Credit Card Defaulters:

Analysis and Prediction Project

Martin Reyes

Business Problem

- Goal: Gain insights from customer info and predict whether they will default their next bill

- Insights:
  - Most relevant feature: recent pay (on-time) history
- Model Results: Optimal model produced a .88 recall score. Out of all defaulters, 88% can be identified.
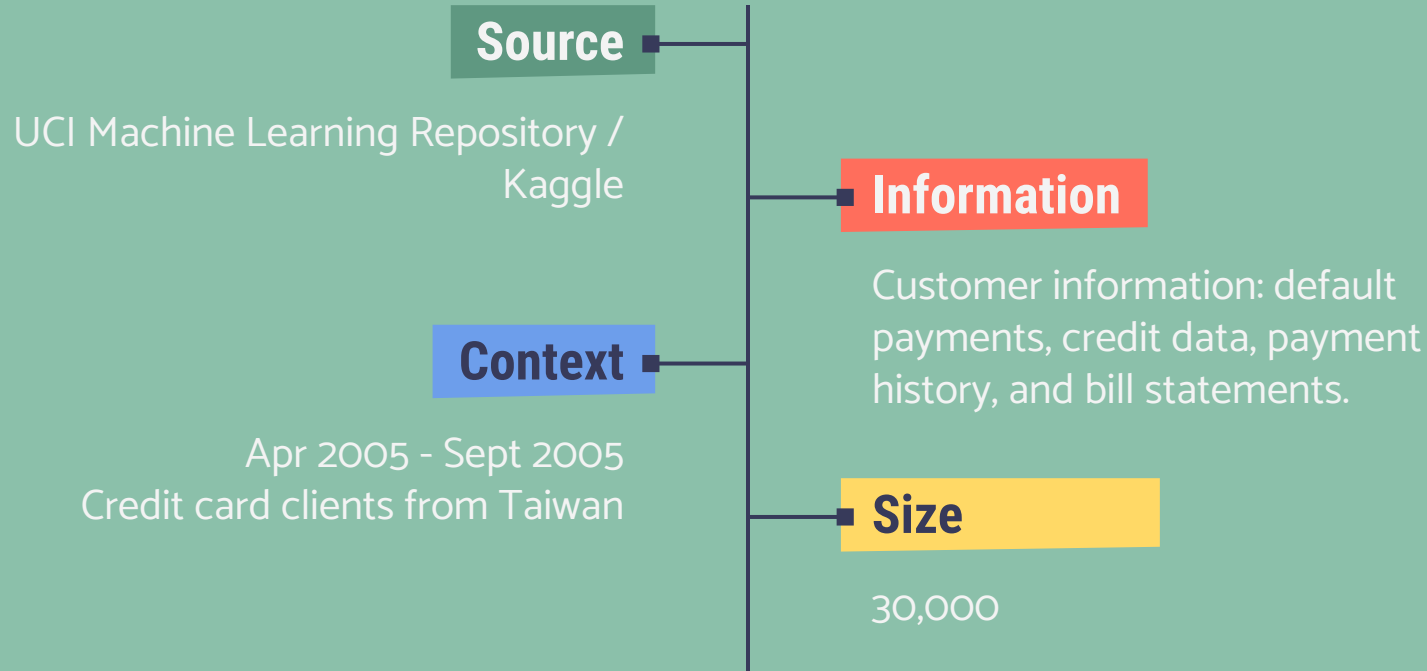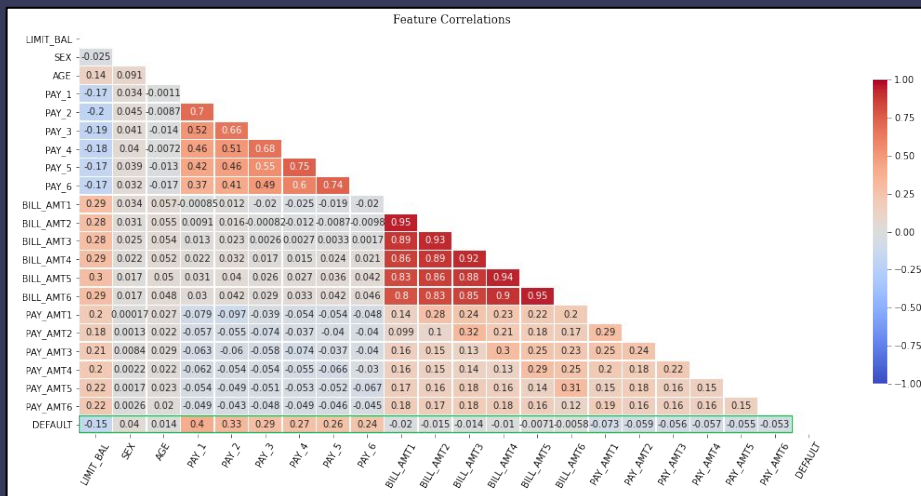
PROBLEM

Results

About the Data

Source

UCI Machine Learning Repository / Kaggle

Information

Customer information: default payments, credit data, payment history, and bill statements.

Context

Apr 2005 - Sept 2005
Credit card clients from Taiwan

Size

30,000

Findings

Feature Correlations

- Payment history correlates most with whether a customer will default
- Other features don't seem to be related
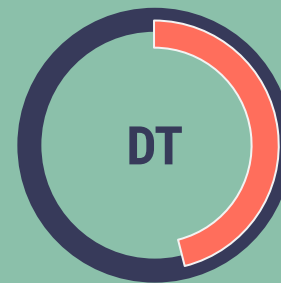
Predictive Model Results
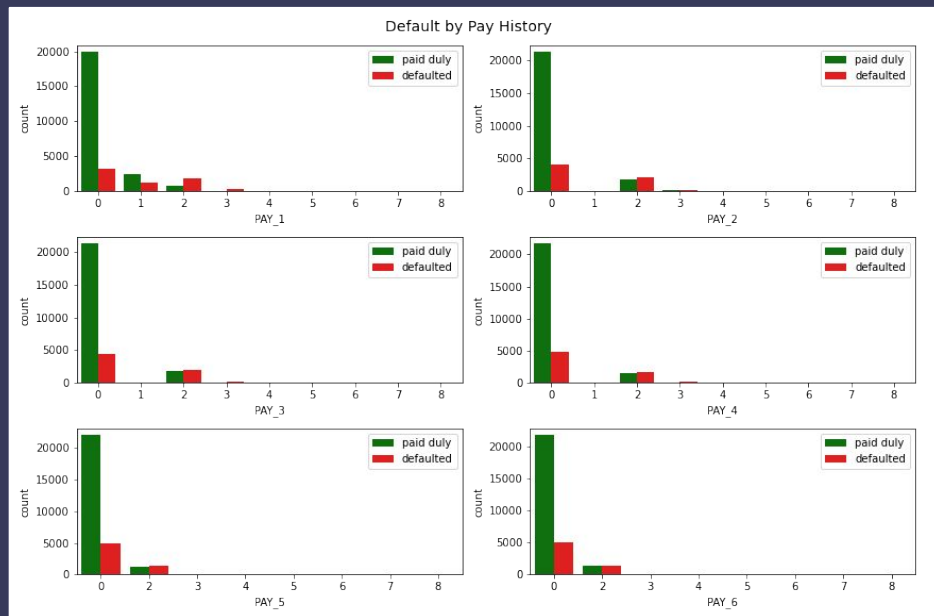
70%

KNN

K-Nearest Neighbors

88%

XGBoost

XGBoost

46%

DT

Decision Tree

Optimal Model

XGBoost Classifier:

- 88% of defaulters predicted
- 59% accurate

10 Most Import Features – XGBoost

Predictions – XGBoost

Summary of Results

Default by Pay History

- Monitor customer's payment history
- Don't focus on age, sex, and/or bill statements
- Optimal Model: XGBoost Classifier
  - Predicts 88% of defaulters

Default by Sex
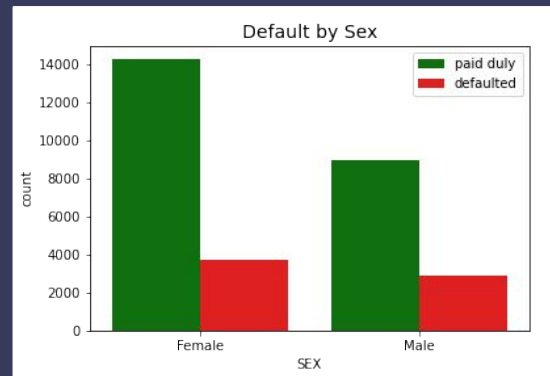
Next Steps

Advanced machine/deep learning techniques

More feature engineering and dimensionality reduction

Explore the data more for insights

Gather more data

Questions?

Email

martinreyes.eng@gmail.com

GitHub

github.com/MartinReyesDS

LinkedIn

www.linkedin.com/in/martin-reyes-ds