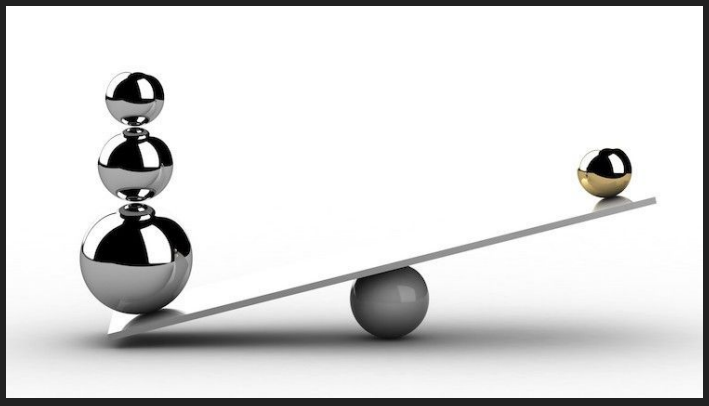
Detecting Attempted Credit Card Fraud

By George Bennett

Image source (https://en.wikipedia.org/wiki/Credit_card)

Problem Statement

- Classification
- Highly Imbalanced
- Flag Fraud
- Don’t Flag Too Much



- This is a classification problem
- The dataset was highly imbalanced. %0.17 Fraud
- My goal is to be able to flag as many incoming transactions as possible
- I must avoid having my models flag too many transactions as that would make the model impractical

Image source
https://towardsdatascience.com/dealing-with-class-imbalanced-datasets-for-classification-2cc6fad99fd9

Business Value

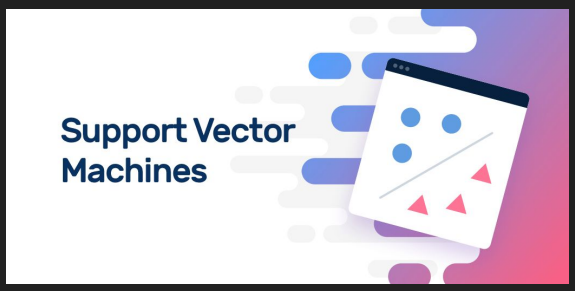
- Prevent As Much Fraud As Possible

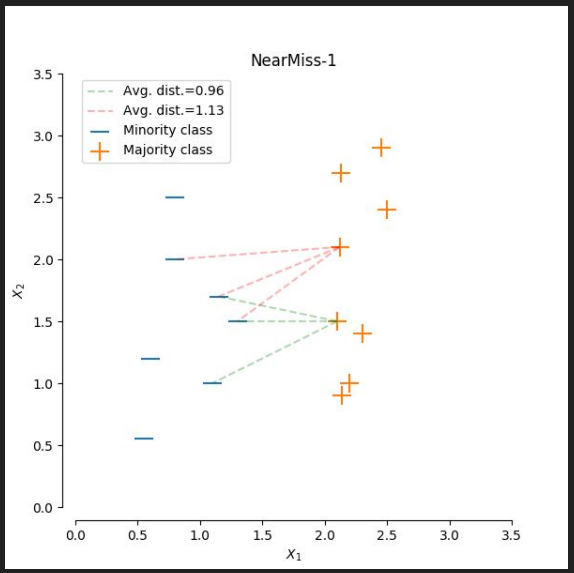
- My model will be able to give lenders added protection against credit card fraud
- It will also help protect the credit card owners from fraud

Image source
https://datafloq.com/read/will-analytics-technology-end-credit-card-fraud/2121

Methodology

- Transform the data different ways
- NearMiss
- Several machine learning models



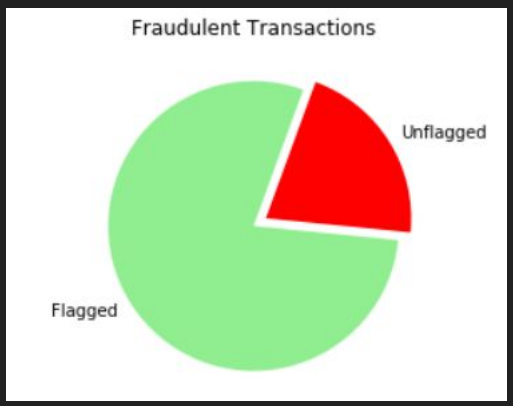


- My methodology was to transform the data several ways using different techniques
- I used NearMiss to resample the training data to make it balanced
- I used several machine learning algorithms

Image source left
Image source right
https://imbalanced-learn.readthedocs.io/en/stable/under_sampling.html

Findings

- Model flags %79 of fraudulent transactions
- Less than %1 of transactions are falsely flagged



- The model had a recall of %79 and a precision of %63

Recommendations

- Decline transactions

- My recommendations for the company would include creating software to process transactions based on this model
- If the model flags the transaction then decline that transaction

Image source https://www.helcim.com/article/understanding-credit-card-declines/

Recommendations

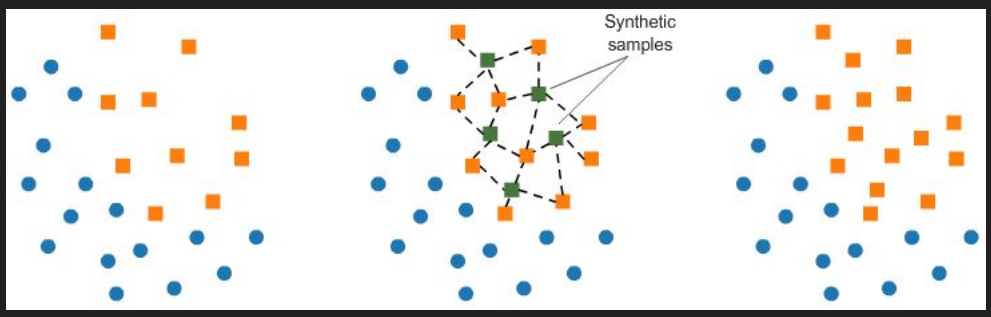
- Make contact with card owner

- Furthermore if a transaction is flagged than call the credit card owner to try and resolve the issue
- And warn them that there is suspicious activity

Image source
http://www.theegreetingsportal.com/things-to-remember-in-choosing-a-conference-call-service-provider/

Future Work

- SMOTE

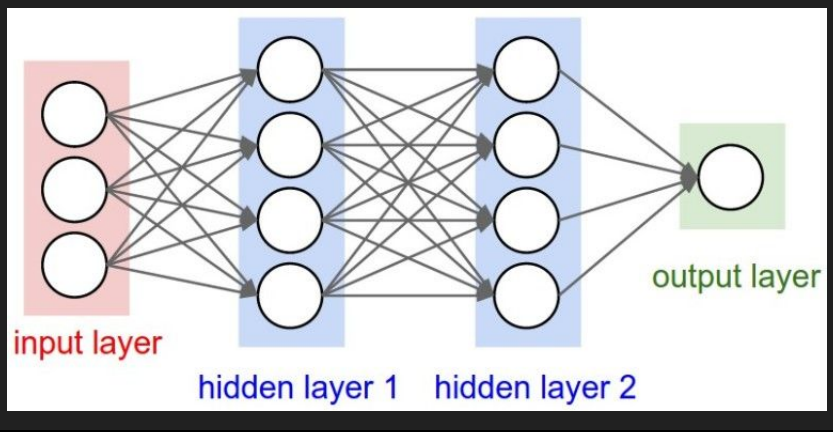


- Given more time and/or computational power I could use synthetic minority oversampling technique to see if it yields better results

Image source
https://towardsdatascience.com/the-5-most-useful-techniques-to-handle-imbalanced-datasets-6cdba096d55a

Future Work

- Additional machine learning algorithms
- Neural Network

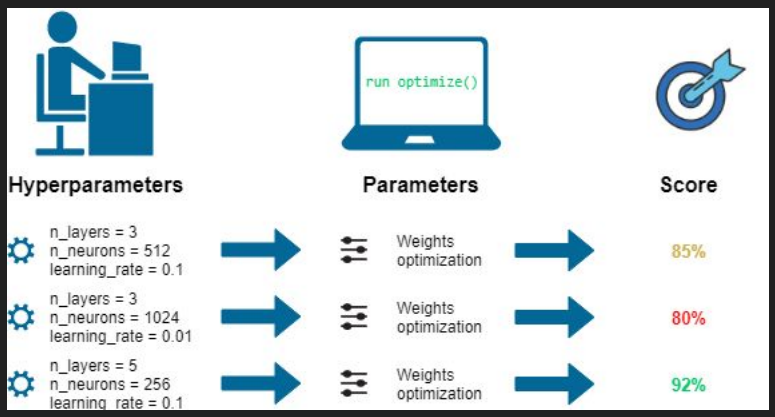


- Given more time I could construct more machine learning models
- A neural network may prove to be more effective if I use SMOTE or have access to more data

Image source
https://www.pyimagesearch.com/2016/09/26/a-simple-neural-network-with-python-and-keras/

Future Work

- More expansive parameter searches



- Along with more machine learning models would be more refined parameter searches

Image source
https://deepai.org/machine-learning-glossary-and-terms/hyperparameter

Thank you