Springboard

Craft a Story from a Dataset

Applying the Storytelling Skills

Presented by: Marwan Ali

Mentor: Tony Paek

Contents of the Report

01

Report Purpose

Reduce fraud (make it at the minimum) to maintain a healthy cashflow position

02

Methodology

- Research design
- Data Source
- Data manipulation
- Analytic goals and methods

03

Analysis Tools

- SMOTE & XGBOOST
- Precision-recall and Receiver operating characteristic

04

Performance results

- Forecasted business annual cost
- Forecasted Vs previous results comparison and savings amount.

Report Purpose

Separate fraudulent transactions from those that are legitimate by training different models to do so.

Reduce fraud

Minimize online scams and reduce credit card skimming

Healthy cashflow

Increase brand trust and enhance reputation

Streamline customer experience

Methodology

Research Design

- Advantages of descriptive research
- Disadvantages (Jha, S., Westland J. C., 2013).

Data Source

- Alternative resource (Kaggle 2019).
- Observational method

Data Manipulation

- Missing data
- Sampling strategies

Analytic Methods

- Supervised statistical methods
- Unsupervised statistical methods

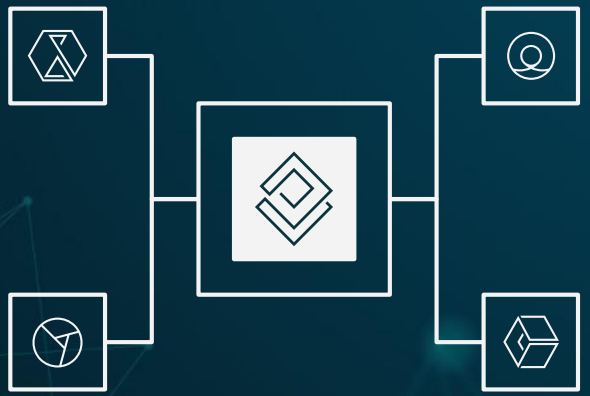
Analysis Goals

Explore features and engineering new features

Data Processing

Sampling Techniques

To make sure that we are using an effective sampling approach



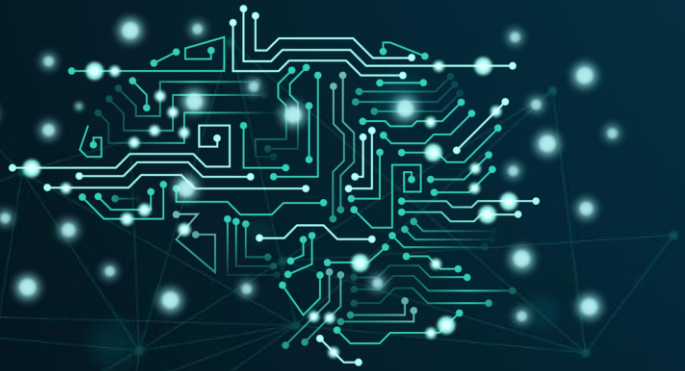
Build an effective XGBoost classifier and select an optimal cutoff

Models Building

Business Impact

Estimate the business impact and potential savings

Analysis Tools



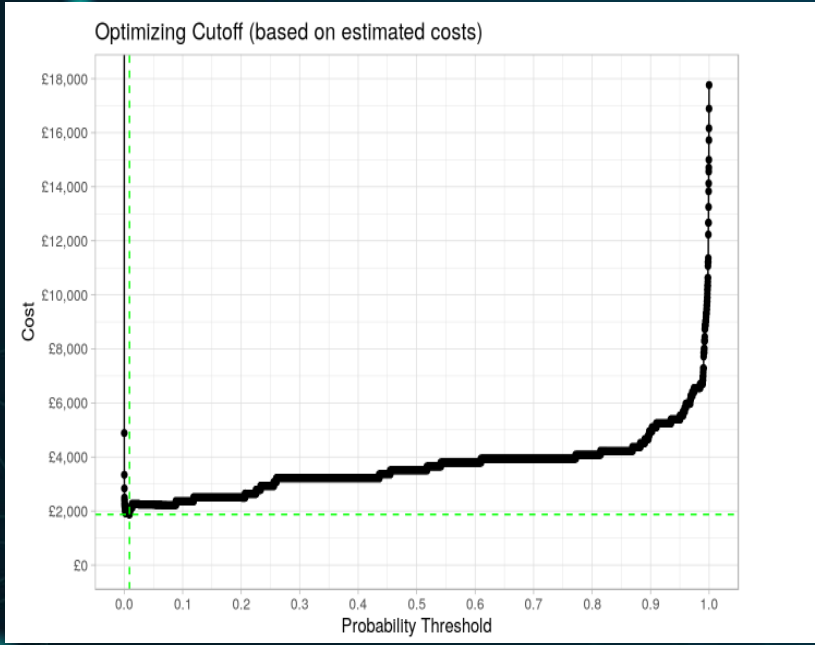
SMOTE — Synthetic minority over-sampling techniques

XGBoost — Extreme gradient boosting

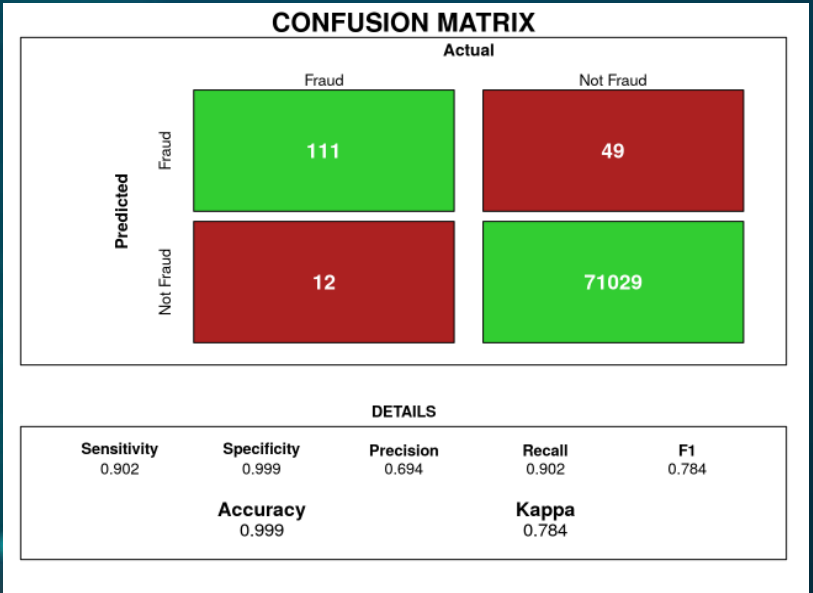
PR — Precision-recall

ROC — Receiver operating characteristic

Optimizing Cutoff (based on estimated costs)



- The goal is to minimize overall costs to the business, (i.e. minimize the following function): Cost = count(FP) . cost(FP) + count(FN) . cost(FN))
- The selected model has found the best balance of these two errors



Final Confusion Matrix/ Half Day

Business values were translated to one full day in the final calculations (Clippinger, Dorinda, and Shirley Kuiper., 2016).

Quantifying Business Value

Full day = **246** fraud transactions

$195 per transaction

Fraud detected

No fraud detection system cost

Previous annual cost

$195 * 246 * 365 = **$17,509,050**

FN & FP errors' effects

24 False negatives per day
98 False positives per day
Business cost per day = **$5,030**

Predicted new annual cost

Annual cost = $5,030 * 365 = **$1,835,950**

Fraud-handling Cost Savings

90%

$17,509,050 - $1,835,950 = **$15,673,100**

References

Clippinger, Dorinda, and Shirley Kuiper. (Ed.). (2016). *Planning and Organizing Business Reports: Written, Oral, and Research-Based* (1st ed.). Business Expert Press, LLC. http://ebookcentral.proquest.com/lib/ncent-ebooks/detail.action?docID=4586439

Jha, S., Westland J. C. (2013). A Descriptive Study of Credit Card Fraud Pattern. *Global Business Review*. https://doi.org/10.1177/0972150913494713

Kaggle (2019). Fraud Detection - SMOTE, XGBoost & Business Impact. [dataset]. https://www.kaggle.com/lmorgan95/fraud-detection-smote-xgboost-business-impact