RIT Business Analytics
Spring 2022 Competition

Team MaTron
Matt Hochman, Tron Schell
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

<table>
<thead>
<tr>
<th>Paid</th>
<th>355005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default</td>
<td>127069</td>
</tr>
<tr>
<td>Name: loan_status, dtype: int64</td>
<td></td>
</tr>
</tbody>
</table>
Exploratory Data Analysis
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
- HPVM: 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

Loan Discrimination based on Age

Percentage of Favorable Loan Status

<table>
<thead>
<tr>
<th>Generation</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gen Z</td>
<td>-68.19%</td>
</tr>
<tr>
<td>Gen Y</td>
<td>-39.28%</td>
</tr>
<tr>
<td>Gen X</td>
<td>28.16%</td>
</tr>
<tr>
<td>Baby Boom 2</td>
<td>26.45%</td>
</tr>
<tr>
<td>Baby Boom 1</td>
<td>25.86%</td>
</tr>
<tr>
<td>Post War</td>
<td>26.36%</td>
</tr>
</tbody>
</table>
Bias Mitigation Solution

3 Techniques:
1. Pre-processing
2. In-processing
3. Post-processing
Reweighing Example

Number of positive outcomes per class

Team 1

Number of positive outcomes per class

Team 1
## Reweighing Effects on Model Performance

<table>
<thead>
<tr>
<th>Before Reweighing</th>
<th>After Reweighing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balanced accuracy = 0.6703</td>
<td>Balanced accuracy = 0.6557</td>
</tr>
</tbody>
</table>
Portfolio Performance
A Healthy Portfolio?

- High Proportion of Paid Loans
- High Loan Profitability Ratio
Is SCB’s Portfolio Healthy?
Portfolio Statistics
Portfolio Statistics

SCB 2017-2018 Loan Portfolio
(Current Loans Predicted)

Paid Loans

Defaul ted Loans

0 100000 200000 300000 400000 500000 600000 700000 800000
### Mean Profitability Ratios

<table>
<thead>
<tr>
<th>Description</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of Paid Loans</td>
<td>1.12</td>
</tr>
<tr>
<td>Mean of Defaulted Loans</td>
<td>0.40</td>
</tr>
<tr>
<td>36 Month Loans</td>
<td>0.46</td>
</tr>
<tr>
<td>60 Month Loans</td>
<td>0.30</td>
</tr>
<tr>
<td>Mean of Paid &amp; Defaulted Loans</td>
<td>0.93</td>
</tr>
<tr>
<td>Mean of Current Loans</td>
<td>0.25</td>
</tr>
<tr>
<td>Mean of Current Loans (Predicted)</td>
<td>1.16</td>
</tr>
</tbody>
</table>
### Portfolio Profitability Ratios

<table>
<thead>
<tr>
<th>Description</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paid &amp; Defaulted Loans</td>
<td>0.91</td>
</tr>
<tr>
<td>All Loans (Current Predicted)</td>
<td>1.10</td>
</tr>
</tbody>
</table>
Factors Affecting Profitability

- **Paid Loans**
  - Interest Rate

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

<table>
<thead>
<tr>
<th>Time to Default</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Months to Default (All)</td>
<td>15</td>
</tr>
<tr>
<td>Average Months to Default (36 Month)</td>
<td>15</td>
</tr>
<tr>
<td>Average Months to Default (60 Month)</td>
<td>16</td>
</tr>
</tbody>
</table>
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
Bias Mitigation

- Changing Data Collection
- Reweighing
Questions?