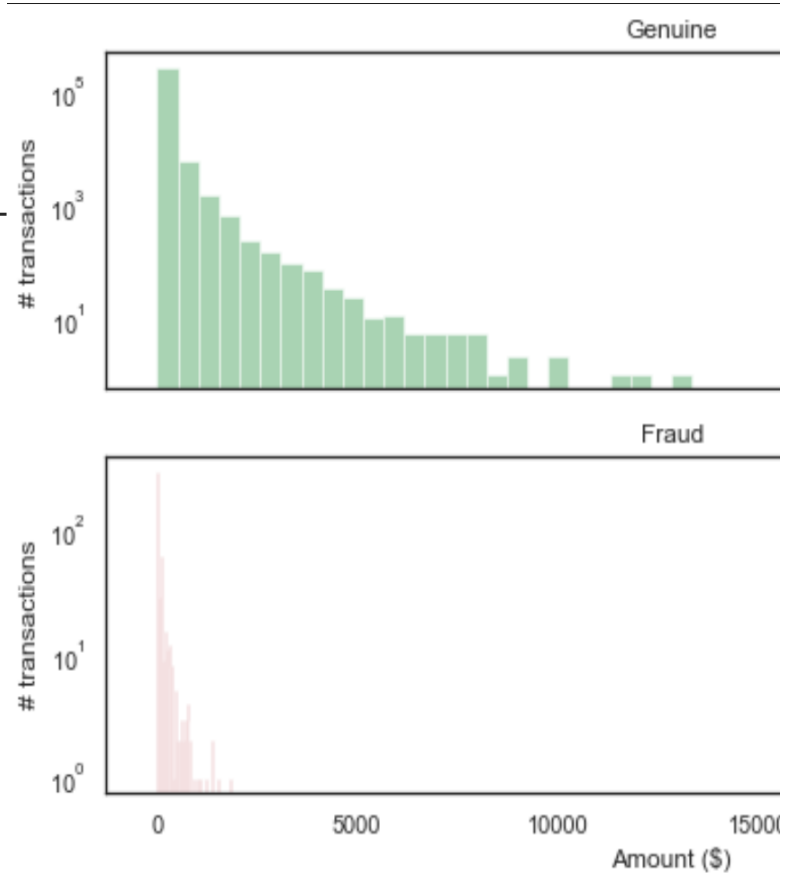
- There is a clear pattern of fraud transactions over 24 hours
- The percentage of fraud transactions is higher during the late night and early morning hours
- Fraudsters might be taking advantage of low transaction volume during these hours



Insights #3

Distribution of Transaction amount

- Fraudulent transactions are mostly skewed towards lower amount whereas non-fraudulent transactions are spreaded across low to high amount.
- Most fraudulent transactions are of small amounts. The maximum amount involved with fraudulent transactions is less than 10K whereas that in non-fraudulent transactions is above 10K.
- We can infer that 'Amount' column has a significant prediction power in this dataset.
- The maximum amount involved in fraudulent transaction is 2125.00 whereas that in non-fraudulent is 25691.00.The "Amount" feature has a wide range of values and needs to be standardized before it can be used in the model.

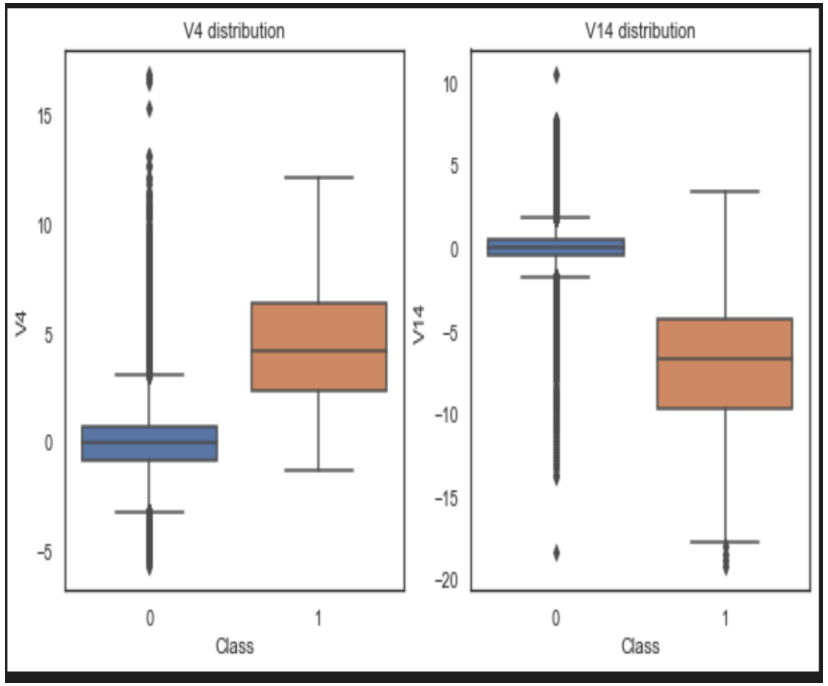


Insights # 4

Transactional Patterns

By analyzing transaction patterns, it is possible to identify fraudulent activities. For instance, if there is a sudden increase in transactions or transactions that are significantly larger than usual, it could be an indication of fraud. By identifying such patterns, companies can investigate further and take appropriate action.

Fraudulent transactions have different patterns from non-fraudulent transactions: A comparison of the distribution of anonymized features between fraudulent and non-fraudulent transactions reveals that the two types of transactions have different patterns. For example, fraudulent transactions tend to have lower values of V4 and V14, and higher values of V2 and V11.



Insights # 5

Application of ML Algorithms for predicting fraudulent transactions.

We can utilize the Machine Learning Algorithms with an accuracy of above 90 % for predicting fraud transactions.Many machine learning methods, including logistic regression, random forests, and neural networks, can identify fraudulent transactions with high accuracy. These models enable credit card firms to automate fraud detection and lessen the burden of manual inspection.We can also improve on this accuracy by increasing the sample size or use deep learning algorithms however at the cost of computational expense.

	precision	recall	f1-score	support
0	0.94	0.95	0.94	98
1	0.95	0.94	0.94	99
accuracy			0.94	197
macro avg	0.94	0.94	0.94	197
weighted avg	0.94	0.94	0.94	197

```
Classifiers:  LogisiticRegression has a training score of 94.0 % accuracy score
Classifiers:  KNearest has a training score of 93.0 % accuracy score
Classifiers:  Support Vector Classifier has a training score of 93.0 % accuracy score
Classifiers:  DecisionTreeClassifier has a training score of 89.0 % accuracy score
Classifiers:  Gradient Boosting Classifier has a training score of 94.0 % accuracy score
```

Summary

- Imbalanced Class Problem
- Class Distribution by Time
- Distribution of Transaction amount
- Transactional Patterns
- Application of ML Algorithms for predicting fraudulent transactions.

Thank You