Research Proposal - CFPB Consumer Complaints Priority Ranking

Nithil - Smrithi - Spencer - Ahaan - Christian - Madeline - Gilbert - Tamas - Adhish

Abstract

This paper introduces a sophisticated tool for prioritizing consumer complaints in the Consumer Financial Protection Bureau (CFPB) database, utilizing Facebook AI's large language model, RoBERTa. We fine-tuned RoBERTa on a manually annotated subset of complaints to identify those with legal significance, and then fine-tuned two separate instances of the model to predict company public and consumer responses. An ensemble approach combined allowed us to create three models that specialized in one natural language task. These insights, combined with complaint timings, resulted in an algorithm that prioritizes complaints based on legal relevance, predicted company public/consumer response, and age. This method significantly allows complaints to be handled in an order cognizant to how the complaint is likely to be handled. Leveraging NLP could optimize financial institution response quality and resource allocation. The study demonstrates the efficacy of large language models in refining customer service operations in the financial sector.

1 Introduction

The Consumer Financial Protection Bureau is a government agency that writes and enforces rules for financial institutions, examines financial institutions, monitors and reports on markets, and importantly, maintains a database of financial product and service complaints. This database holds consumer complaints alongside information like the company type and company response. Complaints in this database are often deemed invalid after submission with accordance to federal regulations. The goal of this project is to develop a complaint screening tool, providing financial entities a prioritized ranking of outstanding complaints so they can address the most pertinent first. This should provide a better service to consumers with valid concerns and save company resources by addressing important complaints promptly. To achieve this we leveraged the CFPB Complaint Database to fine tune three large language models for different purposes. We begin by manually annotating complaints by the presence of a legal basis, and creating a truth set for the LLM. We then fine tuned Facebook AI's RoBERTa on this truth set to create a domain specific model, accurate at labeling complaint legality. Following this, we use a balanced sampling of two types of company responses to create two further fine-tuned models. In using an ensemble approach, we have three models that specialize in in different purposes. The outputs of these models as well as the time of the complaint are given a weight, creating the scheduling algorithm that determines priority and sorts the complaints accordingly.

2 Background

2.1 CFPB Background information

Our study utilizes data collected by the Consumer Financial Protection Bureau (CFPB), a U.S. government agency dedicated to making sure consumers are treated fairly by banks, lenders and other financial institutions. The agency owns a database containing consumer complaints to these institutions. A complaint is formally defined as an issue or dissatisfaction with reason towards a specific company's services or products. When a complaint is made, it is directed towards the company itself, usually the customer services department. They are then responsible for investigating the issue, communicating with the consumer, and attempting to resolve the complaint based on their internal procedures and policies. Two options arise. If the consumer is satisfied with the resolution provided by the company, the matter may be considered resolved and the consumer can be considered "satisfied". However, if the consumer believes their complaint remains unresolved, he/she may choose to escalate the issue further by filing a complaint with the CFPB. This is important to note as complaints reaching this database are coming from customers unsatisfied with the initial resolution.

One specific thing to note about the CFPB is that a companies frequency of complaints can't be used as a sole predictor of regulatory compliance. Factors like complaint severity, company size, market share, and geographical population must be added to the equation. To get a complete picture, it would be beneficial to combine complaint data from other public/private databases.

The CFPB maintains the financial complaint database to analyze the data and provide transparency for consumers, researchers, and the general public. The CFPB performs their own studies of the data to identify areas of concern and prioritize regulatory actions. The public can also search for information about complaints related to specific financial products or services, and financial institutions.

3 Methodology

3.1 Data Collection

We retrieved complaint data from the CFPB Consumer Complaint Dataset. The dataset can be accessed by API or CSV/JSON Download and can be filtered through a graphical interface before download. We downloaded the entirety of the complaint database and created three different data sets for our ensemble model approach. We first sampled 900 user complaints for legal-basis annotation. We then created two cleaned data sets for the training and testing of the multi-class classification models. These data sets included only complaints relevant to that model's purpose, increasing training efficiency. For instance, complaints with the public response "Company chose not to provide a public response" were not used to train the public response classification model.

3.2 Data Annotation

To annotate the data, we split the sample of 900 complaints into three sections of 300 complaints. To minimize bias, the complaints were sectioned such that each was annotated by three different team members. Each complaint was given two columns "Law Related" and "Fraud Related". If a complaint had some sort of legal basis (IE it was not a petty concern) we marked it "Law Related", and if it was related to some sort of fraud (Identity Theft, etc.) it was marked "Fraud Related". Using a Python script, we recombined these annotations taking the majority opinion of the three annotators, creating a truth set to fine tune RoBERTa.

3.3 Legal Basis Classification

We first trained RoBERTa using our custom labels in order to better label the complaint narrative entries by legal basis. This allows the filtration of personal gripes with employee service representatives, company application capabilities, and other less pressing issues. The training dataset consisted of 810 annotated complaints, and the remaining 90 complaints made up the testing dataset. To enhance the performance of our algorithm, we used AdamW and the loss function. AdamW was used to correct the weight decay regularization method. The loss function helped measure the disparity between our model's predicted values and the true values. One change we made to fine tune the model was integrating learning-rate scheduling to adjust the learning rate during training. This helped the model converge faster and avoid overshooting. Another change we made was decreasing the batch size from 32 to 16. This helped improve results as smaller batch sizes are better suited for smaller data sets. These two fine tuning changes to our code resulted in a 3% increase in performance from an initial 86% accuracy.

3.4 Company Response Prediction

To predict company responses to consumer complaints, the initial dataset was subjected to a series of preprocessing steps to ensure data quality and relevance. Two sets were selected representative subsets of the data to streamline the analysis process for the two multi-class models. These subsets were divided into training, testing, and validation sets to facilitate the development and evaluation of our multi-label classification models. Datasets were handled using Python's 'pandas' library and further processed using Hugging Face's 'datasets' library for efficient manipulation. To further enhance the model's generalizability, the test set was divided equally into two shards, one serving as the actual test set and the other as the validation set.

We then fine-tuned the pre-trained RoBERTa model to adapt to the specifics the CFPB Dataset. This fine-tuning process allowed RoBERTa to learn the intricacies and patterns that caused companies to choose different public/consumer responses. Each complaint was associated with one or more types of company responses. To further optimize the model's performance, we engaged in hyperparameter tuning. This included adjusting learning rates, batch sizes, and the number of training epochs. The aim was to find a sweet spot where the model achieved the highest accuracy without overfitting the training data. The performance of the model was rigorously evaluated using a disproportionately large testing dataset. Evaluation is faster than fine-tuning, so we sought to optimize our f1 score over a large testing set while training on only a few thousand complaints. F1-score was calculated to assess the model's effectiveness in correctly classifying the types of company responses. Through this process, we were able to achieve an f1 score of .75.

3.5 Priority Ranking

Using the 3 models previously described and the age of complaints, we created a program capable of sorting the complaints in a priority order. This automatic ranking of complaints has potential to save financial institutions time and money by allowing them to address the most important complaints first. We created a weighting function to act upon the model assigned legality, likely company responses, and age of a complaint. We can now feed randomly ordered csv files of complaints to our program, and receive an output csv with the complaints sorted and a new column "Has Legal Basis?" as determined by the legal basis classification model.

The weighting function was made due to heavy assumptions from our team as to how a financial institution may want their complaints sorted.

3.6 Spearman's Rank-order

To test the robustness of our sorting algorithm, we compared it to a human-like ranking. We tested it on a ranking of 15 complaints (see next subsection for more details). We used Spearman's rank-order to determine the likeness between two rankings. Spearman's rank correlation coefficient (ρ) is calculated using the following equation:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

Here, ρ is the Spearman's rank correlation coefficient, d_i is the difference between the ranks of corresponding pairs of variables, and n is the number of data points. Just like the Pearson equivalent (which tests for linear correlation), the test will yield a figure of between -1 and +1, and the closer the figure is to 1, the stronger the monotonic relationship. Spearman's coefficient is appropriate for both continuous and discrete ordinal variables, which is why we used it on our ranking system.

3.7 Prioritized Ranking Example

Figure 1: Here we have the partial output (hiding pre-existing and irrelevant columns) of a run of our priority ranking program when fed a random sampling in a random order of 2 months of complaints

Date received	Consumer complaint narrative	Has Legal Basis?
8/14/2023	All negative and non beneficial marks on my credit report are illegal so all of the negatives on my report shou	Yes
8/14/2023	When I investigated my credit report, I realized that some of the information was erroneous. XXXX and XXXX	Yes
8/31/2023	By the provisions of the Fair Credit Reporting Act, I am writing this letter to request that the listed account w	Yes
8/14/2023	as a over the XXXX XXXX XXXX I have my personal information stolen from individuals that i let live in my hom	No
8/16/2023	My identity was stolen and used fraudulently. I dispute the account and the debt attached to it. I have since	No
8/24/2023	I was denied services by an airline company and reported the same to the credit card company and requeste	No
8/31/2023	I don't want to see this on my credit report, I'm begging you this is not mine, This is totally Fraud. Please take	No
8/30/2023	XXXX XXXX accepted my 1099 C and sent me copy B for my student loan tender of payment and then they tri	No
9/3/2023	XXXX XXXX sent me a statement that my mortgage was going from XXXX to XXXX due to my escrow being sh	No
8/25/2023	I am filing this complaint on my own behalf to address an issue concerning late payments on my credit recor	No
8/26/2023	Requested contract with my wet signature and company failed to provide this information. Therefore the del	No
9/19/2023	I did not apply for any credit with these companies, nor did I authorize these hard inquiries. These inquiries	No
9/13/2023	The company refused to honor my privacy protection. They said they was going to report what they want wit	No
9/21/2023	I bought my car in XX/XX/2023 with a high interest rate of over 25 %. Credit acceptance is getting sued for hi	No
9/6/2023	The existence of a derogatory rating on my account is causing me significant concern. I am deeply worried ab	No

4 Conclusion

4.1 Findings

The Spearman's rank correlation coefficient that we calcualated from our random sampling of 15 complaints came out to be 0.52. This is indicates a moderately strong likeness between our human ranking and the program's ranking. We acknowledge that there are still improvements to be made to the program's ranking.

4.2 Further Work

There are a few potential avenues to further our work. The largest step forward would to be to partner with a financial institution which could benefit from our work. If industry professionals ranked a series of complaints, then we could train a neural net to form a better weighting algorithm to more accurately combine our model outputs to create sorting weights for the final output. Additionally, we would be able to compare our algorithm to any currently existing method for the ordering of addressing complaints. Partnering with such a company would also allow us to apply our work towards the initial complaints received by these firms rather than only the complaints elevated to the CFPB.

Another step we could make moving forwards would be to try and identify trends in complaint data vs. stock market data. However, some trends may be masked by the fact that only complaints raised to the CFPB would get monitored.

Lastly, we will note that we are currently limited by our use of RoBERTa in that the maximum size of the complaints we can rank is 512 characters. It is not too infrequent to receive complaints beyond this size. Moving forward, we could workout ways to address such issues, likely leveraging models capable of larger character sizes.

5 Code

5.1 Google Colabs

- Priority Ranker
- Legal Basis Classifier
- Public Response Classifier
- Consumer Response Classifier