MOD 5 Final Project
IEEE CIS Fraud Detection Project

Hyungjun Kang

Table of Contents

kaggle

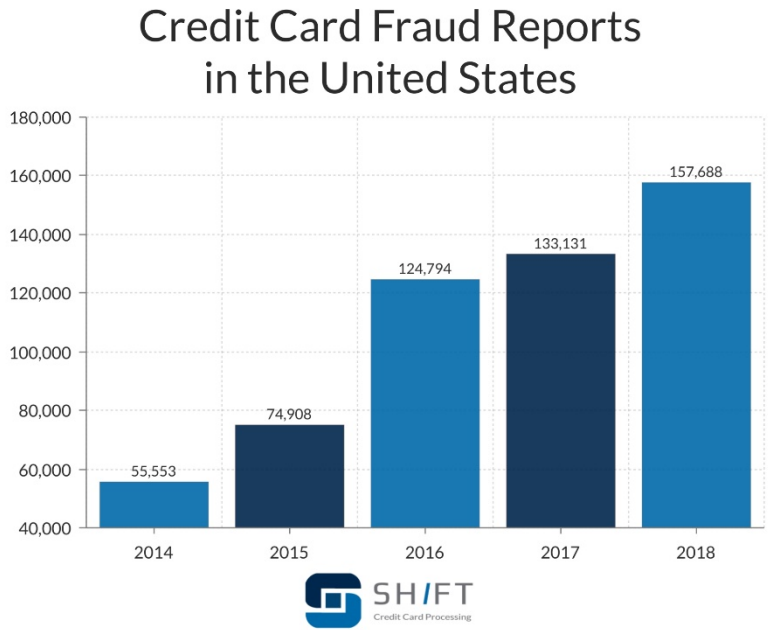
IEEE

Advancing Technology for Humanity

The world's largest professional association for the advancement of technology

Problem Statement

- Credit card fraud is the unauthorized use of a credit/debit card, fraudulently obtaining money or property.
- Cases of Credit Card Fraud are increasing every year in the US.
- Credit Card fraud costs consumers and financial companies billions of dollars every year.
- Developing a fraud detection system is important to prevent losses across the various parties.



Objective

- Developing a data science project based on IEEE CIS dataset
- Benchmarking machine learning models on large scale dataset to predict fraudulent customers.

Methodology

1

Light Gradient Boosting

- Gradient boosting method.
- Tree leaf-wise.
- Solving regression or classification problems.

2

Extreme Gradient Boosting

- Gradient Boosting.
- Decision Tree based ensemble Machine Learning.

3

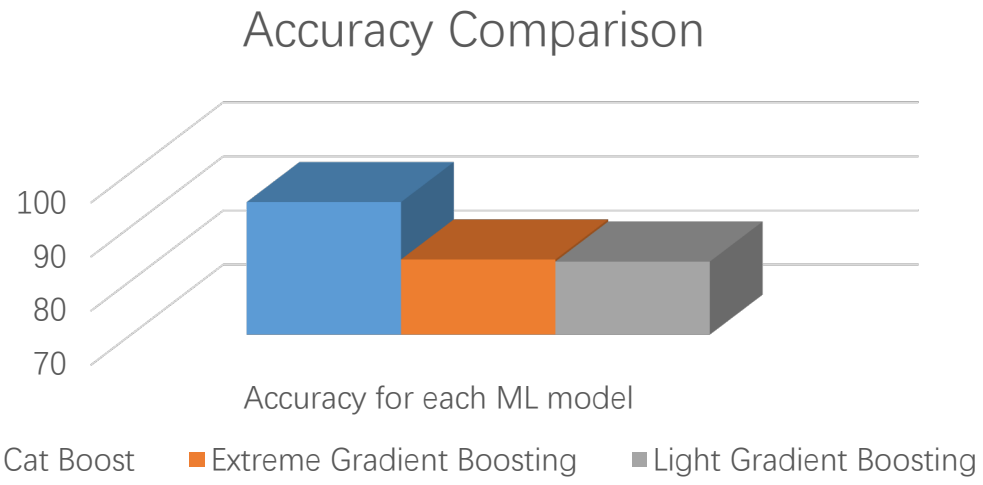
CAT Boosting

- Gradient Boosting.
- Great results with default parameters.
- Faster Predictions.
- Ordered boosting

Results

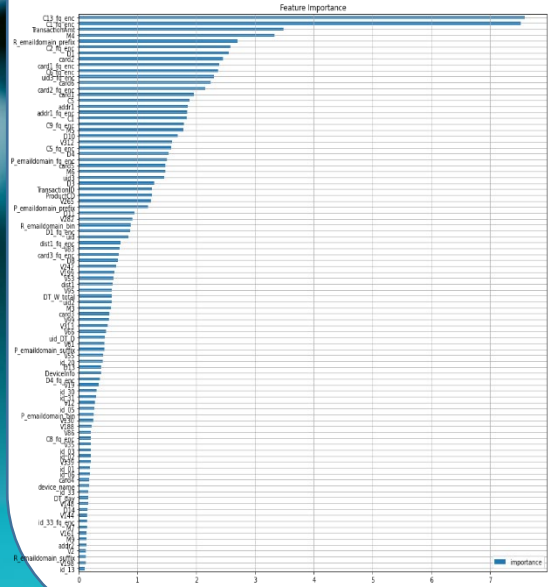
Catboost presented with the best accuracy at 94.9%.

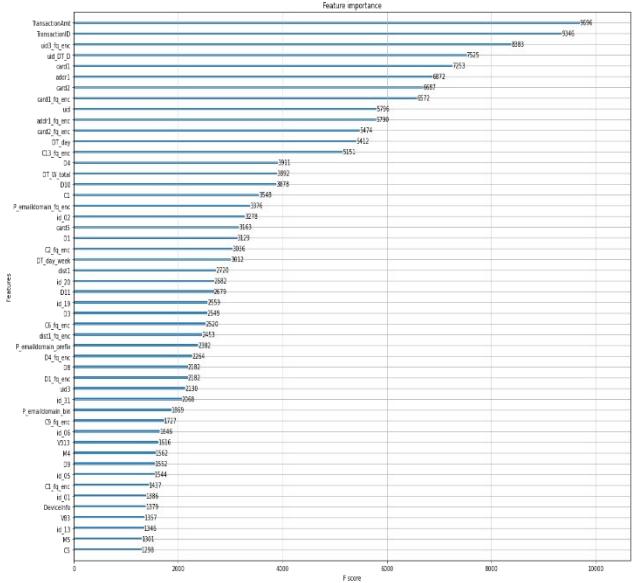
1. Catboost: 94.9%
2. Extreme Gradient Boosting: 83.9%
3. Light Gradient Boosting: 83.5%

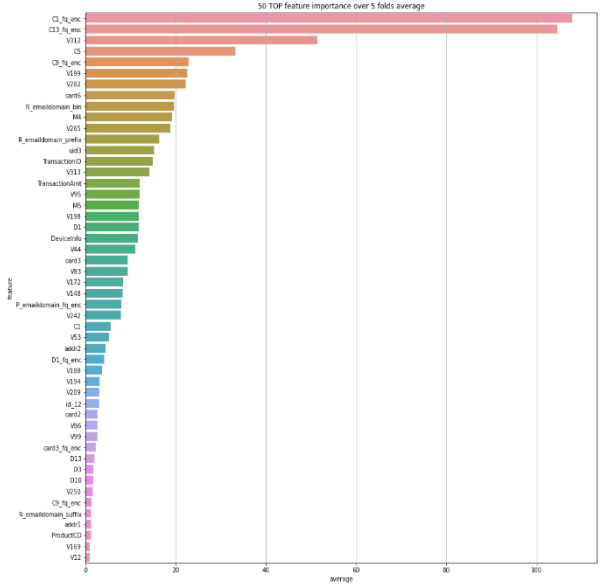


Results

- Different result for feature importance for each models







Conclusion – Continual Learning

Feed new dataset

New customer data would come in everyday. It's always important to keep the data up to date

Feature Engineering

Clean up the data and feed good data into machine learning models for greater results.

Monitor & Repeat

Make sure everything is working correctly and repeat the cycle.

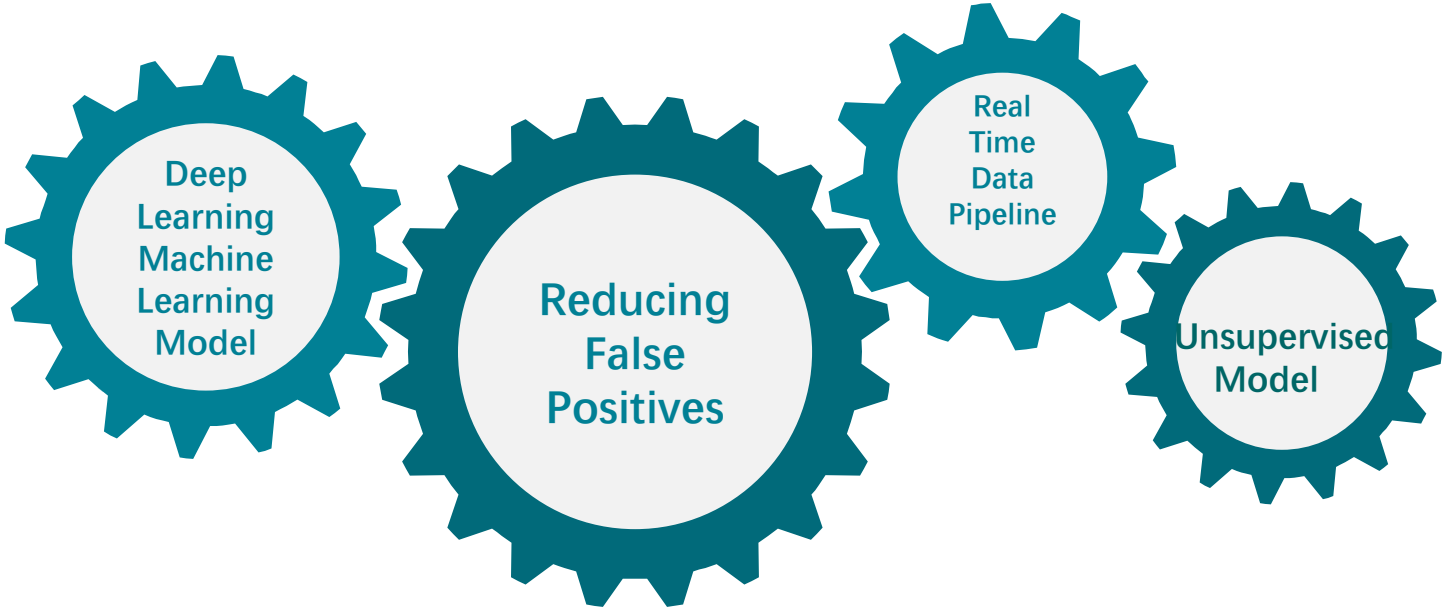
Obtain insights

New data exploration and get the insights for the differences

ML Modeling/Validation

Build a machine learning model based on new data and pre-existing dataset.

Future Works



Questions and Answers

THANK YOU !

Hyungjun Kang