

Project Proposal

Project Title: Logistic Regression for Credit Risk Assessment Using Credit Applications

Project Summary: I propose to analyze the Kaggle Credit Data Set for a credit risk assessment using logistic regression or gradient boosting based off 11 data points per applicant and then create a risk assessment prediction based off my logistic regression model.

Dataset Description: The Kaggle Credit Data Set is an open-source data set that includes thousands of applicants for new lines of credit. Each applicant in the dataset contains the following information:

- SeriousDelinquencyIn2Years → if applicant has experienced 90 days delinquency or worse
- UtilizationofUnsecureCreditLines → Total balance on credit cards and person lines of credit divided by the number of credit lines
- Age → applicant age in years
- NumberofPastDue30-59Days → number of times the applicant has been past due between 30-59 days
- DebtRatio → Sum of debt payments divided by monthly gross income
- MonthlyIncome → Gross monthly income
- NumberofOpenCreditLines → Number of open loans
- NumberOfTime90DaysLate → Number of times applicant has been 90 days or more late
- NumberOfRealEstate → Number of specifically real estate loans
- NumberOf50-89DaysLate → Number of times late between 50-89 days
- NumberOfDependents → Number of dependents in family

Using these features of the data set for thousands of applicants for a new line of credit, I propose to use logistic regression model to predict how risky a candidate would be for future credit.

Data Mining Algorithm: After learning the dataset, I propose to predict a new applicants risk level for a new line of credit based off a logistic regression model and if all goes well additionally based off a gradient boosting algorithm, learned from the data set from Kaggle that will predict my future credit applicants.

Data Analysis: I propose to create an interactive program where an applicant's information can be entered, and a user can instantly see their risk assessment, based on their applicant information. I believe this would be similar to how credit institution operates today and would be useful for predicting which factors are the most influential when assessing the risk of an applicant.