Capstone Project Proposal

US Consumer Finance Complaints Text Mining

Problem

Consumer Financial Protection Bureau is an government agency,it helps consumers complaints heard by financial companies.the goal of the project is to study and identify the inappropriate practices and allowing the government to stop those before it becomes a major issue.This project focuses on the analysis of the complaints over different segments,also providing sentiment analysis of the complaints.

Client

This project benefits CFPB and the fiancial companies to identify the major issues and to find the areas of improvement to better serve the customer.

About the data

The Consumer Complaint Database is a collection of complaints on a range of consumer financial products and services, sent to companies for response. It started receiving complaints from July 2011.The database generally updates daily, and contains certain information for each complaint, including the source of the complaint, the date of submission, and the company the complaint was sent to for response. The database also includes information about the actions taken by the company in response to the complaint, such as, whether the company's response was timely and how the company responded.

Approach

Dataset used for analysis is US Consumer Finance Complaints data from Kaggle.

- Importing and Reading the csv file for further analysis,is the first step in data analysis
- There are 18 variables.Cleaning up the data is a important challenge. This particular dataset contains variables that needs to be converted to factors. Need to reformat dates
- Exploratory Data Analysis step involves the distribution of complaints based on company,product,issue and states
- Performing the sentiment analysis to identify the major issues.
- Analysis of different Text Mining pacakges like tidytext,quantenda and text2vec and identify the best suitable methods for the project.

Deliverables

The final deliverables are

- R code
- R markdown document
- Slide deck