

Project Milestone 1

Harlan Wittlieff

Data Science, Bellevue University

DSC 680: Applied Data Science

Dr. Catie Williams

June 12, 2022

Topic

The small finance company that I work for, Loans Today, is launching an initiative to streamline our loan approval process through leveraging automation. The project is called “Approval AI”

Business Problem

Currently, Loans Today employs a small team of Loan Officers. This was an adequate staffing level when the operation started. However, the manual loan approval process is creating a bottleneck in the loan operations. Due to today’s tight labor market and increasing compensation package demands, our finance company does not have the budget to hire more Loan Officers. Additionally, it would take a great deal of time to train new hires and get them up-to-speed on the loan approval process. Therefore, the company is turning to AI for a solution to this problem. The goal is to create an algorithm that can predict loan status as approved, yes or no.

Datasets

The Loan Approval Data Set on Kaggle will serve as the primary dataset for this data science project. The dataset features the key variables included on the Loans Today loan application. These variables include the following:

- Loan ID
- Gender
- Married
- Dependents
- Education
- Self Employed
- Applicant Income
- Co-applicant Income
- Loan Amount

- Loan Amount Term
- Credit History
- Property Area
- Loan Status

Methods

The target variable for this project is “Loan Status,” as the model will predict whether a loan is approved. If the classes are not balanced, a method such as SMOTE would need to be leveraged to address the issue. As this is a classification scenario, a logistic regression model will be investigated. Additional models, such as the random forest model may also warrant investigation pending the results of the logistic regression model.

Ethical Considerations

Seeing as a loan can significantly impact the livelihood of the applicant, care must be taken to ensure there are not any unfair biases in the model. This due diligence is required not only when building the model, but also following deployment to ensure the results are equitable.

Challenges/Issues

Since we are a smaller organization our data set is modest with 614 entries. This may cause some limitations for our models learning process.

There is a chance that the first model will not reach accuracy and performance standards. If this is the case, the additional models will be evaluated. Since this is the first AI project of Loans Today, any model implemented will be used to supplement and streamline the Loan Officer’s decision-making, not solely make the decision on its own. In future, the company may decide to rely more heavily on the model for decision-making.

References

Ranjith, K. (2020). *Loan Approval Data Set*. Www.kaggle.com.
<https://www.kaggle.com/datasets/granjithkumar/loan-approval-data-set>