

Loan eligibility project

In this project, I built machine learning models to predict the eligibility of intended loan borrower. These models could help money lenders and investors automate the loan eligibility process (real time) based on customer detail provided while filling online application form. These details are Gender, Marital Status, Education, Number of Dependents, Income, Loan Amount, Credit History and others.

I used LendingClub dataset with 614 loans and 13 variables for each loan.

In training the models, I only used features that were readily available on the table. These features included, among others, the borrower's gender, marital status, education, loan amount and credit history. Credit history was found to be the variable that affected loan status the most with 80% of those recorded to have a credit history being eligible.

The modeling process took several steps, including: removing loan features with significant missing data, or that aren't known to investors; exploring, transforming, and visualizing the data; creating dummy variables for categorical features; and fitting four models: logistic regression, random forest, k-nearest neighbors and support vector machines. I use machine learning pipelines to combine imputation, standardization, dimension reduction, and model fitting into one pipeline object.

I found that the four models performed similarly well according to the jaccard similarity scores on the testing data.

Algorithm	Jaccard	F1-score	Log Loss
<i>KNN</i>	0.81	0.78	NA
<i>Decision Tree</i>	0.79	0.77	NA
<i>SVM</i>	0.79	0.77	NA
<i>Logistic Regression</i>	0.80	0.77	0.57

I chose **K nearest neighbors** as the final model, which obtained an jaccard score of 0.81 on a test set consisting of the most recent 30% of the loans. This proved to be the highest scoring model.

I also found that the most impactful variables for predicting charge-off was the credit history variable.

All the analysis is done in a Python Jupyter Notebook, utilizing the packages numpy, pandas, matplotlib, seaborn, and scikit-learn.

