**Analysis of Consumer Complaints**

Web Textual Mining

Zoya Ansari, Dalya Khatun, Maqbool Ansari

Table of Contents

Proposal3

Motivation3

Background 3

Abstract4

Project Analysis5-6

Supporting Details7

**proposal**

We propose to collect data regarding consumer complaint narratives during a range of three or four years in order to analyze the consumer complaints through the use of web textual mining and represent the analyzed informational data in the format of tables, charts, and graphs.

Motivation:

In this project, the motivation is to provide useful and valuable information to companies to further improve their profits, revenue, and customer satisfaction. The goal is to be able to extract and produce valuable information from unstructured textual data.

Background:

In 2011, Congress created the CFPB to ensure the protection of consumer interests in many financial markets. The CFPB receives and processes consumer complaints related to a variety of financial services, including credit cards, mortgages, bank accounts, student loans, consumer loans, credit reports, and debt collection. This database is a public collection of consumer financial complaints that includes basic information about the complaints such as submission date, consumer’s zip code, the company, the product type, the consumer narratives, and how the company has responded to the complaint. Making the data publicly available not only allows the financial institutions and their consumers to view the quality of their financial products and services, but also encourages public users such as analysts, data scientists and others to explore this information, and build on valuable comprehensions accordingly. The data does not include the analysis of *consumer complaint narratives* that certainly would provide more valuable information. As a result, the consumer narratives provide context to complaints upon which more interesting information could be gathered. This information can be extracted by a deep analysis of the narrative data using text mining techniques. Hence, the project will determine the benefit of text mining methods in analyzing consumer complaint narratives.

Abstract:

As technology is advancing, the complexity of Websites and the information within is advancing as well. Web pages contain numerous amounts of information that may be beneficial to that particular company. For example, various companies allow their customers to relay their input through the use of comments or complains in their website. Web text mining is a particular tool that is extremely helpful in not only extracting useful information out of all the text but also in formatting the data accordingly into charts or graphs. Text mining allows business owners to comprehend and predict valuable information. Web Text mining processes unstructured textual data and outputs easy to access and read information.

Businesses further their brand through improvements, which are most likely recommended by their consumers. Consumer complaints or recommendations allow business owners to gain valuable information to improve their brand. The research paper applies to the application of Web Text Mining to not only reproduce useful information from textual data but also represent the data in such a manner that is helpful to business owners who wish to further improve their company brand.

The data named Consumer Complaints was collected from Data.Gov. The data was stored and processes as a CSV file. The Consumer Complaint narratives stored in the data were collected from the years 2013 to 2016. Web Textual Mining is an extremely useful process to extract valuable information from unstructured data that can be represented in a variety of ways such as tables, charts, graphs, and many other alternatives.

Project Analysis:

The text mining method allows analysts to recap automatically unstructured consumer complaints into an assortment of topics through which the principal semantic structure would be discovered. This assignment would not be possible by human annotations as it requires reading a large volume of consumer narratives and identifying semantic similarities among them, a process both arduous and time-consuming.

The first step in the proposed outline is data preprocessing to prepare the text data for analysis. Characteristically, text data is noisy and unstructured. Data preprocessing provides a stable format for the data, and removes non-critical words that are not contributing to the text analysis. Our data preprocessing includes the following five tasks: (1) convert to lowercase, (2) remove special characters and tokenize them into terms, (3) remove stop words, (4) stemming, and (5) construct term-document matrix. The output of preprocessing is the term-document matrix representing the frequency of the critical terms in the consumer complaints.

The second step in the proposed outline is running text mining analysis on the term-document matrix to discover the pattern of the word use in the documents to construct treasured date through graphical representation.

As each consumer complaint could cover multiple problems (e.g., credit reporting and harassment), text mining can be used as an efficient method to take these issues by summarizing each complaint into a variety of topics.

The proposed outline absolutely can be useful in observing the consumer complaint narratives for developing topics. It can be prepared easily to automate the prediction of topic assignments of future complaint narratives. New evolving topics or issues can be determined through the process previously stated through data collection and analysis for a period of time. The proposed text mining outline is never projected to replace human analysts. Instead, it is designed as a corresponding analytic tool for the existing analysts to examine consumer complaints more efficiently and effectively, eventually improving consumer protection from unfair, deceptive or abusive practices in the financial markets.