

# PROJECT PROPOSAL

## Credit Risk Analysis

**TITLE:** Credit Risk Analysis

**NAMES:** Chahhat Jagdish Pandit, Vijay Kinnera Thota, Aatusa Mehdiyan

### ABSTRACT:

Financial institutes like banks spend significant effort in identifying credit-worthy consumers for lending. The objective of the credit risk analysis is to assess the borrowers' creditworthiness by quantifying the risk of loss to which the lender is exposed. The probability of default, loss due to default, and exposure to default are the three measures that lenders use to measure credit risk.

This is the data set of mortgage observations for 50,000 residential U.S. mortgage borrowers over 60 periods, 600,000 records in total. The data set is a randomized selection of mortgage-loan-level data collected from the portfolios underlying U.S. residential mortgage-backed securities (RMBS) securitization portfolios and provided by International Financial Research ([www.internationalfinancialresearch.org](http://www.internationalfinancialresearch.org)).

We have applied SVM, Random Forest, KNN, Logistic Regression Classification models to determine whether the borrower is able to pay their loan(0) or not(1) at the end of the maturity period.

### DATASETS:

Key variables include:

- id: Borrower ID
- time: Time stamp of observation
- orig\_time: Time stamp for origination
- first\_time: Time stamp for first observation
- mat\_time: Time stamp for maturity
- balance\_time: Outstanding balance at observation time
- LTV\_time: Loan-to-value ratio at observation time, in %
- interest\_rate\_time: Interest rate at observation time, in %
- hpi\_time: House price index at observation time, base year = 100
- gdp\_time: Gross domestic product (GDP) growth at observation time, in %
- uer\_time: Unemployment rate at observation time, in %
- REtype\_CO\_orig\_time: Real estate type condominium = 1, otherwise = 0
- REtype\_PU\_orig\_time: Real estate type planned urban development = 1, otherwise = 0
- REtype\_SF\_orig\_time: Single-family home = 1, otherwise = 0
- investor\_orig\_time: Investor borrower = 1, otherwise = 0
- balance\_orig\_time: Outstanding balance at origination time
- FICO\_orig\_time: FICO score at origination time, in %
- LTV\_orig\_time: Loan-to-value ratio at origination time, in %
- Interest\_Rate\_orig\_time: Interest rate at origination time, in %
- hpi\_orig\_time: House price index at origination time, base year = 100
- default\_time: Default observation at observation time

- `payoff_time`: Payoff observation at observation time
- `status_time`: Default (1), payoff (2), and nondefault/nonpayoff (0) observation at observation time

### **Maturity Time Period**

In this dataset there are two variables - `time`: Time stamp of observation and `orig_time`: Time stamp for origination.

Based on these columns we have derived a column `maturity_time_period_days` which is the difference between these two variables. By adding this column we can know how many days the loan has been paid or how many days are there for the loan to reach maturity.

### **First day of observation during Maturity Time Period**

The variable `orig_time` has values of when the observation of the loan started. So, we added a column named `start_day_observation`. This new column contains the subtraction of `orig_time` and `first_time`. This new column has the number of days on which the loan observation started during the Maturity Time Period (`maturity_time_period_days`).

### **REFERENCES:**

- [1]. Trilok N Pandey, M. Suman Kumar, J. Alok Kumar, Satchidananda Dehuri, "Credit Risk Analysis using Machine Learning Classifiers" International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017).
- [2]. Cheng-Lung Huang, Mu-Chen Chen and Chieh-Jen Wang, "Credit scoring with a data mining approach based on support vector machines", Expert systems with applications, vol. 33.4, pp. 847-856, 2007.
- [3]. S.J. Shiv; Srinivasa Murthy; Krishnaprasad Challuru, "Credit Risk Analysis Using Machine Learning Techniques" 2018 Fourteenth International Conference on Information Processing (ICINPRO), 10.1109/ICINPRO43533.2018.9096854