

Proposal for Capstone Project 1

Introduction

Lending Club Data Set:

<https://www.lendingclub.com/info/download-data.action>

Lending Club provides data about loan applications' performance of loans that it issued. The data is from 2007 to the second quarter of 2017 including ten separate data sets. To avoid the economic depression period we choose the relatively recent data set in 2015.

Local Area Unemployment Rate in US:

<https://www.kaggle.com/jayrav13/unemployment-by-county-us>

<https://www.gaslampmedia.com/download-zip-code-latitude-longitude-city-state-county-csv/>

The first data is available on the website of Kaggle and represents the Local Area Unemployment Statistics from 1990-2016, broken down by state, month and county. There is no unemployment rate in US for Zip code level. So I chose the second data set which contains zip code list and corresponding county information. Linking the two tables together with Lending Club data sets I can explore the loan status' relationship with unemployment rate in different zip code area.

Problem to resolve:

To predict if loan performance associates with unemployment rate. And to build a model to predict the loan performance using machine learning algorithms.

Clients and Audiences

Bank, Insurance company, Financial industry may have interested in this kind of problem.

Approach

1. Downloading training set and test set from Lending Club website and US Labor Department.
2. Be familiar with the variables in the data set.
3. Cleaning data to prepare the data set for analyzing.
4. Applying Decision Trees, Random Forest, AdaBoost, logistical regression algorithms to build classifier for the prediction of loan performance.

Project Deliverables

The project deliverable will contain following:

- This document explaining approach of the project.
- Presentation created on this project.
- Python code for data wrangling with comments and details.
- Graphs and visualizations created for this project.
- Python code for model building with comments and details.