Anilkumar Panda

Made for **Adyen**

Version 1.0

Risk Scores Suggestion

Identify the risk scores for least no.of fraudulent chargeback

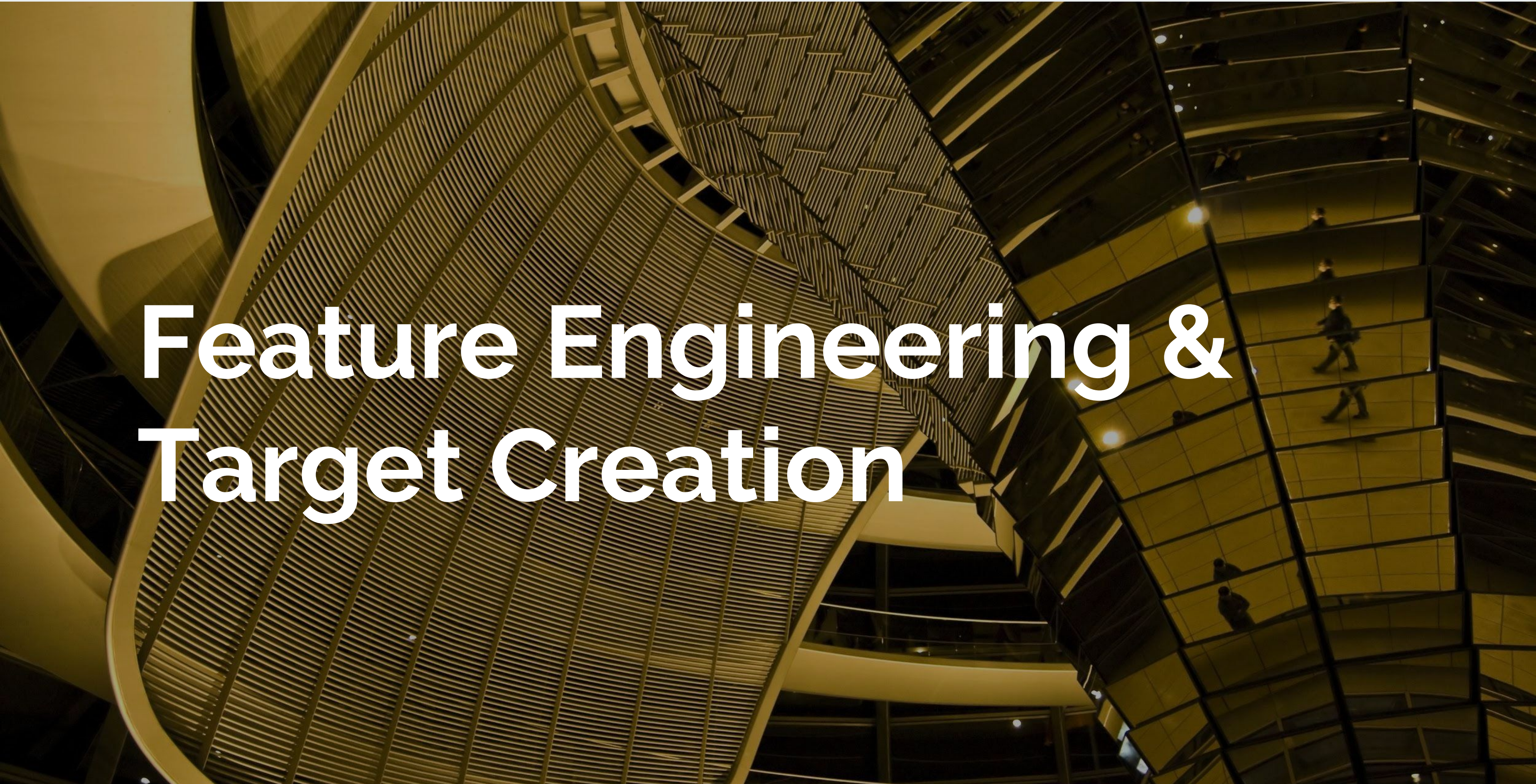
TOC

Problem Statement

Adyen is looking to help a large Food ordering merchant optimize its risk settings. Adyen has a Rule Based Risk Engine with 78 different rules. A merchant can assign a score to each of the rules, if the combined score of the rules passes a threshold of 100, Adyen's risk engine will automatically block the transaction.The merchant has asked us if we could **determine the optimal risk score** for them.

- Fraudulent chargebacks should be less than 1 pc (or as low as possible)

Fig : POS Sales



Feature Engineering & Target Creation

Feature Engineering & Target Creation

Target :

Classification Problem : (Fraud = 1, Not Fraud = 0)

1. Highly Imbalanced dataset.

Features :

1. Rules from 1-78 are only considered.
2. More features can be created too .

   Eg (recency,amount of transaction ,rating of the bank for fraud usage, origin country etc) . However since we want to assign weights to the rules, not considering other features.

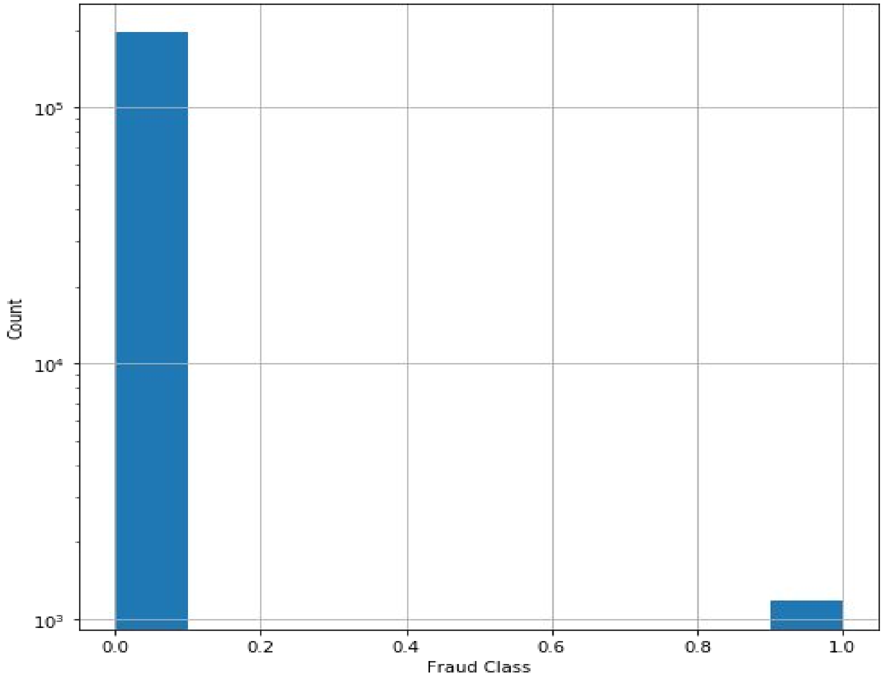
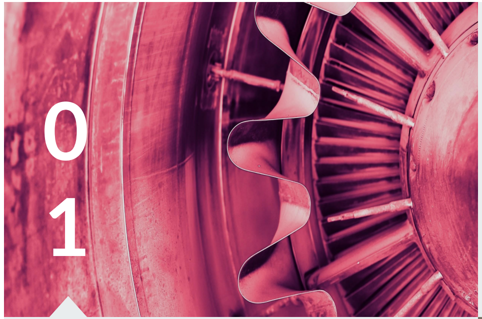


Fig : Fraud Histogram (Log Transformed)

Experiment Design

Experiment Design



Data Cleaning
Target & Feature Creation

In this step, we read the data ,create targets and features .

Handle Data Imbalance

Since we are dealing with highly imbalanced datasets,we try various sampling techniques .

Classification

On, the sampled datasets , we try multiple classification techniques namely (Logistic Regression,Bagging and Boosting Trees) . We use scoring metrics like AUC,Precision & Recall,Cohen-Kappa.

Interpretation

The classifiers give us a sense feature importance. These feature importance can then be converted to weights to be given to the rules.

Getting Ready- Common Settings

Training & Testing datasets

01

Test data ratio : 0.3

Model Evaluation

02

High Precision on Majority Class - 0/High Recall on Minority Class - 1.

AUC

Cohen-Kappa Score

Imbalanced Class

- **Modify class weights**

  01

  Assign weights to class
- **Sampling Techniques**

  02

  Random Under & Over Sampling
  NearMiss
  SMOTE
- **Ensemble Methods**

  03

  Bagging Classifier
  SMOTE+EEN

Classification

Decision Tree

01

Though Decision tree classifiers suffer from the disadvantages such as growing complex and still unable to generalise the data well.

It is simple to understand and trees can be visualised .

Test AUC / Cohen-Kappa score : 0.73/ 0.049

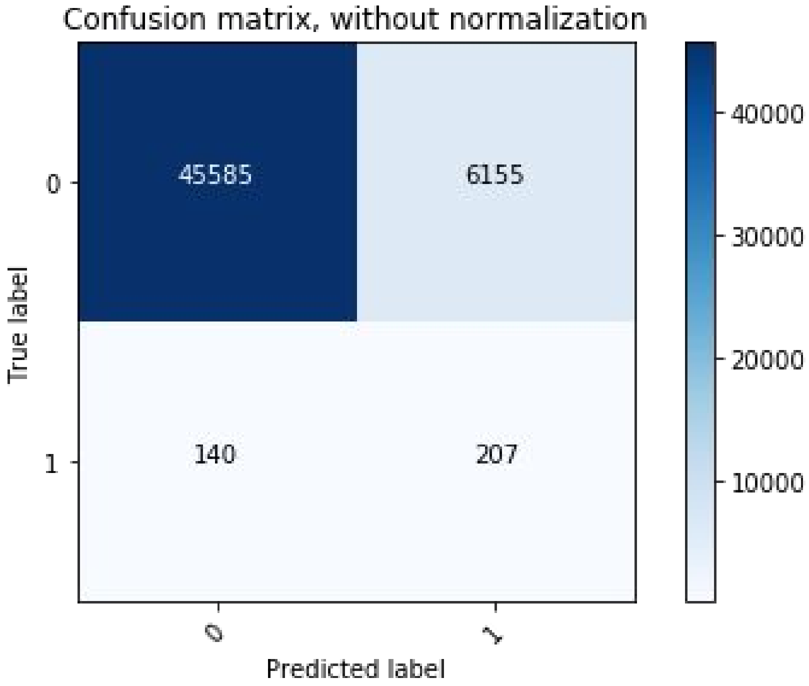


Fig : Classification Report

Classification

Random Forest & Extra Trees Classifier

02

One of the most robust set of classifiers, inherently reduces overfitting .

Provides a list of feature importance, which can help in identifying entity attributes.

Random Forest :
AUC / Cohen-Kappa : 0.72/0.049

Extra Trees Classifier :
AUC / Cohen-Kappa : 0.73/0.05

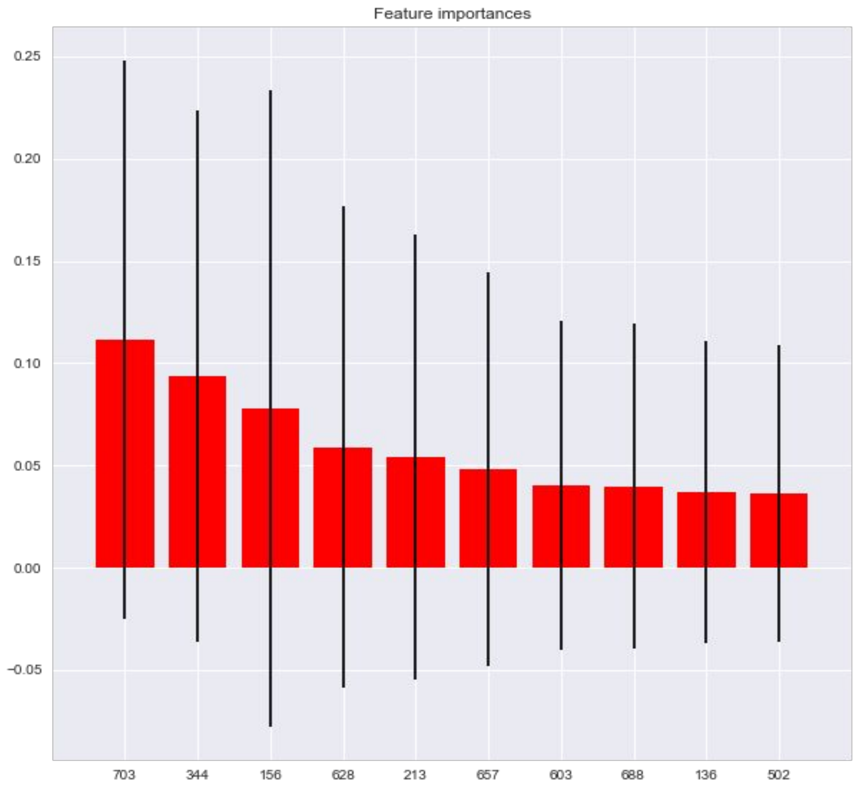


Fig : Extra Trees Classifier Feature Importance

Results

Performance Analysis

A run is considered good enough for usage if :

- Train & Test ROC-AUC is >= 0.72
- Training AUC and Testing AUC <= 5 percentage points
- Cohen - Kappa is > = 0.04
- Recall on class 1 > = 0.6

Sampling Technique	Class Weights					
ML Algorithm	LR	DT	RF	ET	GBT	ABT
Metrics (Test Set)						
Precision on 0 class	1	1	1	1	0.99	0.99
Recall on 1 class	0.61	0.62	0.58	0.62	0.02	0.01
F1 Score	0.06	0.07	0.06	0.07	0.03	0.03
AUC	0.73	0.74	0.73	0.73	0.52	0.52
Cohen-Kappa score	0.04	0.04	0.05	0.05	0.02	0.02

Complete Result Set : https://docs.google.com/spreadsheets/d/1OSu_MyEJ_d_kjv1hw1LVswEhpdcXLttGpWLzbXKC5-4/edit?usp=sharing

Most Important Rules

Rule No	Importance(%)
Rule 49	19
Rule 55	8
Rule 63	14
Rule 74	11

Complete Result Set : https://docs.google.com/spreadsheets/d/1OSu_MyEJ_d_kjv1hw1LVswEhpdcXLttGpWLzbXKC5-4/edit?usp=sharing

Future Work

- ## Better Features & Grid Search

01

Create more features that can detect fraudulent transactions .
Grid search for hyper parameter tuning .
SMOTE with KBest features

- ## Anomaly Detection

02

Treat fraud cases as anomaly .

- ## Modify Threshold

03

Modify default threshold probability of 0.5 .

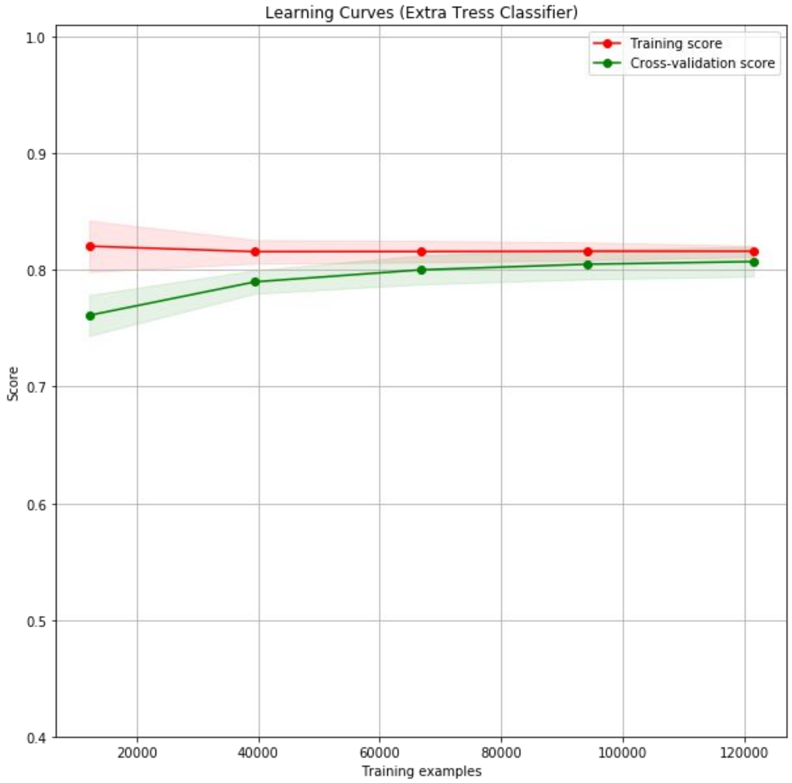


Fig : Extra Trees Classifier Learning Curve

Thank you.