Abuabbas Qureshi

Capstone Project 1 Proposal

The project I will be working on will be based off a Lending Club loan application dataset. This project will investigate the risk factors for an applicant. The idea is to determine which of the applicants could be risky clients to grant a loan to. This would be done primarily through exploratory data analysis and touch upon basic machine learning concepts to derive an answer to the main question. While pursuing the question, I am interested in finding out some more supplementary answers to questions such as what kind of applicants who have had delinquencies in the past still prove to be good clients, what kind of applicants who are high earners still prove to be risky clients. I would also be very interested in finding out whether location would be a factor in determining the risk factors in a loan applicant. I am also curious to see whether the 2008 financial crisis has anything to do with the success or failure of a loan application. Some of these queries would be whether a high income applicant suddenly fell on hard times or not, or whether a low income applicant has a better chance of prospering now.

Along with this, I believe I will have to perform some data munging to be able to perform feature selection better. Apart from data munging, there will be plenty of data wrangling to get an in depth analysis on this data set. The variables are plenty however prior to starting this project I have see many of them containing missing values over 70%. The TARGET variable will be determined from the ‘loan\_status” column, I will differentiate between a ‘Good Loan’ and a ‘Bad Loan’ according to the values in the column. I suspect apart from “Fully Paid” and “Current” the other values will be bad loans. This will be determined once I begin the analysis.

For the machine learning aspect of the project, I will be applying three algorithms to build the predictive models. These will be Logistic Regression, Decision Trees, and Random Forests. These are classification based algorithms that will help us achieve our goal in this project.