

Business Understanding Document

Daksh Varshneya(IMT2012012)

Praseeda Kalkur(PH2016008)

Ujjwal Jain(IMT2013056)

Udbhav Vats(IMT2013055)

Group 21

13th October 2016

1 Background Information

Lending Club is a marketplace to borrow and lend money. It allows a user to borrow money in the form of personal loans, business loans, and financing for elective medical procedures. It's an online platform with no physical infrastructure present. Technology helps them lower the cost and deliver an amazing experience. They optimize the interest rate by sharing the cost savings amongst the borrowers and investors in the form of attractive returns.

A borrower can get an instant quote in minutes with no impact on his credit score. Once he agrees on an interest rate the investor is invited to commit funds towards the loan application. An investor can open an account and build a portfolio of hundreds or thousands of loans made to quality borrowers. He will receive monthly payments of principal and interest, which he can withdraw or reinvest.

2 Assessing the Situation

This is the Lending Club's first attempt at extracting some useful information from the data and hence they have published a large chunk of their data on an open platform for users to find some interesting patterns in the data. Since, there is no direct contact to the publishers of the data, we cannot assess the company situation as of now.

2.1 Data

Since, Lending Club is a big player in this field of P2P borrowing and lending money, the data is quite extensive with plenty of dimensions to look upon. The study would involve looking at all loans provided through the platform from 2007 to 2015. The dataset can be accessed at [Kaggle Dataset web page](#). If the results are satisfying, the model can be tested on various other similar datasets.

2.2 Risks

There is no risk involved as this is academic study of the dataset with no scope of actual implementation in the real world.

3 Data Mining Goals

We plan to answer the following questions through various data mining tasks -

- Is there any correlation between annual income, loan amount and defaulting pattern (Exploratory analysis).
- Is there any correlation between debt-to-income-ratio(dti) and loan defaulting?
- Is there any correlation between interest rate applied on a loan and purpose defined for which the loan was requested?
- Find out the aggregated amount borrowed, number of loans approved across various states of America (Data Visualization)
- Given a new application of a borrower can we predict if his application will be accepted or not?(Predictive analysis)

4 Data Mining Success Criteria

The study would be considered successful if we are able to find relevant and interesting answers to the above targeted questions.

- Exploratory analysis would be successful if it helps in finding a target population for Lending Club to focus on. Specifically, finding correlations can lead us to even better models for predictive analysis tasks.
- While assessing our predictive model to check if the loan application should be accepted or not, an important criteria of success would be to bring down the number of false positives as from Lending Club's perspective false positives would be costlier.

5 Project Plan

Phase	Time	Resources	Risks
Business Understanding	2 weeks	All 4 members	Wrong business model assumptions
Data Understanding	1 week	All 4 members	Data Problems, Technology Problems
Data Preparation	2 weeks	All 4 members	Data Problems, Technology Problems
Modeling	2 weeks	All 4 members	Technology problems, inability to find meaningful correlations
Evaluation	1 week	All 4 members	Technology problems, Economic change, inability to find meaningful correlations results
Deployment	1 week	All 4 members	Technology problems, Economic change, inability to implement results