In this project, I build an algorithm to predict whether a particular in lending club will be a success or a failure. A successful loan is a loan that is fully paid off. A failure loan is a loan either in default state or charged off state, where there is no reasonable expectation that the loan will be paid off.

In loans funded in Lending Club, about 15 percent of loans are a failure, and 85 % percent of the loans are successful.

For loans of a large amount, the failure rate is particularly high.

The goal of my project is to predict whether a loan will be successful or not.

To do that I downloaded data from lending club, selected several features to build a prediction model, including both numerical features, such as the annual income, the fico credit score, and also categorical features, such as the grade of the loan assigned by lending club.

Using these features, I built a random forest model to predict whether the loan will be successful.

The idea is to build a large number of decision trees, each is used to generate a prediction, and the final prediction is the average of these predictions from each tree.

The strategy by my random forest model significantly decreases the chance of funding a failure loan, compared to the strategy by lending club, with a improvement of 40 %.

Moving forward, I would like to incorporate more features to further improve the model. One feature I am particular interested in the description of the loan provided by the borrower. It contains lots of details about the borrowers. I would also like to build a user interface to provide recommendation about whether or not one should fund a particular loan.