**Capstone Project Proposal**

Lending Club is no new name in the lending industry. It has shifted the lending paradigm with the use of technology devising efficient, convenient and the smartest way about investing and borrowing.

It has revolutionized peer to peer loan lending platform by operating a credit marketplace which is overpowering the loan programs offered by traditional banking institutions. With their lending model, borrowers can take advantage of lower interest rate where as investors can also be benefitted with better returns on their investments.

With Lending Club, an investor can invest in a portfolio of loans. But how will those loan perform, and which loans to pick from? In order to say anything meaningful about what loans to choose, we must first estimate how loans will do over time. What percentage of loans will default? What percentage of loans will get paid off in full?

Through this project we are interested in making a guess at the probability of default, which is intended to support company’s decision in approving/not approving the loans given the features that are collected from the lenders through loan application.

I can access company’s dataset which is publicly available on their site.

I plan to base my analysis on 2014-15 dataset which comes with over 650000 records and 122 features.

This is the basic flow I am going to follow.

Process Flow

• Exploratory Analysis

• Dealing with nulls / missing values

• Feature Cleaning

• Preparing Training/Test data set

• Applying Modelling techniques

I plan to use Python framework making an effective use of Panda library for basic data frame wrangling. To handle numerical calculations, we will use libraries like NumPy, SciPy. We will also use ggplot, matplotlib for exploratory analysis and visualization purposes. To benchmark our results against robust machine learning code, I can employ scikit-learn library, which offers out of the box functionality for the Logistic regression, kNN, Random forest and Naive Bayes algorithms implementation.