Finally Report

By Ibrahim Nasser Alothimein

Machine-Learning-Bank-Loan-Dataset

Table content :

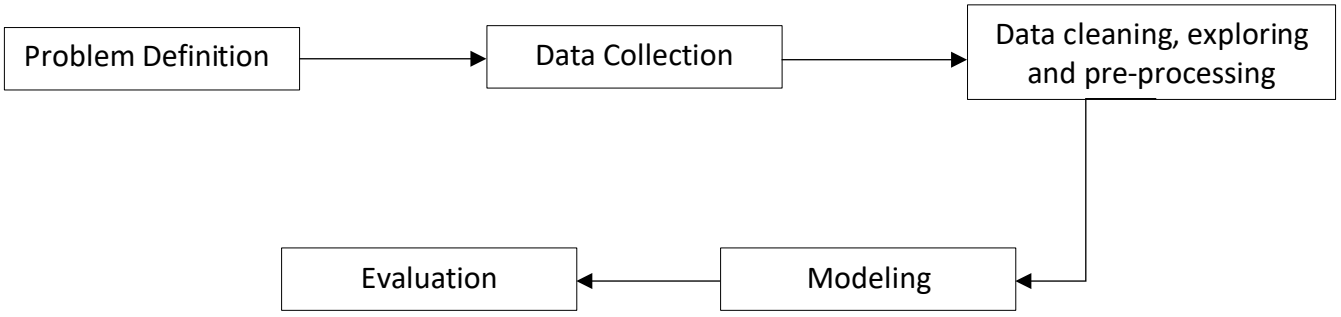
Abstract

This project will use the 'Bank Loan Status Dataset' uploaded by Zaur Begiev on Kaggle (Begiev, Z. 2017), which contains lending data for a particular lending institutions's historical accounts that are no longer active - meaning that the accounts have closed either due to the loans having been repaid, or that the loans were written off.
The purpose of this project is to use this bank loan dataset to predict whether a borrower will repay their loans or not based on the dataset's descriptive features. The target feature of interest will be the 'Loan Status', which determines whether a borrower has fully repaid their loan, or have been charged off (when a bank has determined that the borrower will never repay their loan).

Data

The Bank Loan Status dataset is loaded and assigned to df. As seen below,
the dataset contains 100,514 observations and 19 features.
Looking at the data types for each features are , 'Loan Status', 'Term', 'Years Current Job', 'Home Ownership', and 'Purpose' are objects, while the remaining features are floats. This is not entirely appropriate as 'Months Since Last Delinquent', 'Number Credit Problems', 'Bankruptcies', and 'Tax Liens' are ordinal features and should not be floats. This will be dealt with in later sections. Additionally, loan and customer IDs are objects due to containing letters, however they will be examined more closely in the next section.

Design



Algorithm

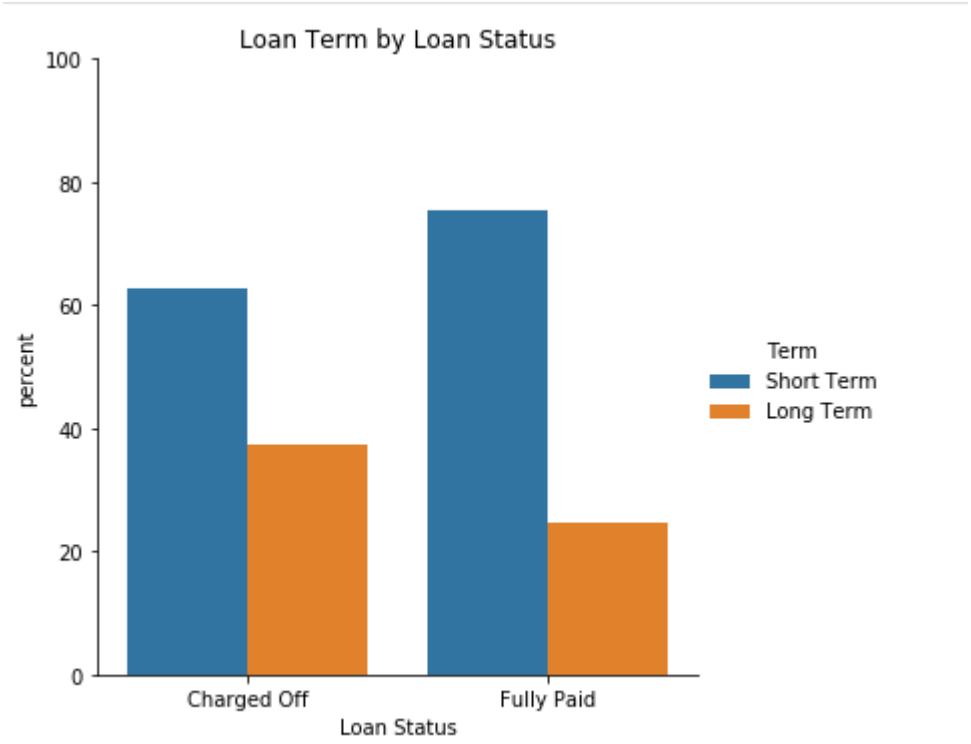
To predict the target feature, the following binary classifiers are considered:

- K-Nearest Neighbours (KNN).
- Decision Tree (DT).
- Naive Bayes (NB).
- Logistic Regression (LR)
- Random Forest (RF).

Tools

- K-Nearest Neighbours (KNN).
- Decision Tree (DT).
- Naive Bayes (NB).
- Logistic Regression (LR)

Communication



When comparing loan status and loan term, it seems that there is a higher proportion of long term loans for those who have been charged off compared to those that fully paid. This suggests that borrowers tend to not meet their loan obligations when the repayment period is longer (which usually means a larger loan).