AI Boot Camp Project 2

Predicting Loan Approval

Team Members:
Solomon Ajayi
Ben Cross
Lewis Hill
Robert Rodriguez

Project Overview

Loan Approval Prediction: Weeding Out the "Wallet Wizards" from the "Financially Foolish"

- Ditch the dartboard - this machine learning model predicts loan approvals with laser focus.
- Credit history and current income: the secret sauce (no crazy collateral required, unlike that time Uncle Fred offered his prized porcelain pig collection).

Project Overview

Goals to be addressed

- **Goal 1: Enhance Loan Approval Accuracy** - Emphasis on the model's ability to improve the precision of loan decisions.
- **Goal 2: Mitigate Approval Bias** - An objective of reducing prejudice in the loan approval process.
- **Goal 3: Foster Responsible Lending Practices** - Focused on the model's role in promoting sound financial decisions for both the borrower and lender.

Project Overview

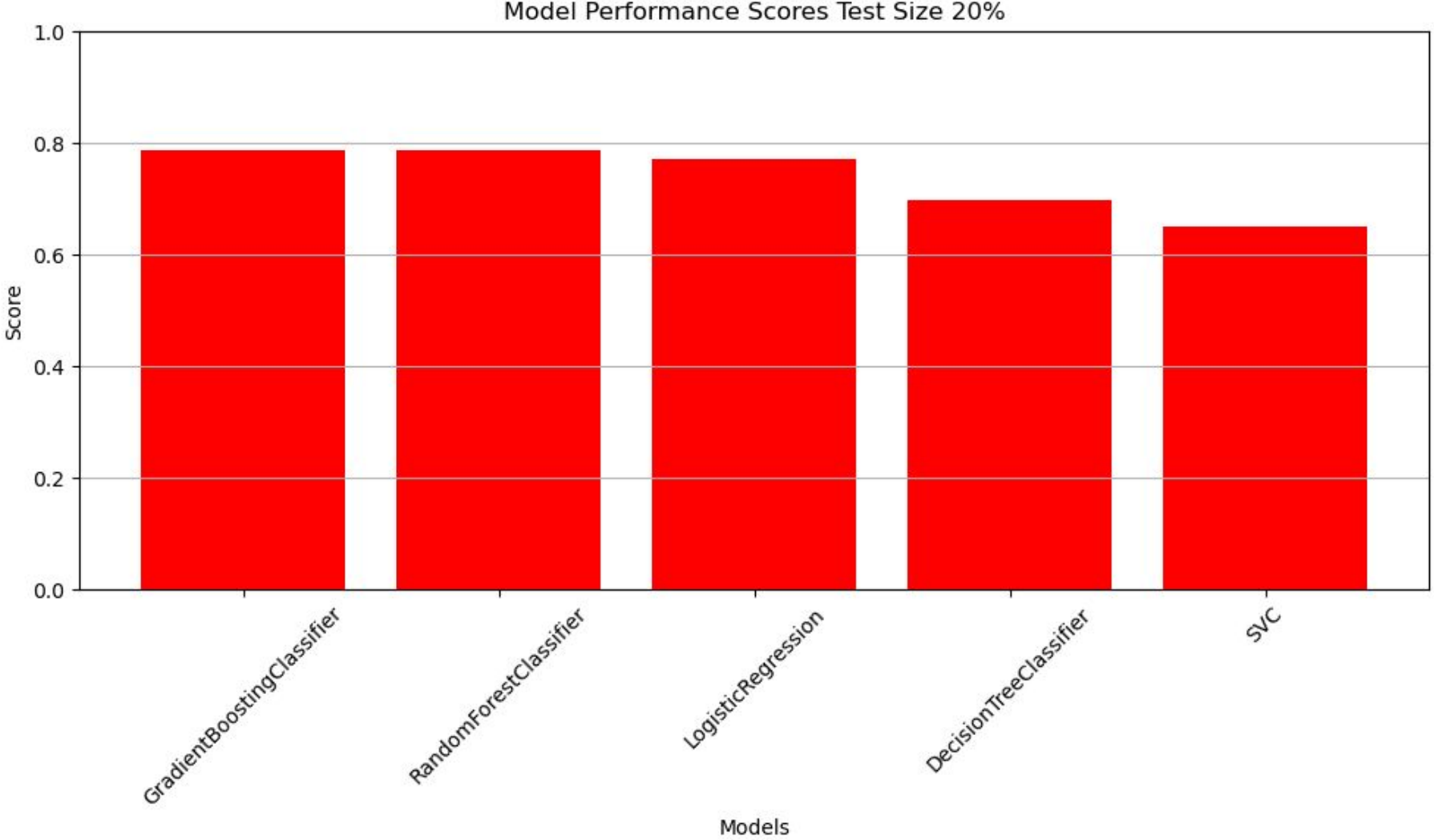
Overview of data collection, cleanup and exploration process

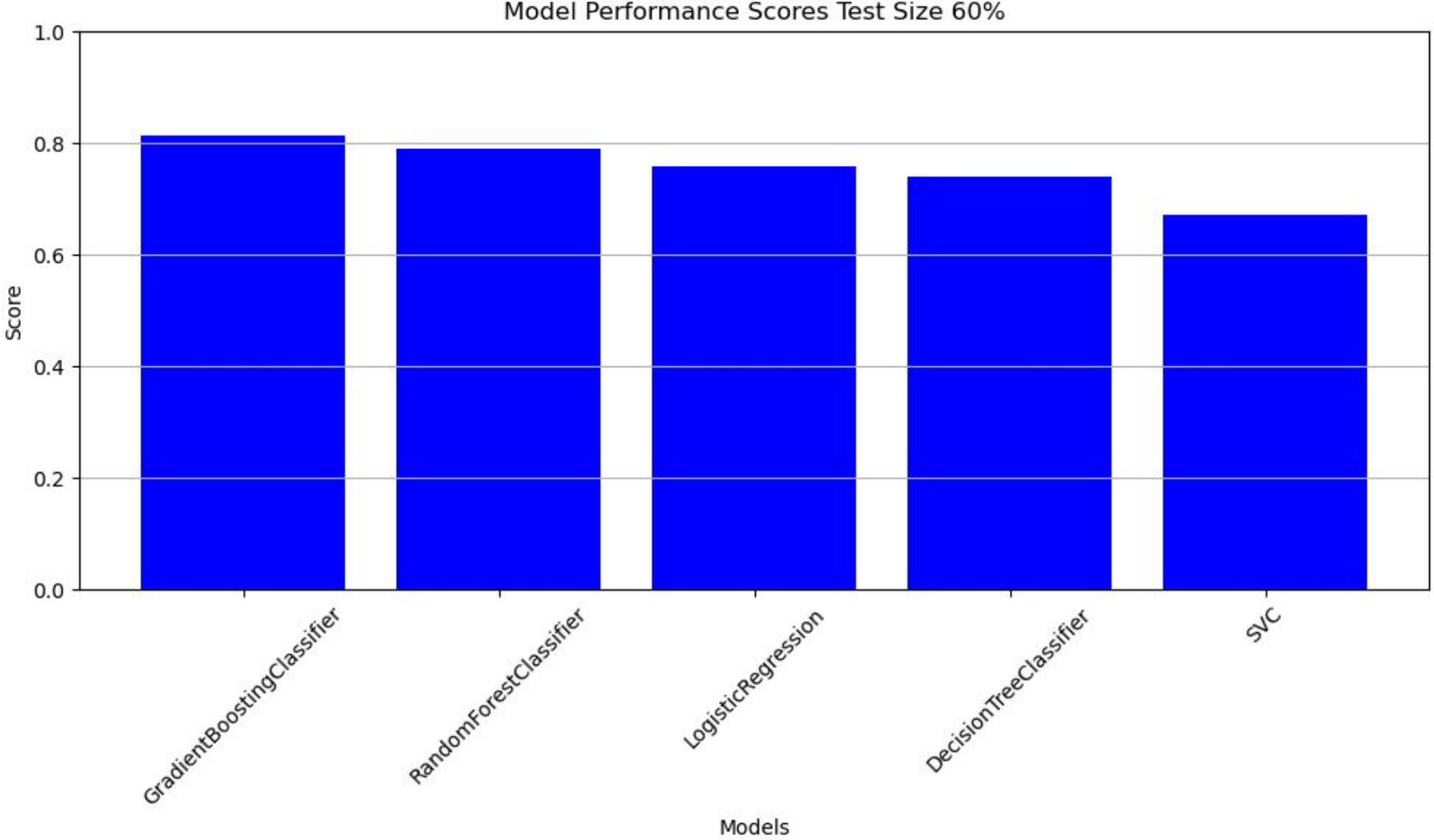
- Leveraged Kaggle, a reputable data repository, to acquire a comprehensive loan applicant dataset.
- Conducted a thorough review process to ensure data alignment with project requirements and suitability for machine learning modeling.
- Cleaned the data to ensure accuracy and usability
- Filled NaN data with Mode (categorical) and Mean (numerical)
- Changed the Test/Train percentages at increments to find the best accuracy for each model.
    - Utilized 20% and 60% Test Sizes

Project Overview

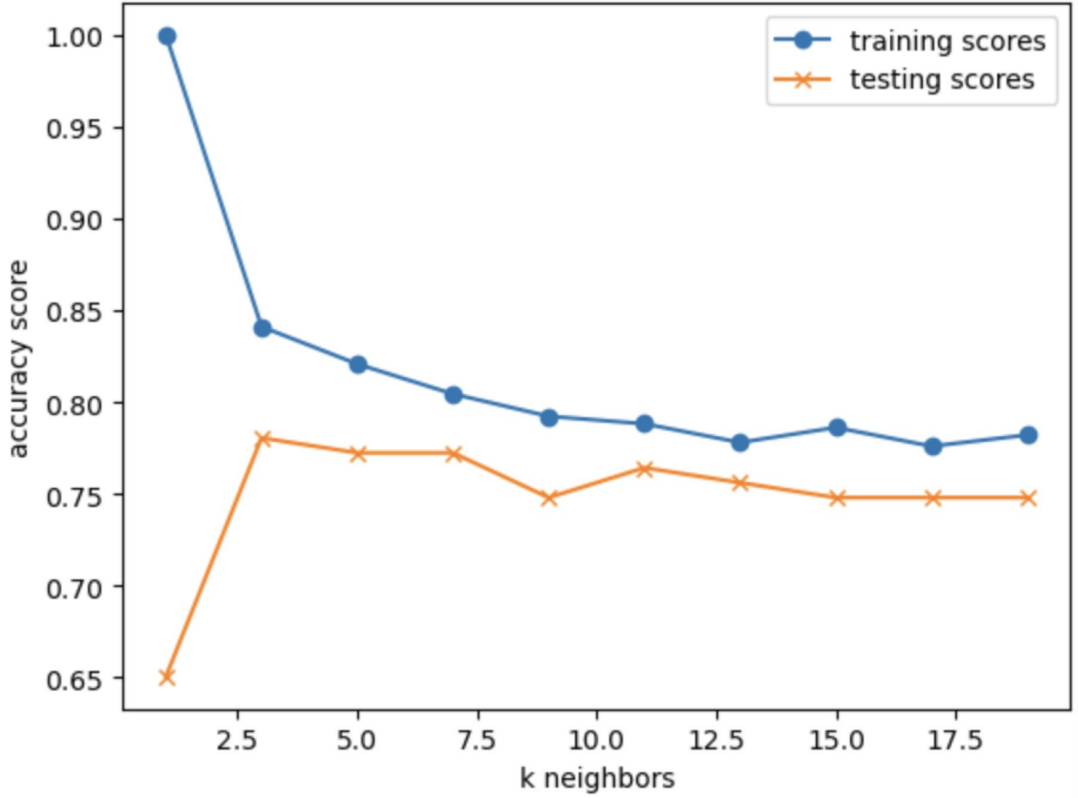
Approach taken to achieve goals

- Implemented a multi-model evaluation approach, employing Logistic Regression, Decision Trees, Random Forests, Gradient Boosting and SVC.
- Executed multiple iterations of Test/Train percentages in order to identify most accurate model
- Identified Gradient Boosting as the optimal model, achieving a robust accuracy of 81% on a 60% data test set.


Model Performance Scores Test Size 20%
1.0
0.8
0.6
0.4
0.2
0.0
Score
GradientBoostingClassifier
RandomForestClassifier
LogisticRegression
DecisionTreeClassifier
SVC
Models

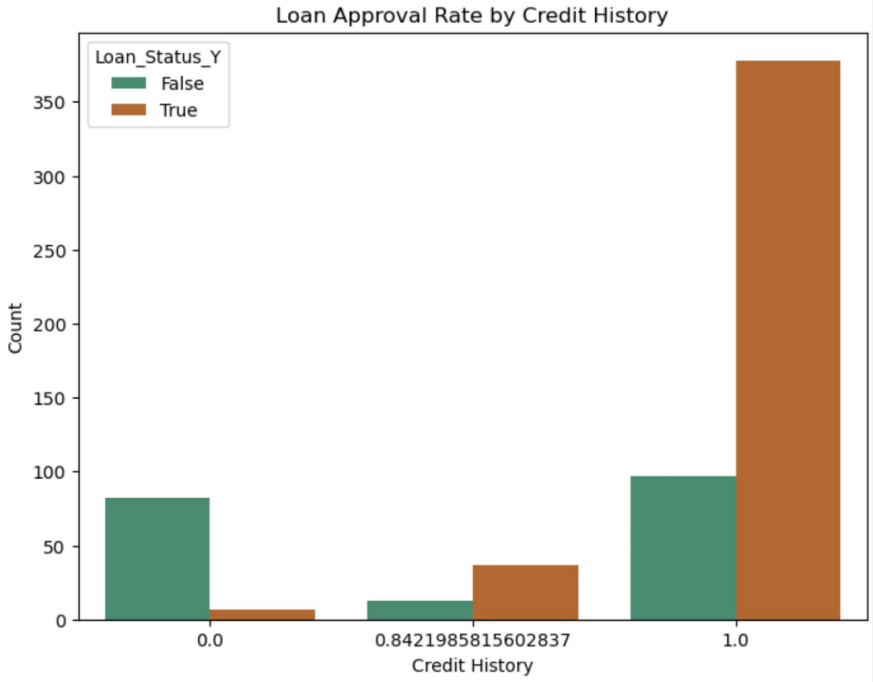

Model Performance Scores Test Size 60%
1.0
0.8
0.6
0.4
0.2
0.0
Score
GradientBoostingClassifier
RandomForestClassifier
LogisticRegression
DecisionTreeClassifier
SVC
Models

KNeighbors

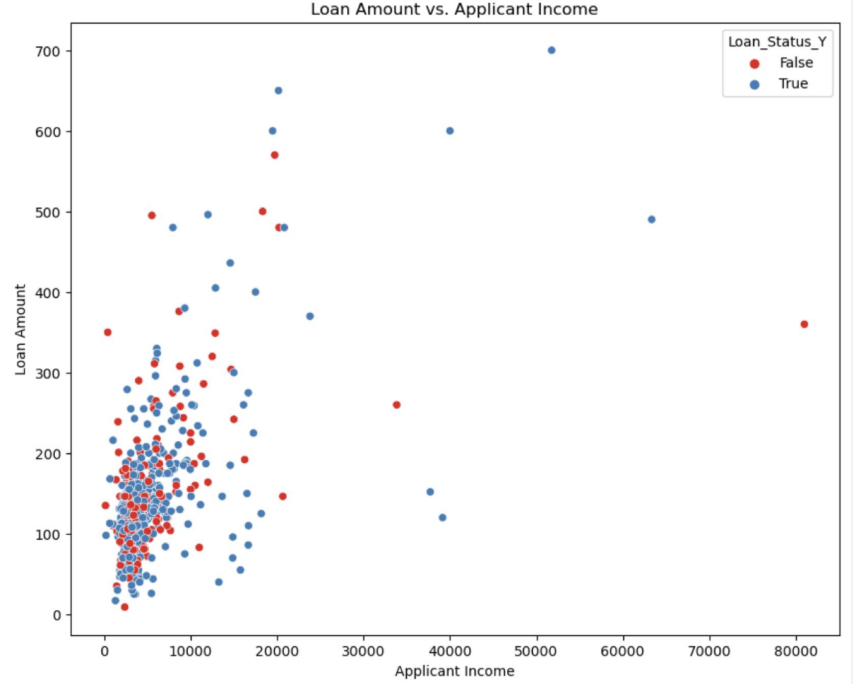



```
k: 1, Train/Test Score: 1.000/0.650
k: 3, Train/Test Score: 0.841/0.780
k: 5, Train/Test Score: 0.821/0.772
k: 7, Train/Test Score: 0.804/0.772
k: 9, Train/Test Score: 0.792/0.748
k: 11, Train/Test Score: 0.788/0.764
k: 13, Train/Test Score: 0.778/0.756
k: 15, Train/Test Score: 0.786/0.748
k: 17, Train/Test Score: 0.776/0.748
k: 19, Train/Test Score: 0.782/0.748
```

Summary: Highest Correlation



➔ Credit History



➔ Applicant Income

Problems Encountered

- Lack of Subject Matter Expertise in both Data Analytics and Finance
  - Limited experience by group in both fields
- Lack of Expertise in Data Bias
- NaN and Missing Data

Future Considerations

- Potentially seek a second data set to run the same models to see if we can get improved accuracy scores and/or combine datasets
- Locate a data set that contains our key columns as well as potential columns where bias could affect the model and compare accuracies
- Advise Stakeholders to add a data input section that gives potential Loan Applicants a place to input monthly expense
    - Applicant income amount could be subtracted by this data and refine and reduce potential future defaults