RIT Business Analytics Spring 2022 Competition

Team MaTron
Matt Hochman, Tron Schell

RIT

Problem Description

Problem Description

Small Capital Bank (SCB) is a novel loan company.

We have been tasked to advise SCB on their current loan decision making process and loan portfolio health.

Problem Description

- **Decision Boundary (Model)**

- **Mitigation of Bias**

- **Portfolio Performance**

Data Preprocessing

Datasets

Dataset 1

Loan portfolio (2017-2018)

Used to build our model

Dataset 2

Current loan applications

Where our model will be applied

Datasets (cont.)

There were features in Dataset 1 that were not contained in the Dataset 2 and needed to be removed to avoid what is known as "Data Leakage".

Dataset 1

877986 rows × 47 columns

Dataset 2

495242 rows × 39 columns

Datasets (cont.)

Because we are predicting if a person will/will not default, this is called a classification problem.

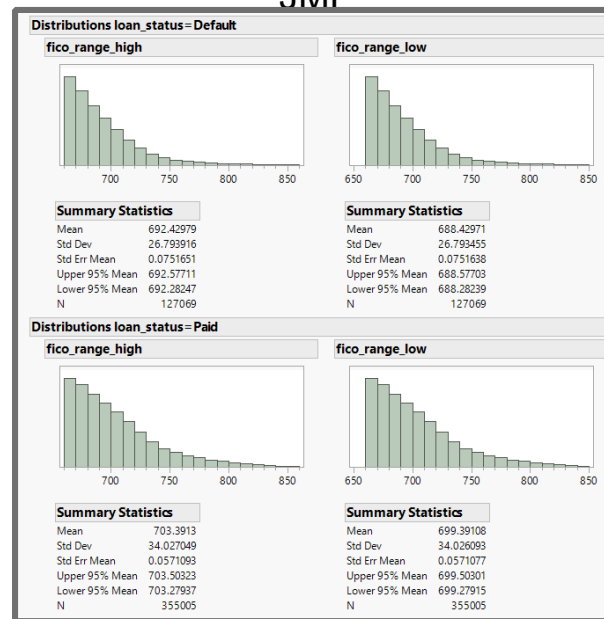Unbalanced classification problem ⟶

```
Paid        355005
Default     127069
Name: loan_status, dtype: int64
```
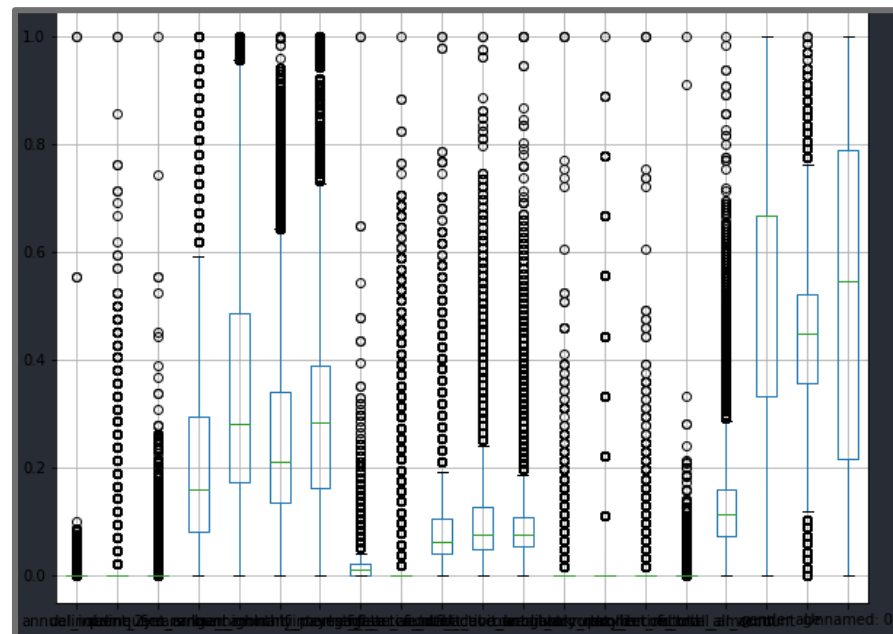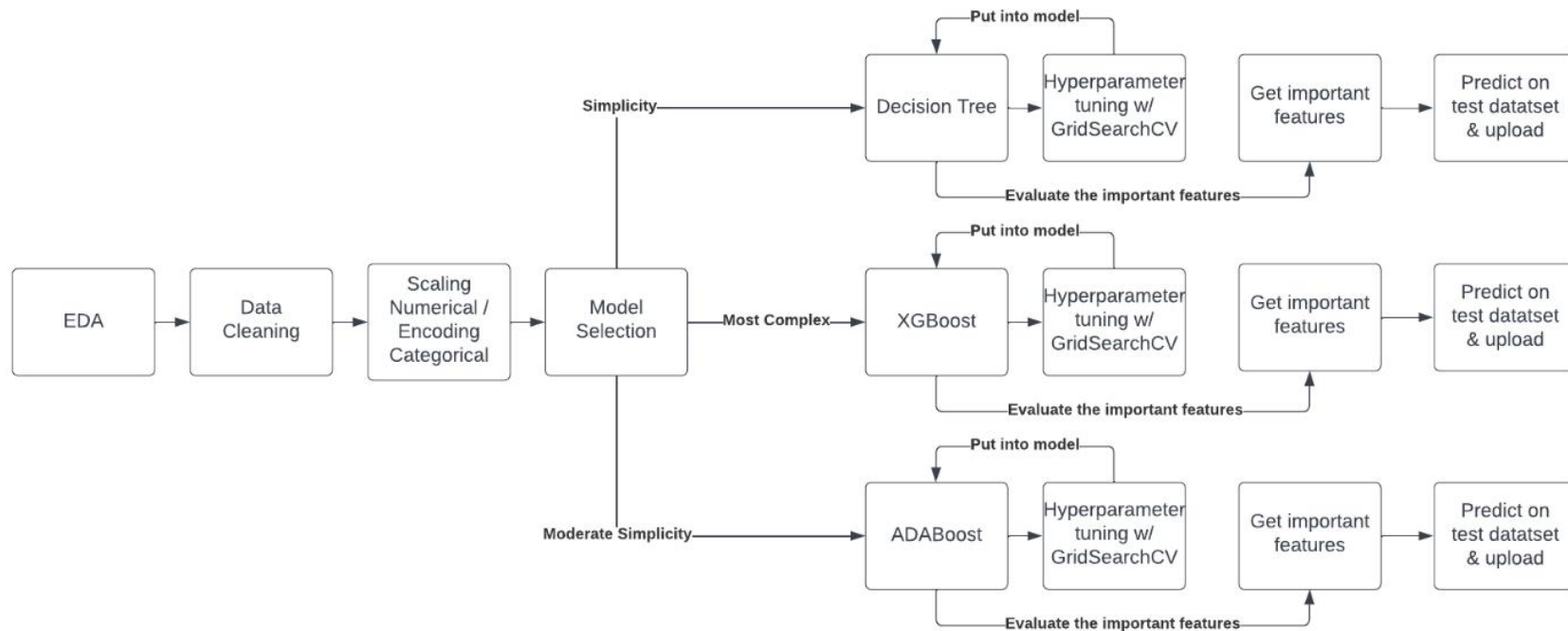
Exploratory Data Analysis

Tableau

JMP

Exploratory Data Analysis (cont.)

Python

Predictive Model

Predictive Model

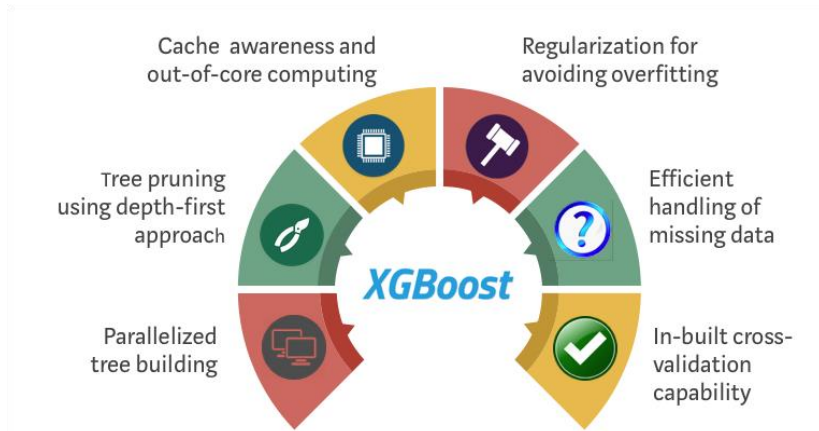# Predictive Model Selection

Three models were constructed

- **Decision Tree**
  - Simplicity
- **AdaBoost**
  - Worked well with unbalanced datasets
- **XGBoost**
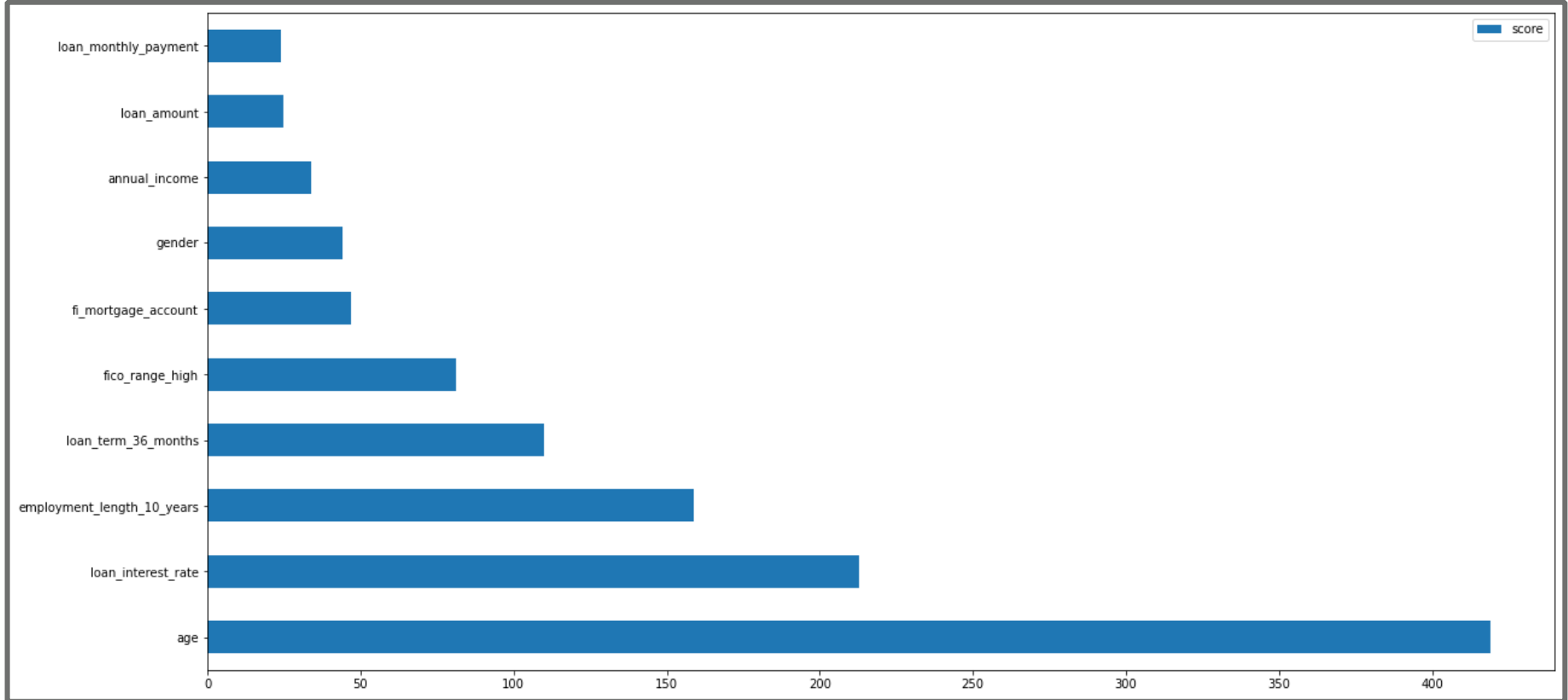 - A "good at everything" machine learning algorithm

Final Predictive Model Selection
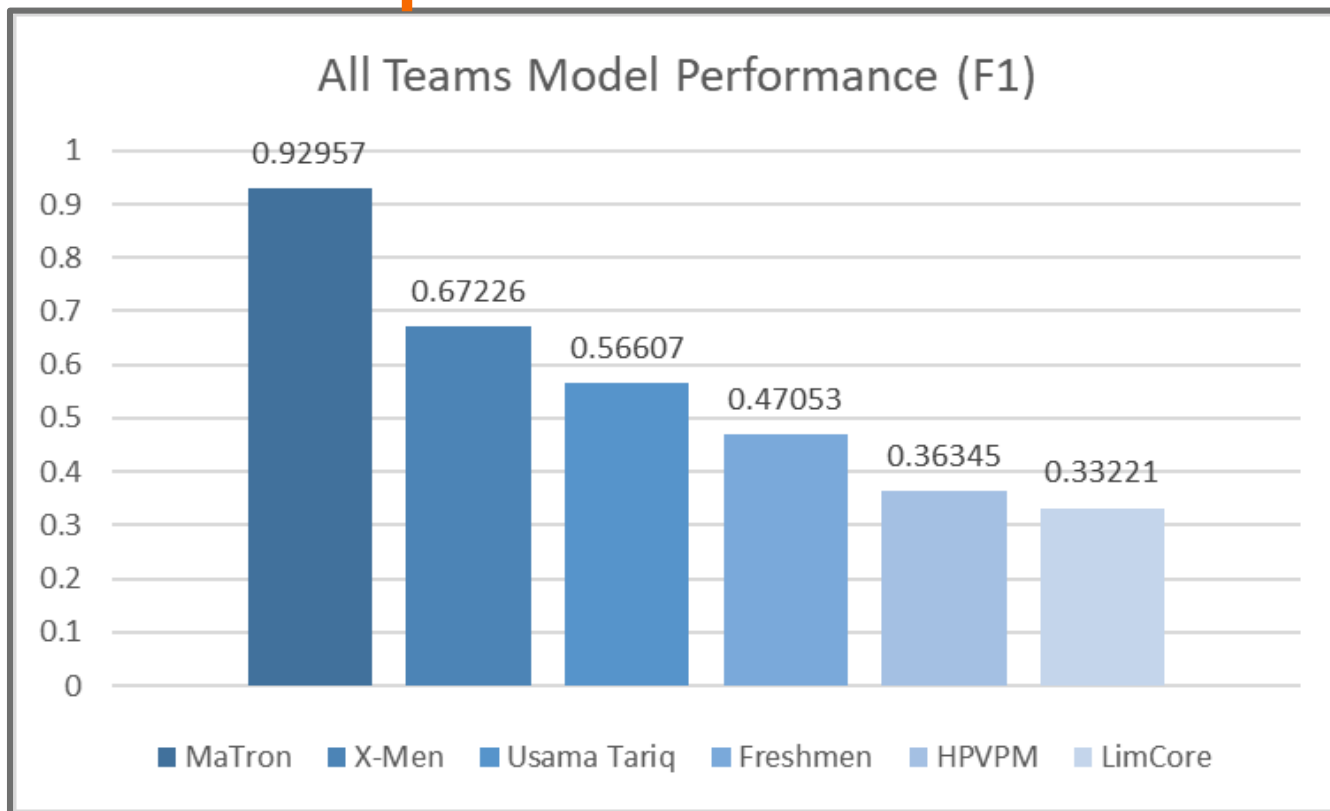
- **XGBoost**
  - High base F1 score (0.93)
  - Hyperparameter tuning
    - "scale_pos_weight"
  - Industry standard ML model

What factors Influenced the Model?

Competition Comparison

All Teams Model Performance (F1)

MaTron: 0.92957, X-Men: 0.67226, Usama Tariq: 0.56607, Freshmen: 0.47053, HPVPM: 0.36345, LimCore: 0.33221
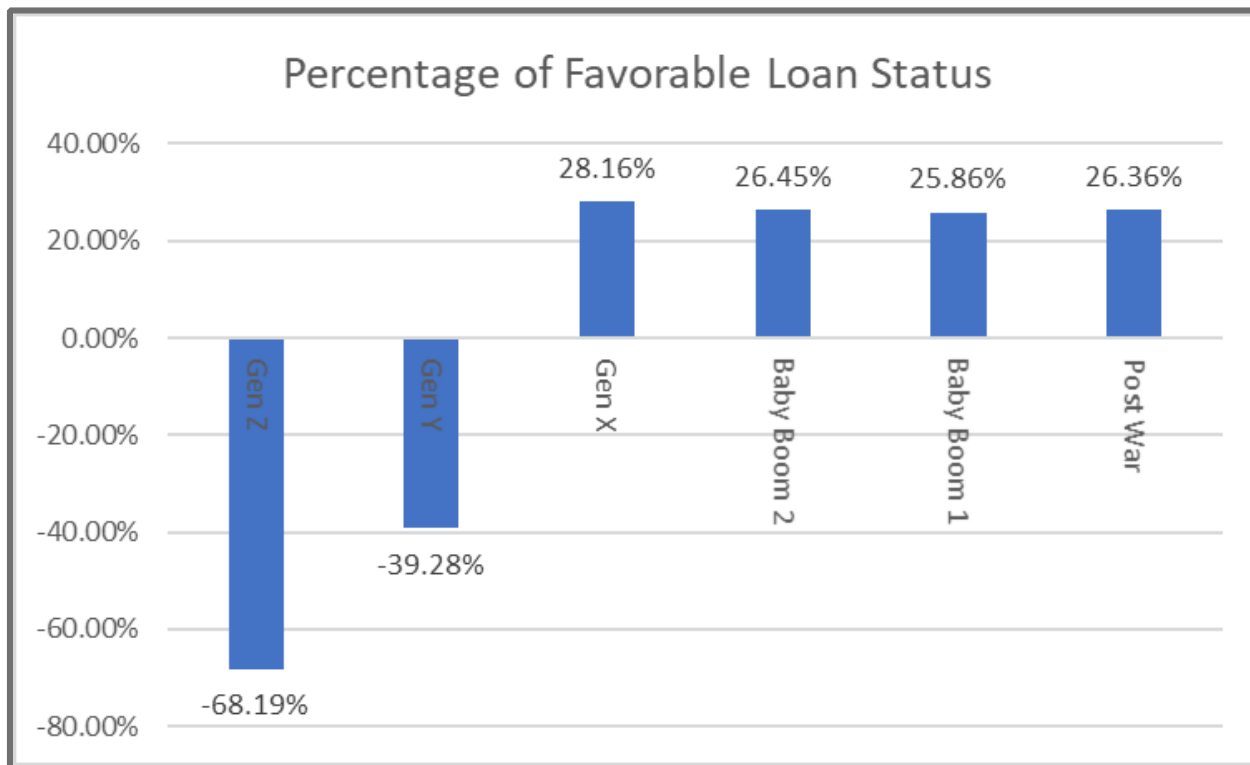
Bias

Protected Classes

- **Regulation of AI and ML Models**
  - Use in employment decisions
  - Future regulation
- **Protected Classes**
  - 8 major categories
    - Focus on Age

Source: https://www.eeoc.gov/newsroom/eeoc-launches-initiative-artificial-intelligence-and-algorithmic-fairness#:~:text=%E2%80%9CBias%20in%20employment%20arising%20from,anti%2Ddiscrimination%20laws%20still%20apply.
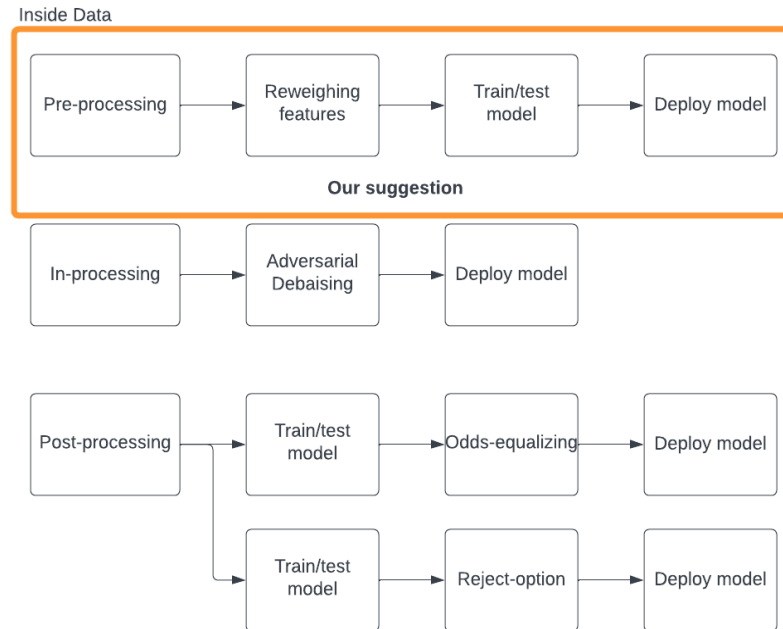
Loan Discrimation based on Age
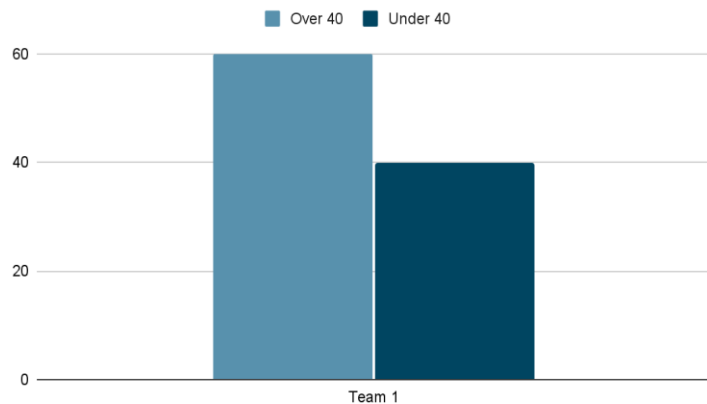
Bias Mitigation Solution

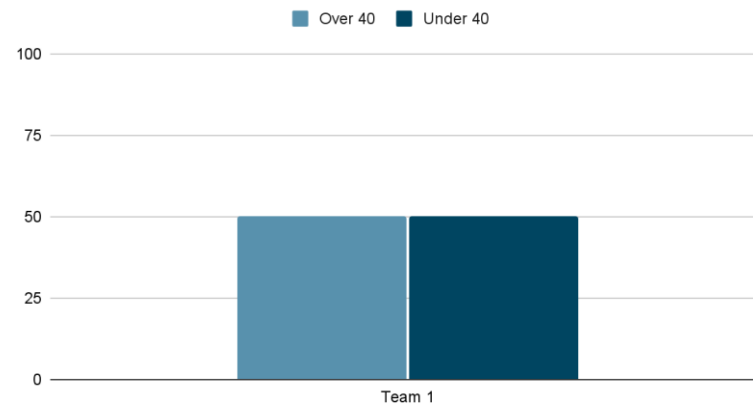## 3 Techniques:

1. Pre-processing
2. In-processing
3. Post-processing

Reweighing Example

Reweighing Effects on Model Performance

Before Reweighing

After Reweighing

Balanced accuracy = 0.6703

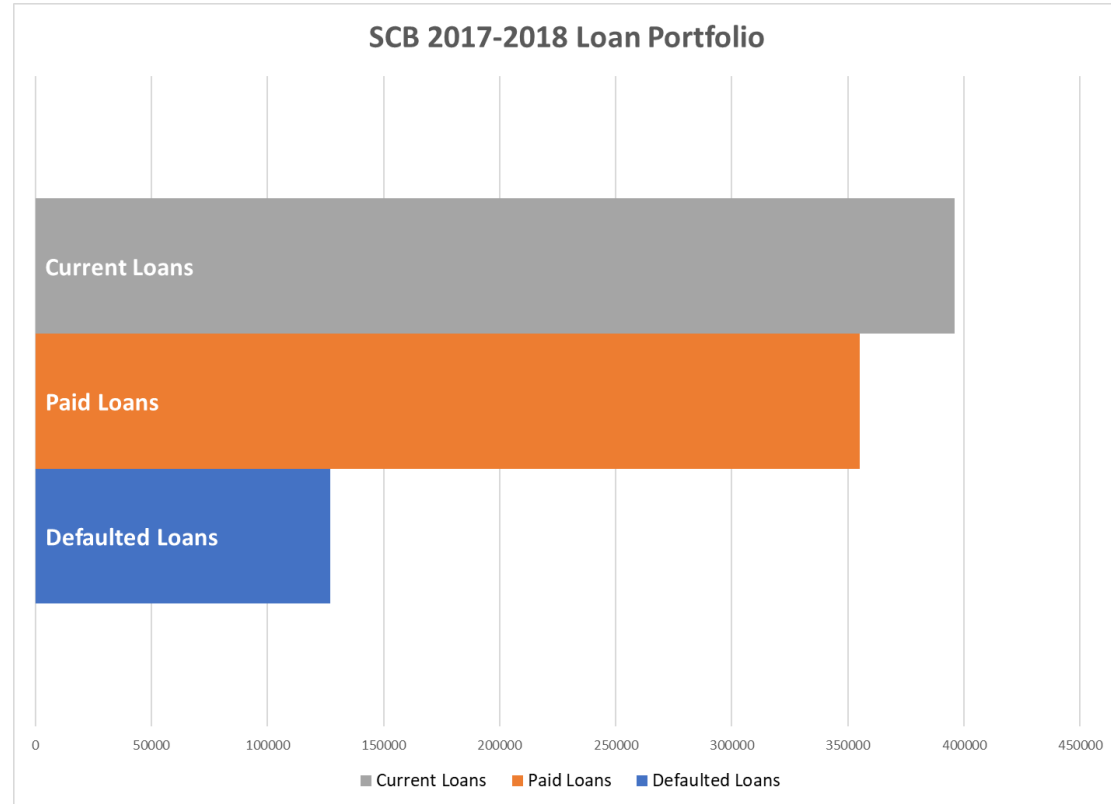Balanced accuracy = 0.6557

Portfolio Performance

A Healthy Portfolio?

Portfolio Performance

- High Proportion of Paid Loans
- High Loan Profitability Ratio

Is SCB's Portfolio Healthy?

Portfolio Statistics

SCB 2017-2018 Loan Portfolio

Portfolio Statistics

SCB 2017-2018 Loan Portfolio
(Current Loans Predicted)
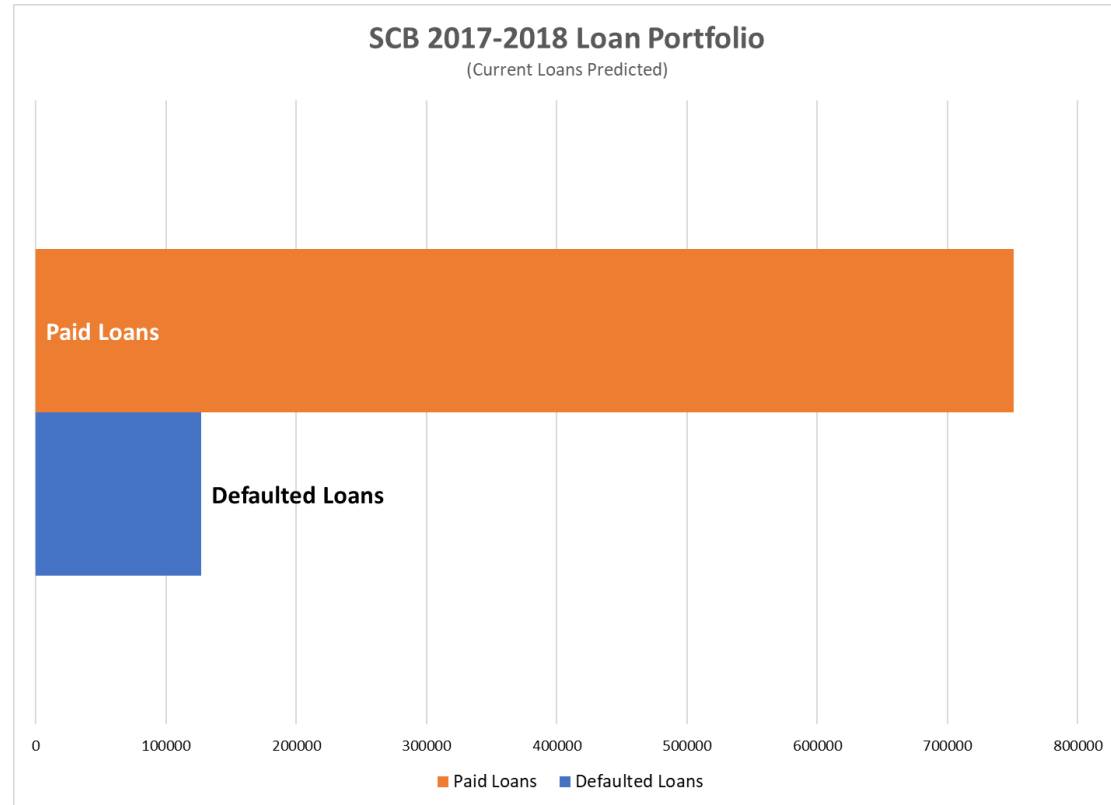
Mean Profitability Ratios

Mean Profitability Ratios	
Mean of Paid Loans	1.12
Mean of Defaulted Loans	0.40
36 Month Loans	0.46
60 Month Loans	0.30
Mean of Paid & Defaulted Loans	0.93
Mean of Current Loans	0.25
Mean of Current Loans (Predicted)	1.16

Portfolio Profitability Ratios

Portfolio Profitability Ratios	
Paid & Defaulted Loans	0.91
All Loans (Current Predicted)	**1.10**

Factors Affecting Profitability

- **Paid Loans**
  - Interest Rate

- **Defaulted Loans**
  - Loan Completion
  - Interest Rate

Time to Default	
Average Months to Default (All)	15
Average Months to Default (36 Month)	15
Average Months to Default (60 Month)	16

Is SCB's Portfolio Healthy?
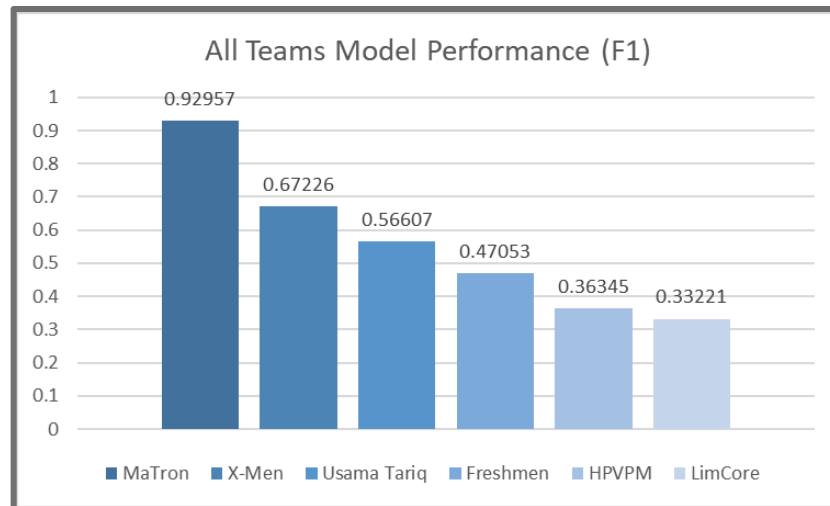
Recommendations

Increase Profitability

- **Ensure Loan Repayment**
  - Avoid high risk borrowers

- **36 vs 60 Month Loans**
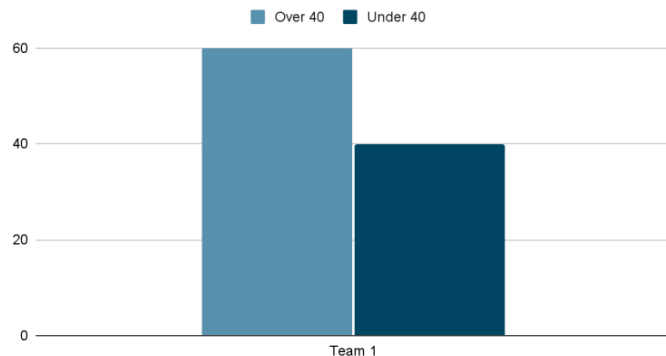 - Difference in default profitability

Using our Model

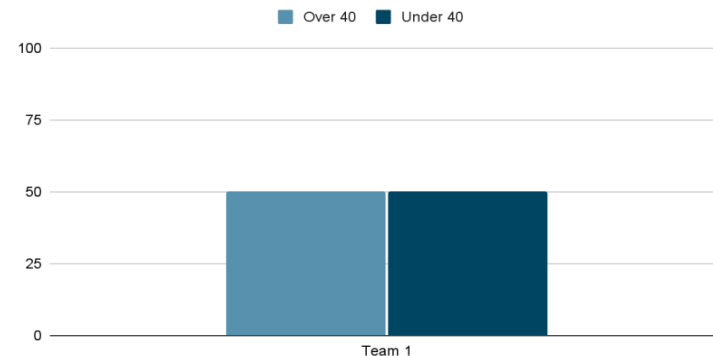## Our XGBoost model is an effective prediction tool

All Teams Model Performance (F1)

- MaTron: 0.92957
- X-Men: 0.67226
- Usama Tariq: 0.56607
- Freshmen: 0.47053
- HPVPM: 0.36345
- LimCore: 0.33221

Bias Mitigation

- **Changing Data Collection**
- **Reweighing**

Questions?