Practical 2: IEEE-CIS Fraud Detection

Team: It's My Party and I'll Sing if I Want To

Executive Summary

Methods to detect fraud have helped save consumers million dollars per year in fraud prevention but even still, in 2019 £1.2 billion was successfully stolen through fraud in the UK (Worobec, 2020). One place where fraudulent activity can occur is through online transactions, which more and more people are using every day. Fraud detection in these online transactions relies on machine learning techniques that use information about the transaction and the identity of the individuals involved to predict if it is likely to be fraud. We used this data and various forms of a boosting model (specifically XGBoost and LightGBM) to predict whether a transaction is likely to be fraud and found that XGBoost performed the best. The final model we decided on had an accuracy that suggests we could now successfully prevent over 90% of fraud cases from happening. Information about the time of transaction, the product, the card used and information about the purchaser and recipient have proved to be particularly useful in identifying fraudulent activity. We especially found that hour of day, product code and email domain of the purchaser were important information that needs to be included when predicting if the transaction is fraud. The proposed model should be used by online services to identify and prevent fraudulent activity before the consumer can lose money and therefore reduce the amount of successful fraud schemes carried out per year.

Proposed Model and Methodology

EDA: The exploratory data analysis led to many valuable insights into the nature of fraudulent activities. Due to the size of the dataset and features, there are far too many observations to list them all, hence only the most informative will be mentioned. For instance, of the transactions provided only 3.5% were fraudulent. This has potential to lead to issues while modelling due to the large imbalance. Additionally, as an exception, the missing values proved to be valuable and provided a hint as to the high correlation between the V columns, allowing the data to be slimmed down. One can also see that Product C clearly has the highest percentage of fraudulent activity and therefore is a valuable, indicative variable. Furthermore, through manipulation of the feature “TransactionDT”, it was shown that the level of card fraud is related to time of day as well as day of the week. Fraudulent activity is highest over the weekend, dropping down towards the middle of the week and is lowest in the morning and increases towards the evening.

Model Fitting: Using this information to feature engineer the training data we were able to create a more optimized and informative dataset to base our initial model on. After investigating a few options such as randomforest and lightgbm models it was confirmed that an xgboost model would in fact be most suitable. Clearly a boosting model was chosen over a simple randomforest. While both lightgbm and xgboost were very effective in producing a high accuracy and lightgbm has a much higher training speed it also was a lot more prone to overfitting, which finally led to the choice of the xgboost model. Using randomized search this model was then tuned. Using these parameters, the final model was produced which demonstrated 94% accuracy.

Business Case

With the **gradual disappearance of cash**, credit and debit card transactions play an increasingly important role in our everyday lives. The trend towards increasing card transactions was accelerated significantly by the corona pandemic. For example, the **number of all credit cards worldwide** has increased by **10.5%, from 22.8 billion (2019) to 25.2 billion (2021)**. It is predicted that this trend will intensify, and the number of all cards will increase by **another 16% by 2023** (Statista 2021). By **2021**, **98% of all Britons** were

already using a debit or credit card regularly, processing a total of **2 billion payments** (UK Finance 2022). However, as transactions increase, so does the incidence of credit card fraud. The Nilson Report, which examines fraudulent transactions, predicts that the volume of fraudulent transactions will be around **$33 billion in 2023** and will grow to at least **$38.5 billion by 2027** (Nilson Report 2022). In the Consumer Credit Act of 1974, it was regulated in the UK that the bank or credit card provider and not the consumer is liable. For this reason, it is more important than ever for credit and debit card providers to **directly detect and prevent fraudulent transactions**.

Our company, ***St Andrews Analytics***, offers bespoke credit card fraud identification software that can detect a fraudulent transaction with **95% accuracy**. Our statistical market analysis of over **500,000 credit card transactions** identified that approximately **3.5% of all transactions are fraudulent**. For example, we were able to analyse with the help of our model that credit card fraud is more likely to happen at night, 90% of the amounts are smaller than €335, and the distance from the home address is unnaturally far. Using **machine learning**, our **excellent fraud detection predictive model** can analyse **over 300 variables** to avert a potential fraud attempt. Many major credit card issuers already benefit from our **accurate and efficient software**.

For this purpose, we are happy to **advise you first** and adapt our model to your wishes. In the second step, we **implement the model** and **train your employees** in the use of the software. **Finally**, you only have to pay a **monthly licence fee** to use the software. Of course, we will also be at your disposal in the future for any questions you may have and will continue to develop our software.

daily transactions	100.000
fraud rate	3.5 %
mean transaction amount	150 $
accuracy of our modell	95 %

Potentially fraudulent transactions daily	$3.5\ \% * 100\,000 = \mathbf{3,500}$
Potential loss daily	$3\,500 * 150\,€ = \mathbf{535,00}\ \$$
Possible loss saved daily	$0.95 * 525\,000\,€ \approx \mathbf{500,000}\ \$$
monthly loss saved	$30 * 498\,750\,€ \approx \mathbf{15,000,000}\ \$$

Consulting	$= \mathbf{1,500,000}\ \$$
Implementation fee	$= \mathbf{1,200,000}\ \$$
Monthly license fee	$= \mathbf{1,400,000}\ \$$

Savings in the first month	$\approx \mathbf{10,900,000}\ \$$
Savings in the following months	$\approx \mathbf{13,600,000}\ \$$

References

Katy Worobec (2020) : *Fraud – The Facts 2020,* UK Finance **Fraud-The-Facts-2020-FINAL-ONLINE-11-June.pdf (ukfinance.org.uk)**
https://www.ukfinance.org.uk/sites/default/files/uploads/Data%20(XLS%20and%20PDF)/Card-Spending-Update-Dec2021.pdf

https://nilsonreport.com/publication_newsletter_archive_issue.php?issue=1209

https://www.statista.com/statistics/1080756/number-payment-cards-in-circulation-worldwide/