

CAPSTONE PROJECT 1 – PROJECT PROPOSAL

1. Problem to Solve:

Many borrowers of home loans default or fail to repay back the due amounts on time and acquire a default status, which is a value corresponding to the number of days the borrower is delinquent. This value is based on the due date of last paid installment reported by services to Freddie Mac. The objective of this project is to predict this default status of new home loans that Freddie Mac is looking to acquire and hence help them make an informed and data-driven decision about whether to buy or reject a loan.

2. Client and why this project is important for them:

The client here is Freddie Mac which is a Federal Home Loan Mortgage Corporation. This project is important for them as they have to continually decide if or not to buy a home loan by assessing the risk involved in the venture. One of the primary risks involved in this business is borrows defaulting on loans, and therefore predicting their default status beforehand at the tie of acquiring the loan will reduce the risk significantly.

3. Data:

Freddie Mac releases two datasets – ‘Origination Data’ which is the data on loans it acquires, and ‘Performance Data’ which is the data on how those loans perform over time. ‘Origination’ data, which is published when the loan is acquired by Freddie Mac, contains information on the borrower and information on their loan and home. ‘Performance’ data, which is published every quarter after the loan is acquired, contains information on the payments being made by the borrower. Some of the seemingly important features are “credit score”, “original interest rate”, “loan term” from the Origination table and “default status” from the Performance table.

The data is acquired from the Freddie Mac website (link: http://www.freddiemac.com/research/datasets/sf_loanlevel_dataset.html) where the datasets are available for every year and in “.txt” format. All the datasets are downloaded and read in as csv files for the purposes of analysis.

4. Outline of Solution:

The purpose of this project is to classify if a loan will be defaulted on or not and by what measure. Hence a classifier machine learning algorithm will be a right way to start with the analysis. Classification can be done using Logistic Regression, Decision Trees or Support Vector Machine algorithms to name a few.

5. **Deliverables:**

Deliverables for this project include a Github Repository containing all the code written for prediction in Jupyter Notebooks and a Visualization Dashboard displaying the summary of the project and the end-results.