Abstract

With the development of the E business, more and more transactions are made on the internet and also the loaning. The basic topic that we are discussing is lending club, which is one of the most famous and have the largest scale P2P platforms in United States. Based on the raw data, we processed it by deleting redundant columns and observations.

Combined python with SPSS, we managed to figure out the factors that influenced most on loan status, which has only two results – fully paid or charged off. The innovation of the project is that we utilized the k-fold algorithm in python to optimize the sub grading criteria in order to increase the accuracy.

The result of the innovation turned out to be cool, which significantly increased the accuracy and made the chart generated better shaped and closer to the ideal trapezoid.

Introduction

The topic of our project is analyzing loan data from lending club. Since based on the report, the increasing rate of the company is getting flat so we are trying to figure out if the reason is that there are flaws inside the system.

The basic scheme of the platform is that it charges commission of every single transaction which means the lending club will benefit from the peer to peer loaning.

Also, they reduce the risk by splitting the loan into like twenty slices and make a large amount of money into smaller portions. There is another way that the platform implied to minimize the risk- grading and sub grading its clients based on the credit record as well as the other factors including annual income or employment.

It has been growing with a rapid pace since it was launched in 2007 and the profit has been increasing so fast since then.

This report presents the analyzing process of the lending club and manages to figure out the main factor that may influence the loan status of it. Later on further research will be introduced about modifying the sub grading criteria to minimize the risk and to keep standard deviation at a relatively low level.