Credit Approve Using Machine Learning and Python

One-of the largest business sector for any financial institutions is mortgage. It is very important to determine which applicant to approve and which to reject. This decision often affects the company's revenue performance and sometimes even determines the company's life and death.

However, it would be difficult for every business to investigate every aspect of the credit record of an individual customer in today's society. So, our goal is to use a few features to create a model for the decision-making process. In our vision, the model should help to filter out weak and fraudulent applicants and therefore increase the investment returns for the company and reduce the occurrence of bad debts.

The dataset we use for this project is about 21st Mortgage Corporation in California, codenamed 549300XQVJ1XBNFA5536. We filtered out many variables that were considered to have less impact on loan approval, and considered only the following popular variables:

action_taken: The action taken on the covered loan or application

loan_type: The type of covered loan or application

lien_status: Lien status of the property securing the covered loan, or in the case of an application, proposed to secure the covered loan.

construction_method: Construction method for the dwelling

ageapplicant: The age of the applicant.

derived_race: Single aggregated race categorization derived from applicant/borrower and co-applicant/co-borrower race fields.

derived_sex: Single aggregated sex categorization derived from applicant/borrower and co-applicant/co-borrower sex fields

derived_dwelling_category: Derived dwelling type from Construction Method and Total Units fields for easier querying of specific records.

Since this dataset contains large pools of information of the applicants across many perspectives, we are confident to obtain insight into general trends including potential new customers and fraudulent patterns. By narrowing down our features to 7 variables and apply proper feature engineering, we can eliminate less powerful predictors and extract valuable information more efficiently and build a model that is both accurate and practical.