CONSUMER COMPLAINTS CLASSIFICATION USING NATURAL LANGUAGE PROCESSING

A PREPRINT

Jithendra Bojedla jbojedla@pdx.edu

Rutuja Padgilwar rpadgil2@pdx.edu

Dileep Kumar Boyapati dileepkb@pdx.edu

December 4, 2023

ABSTRACT

Develop an accurate and effective Natural Language Processing model that aims to address real-world challenges in consumer complaint classification. The primary objective is to enhance the Consumer Financial Protection Bureau(CFPB) capability to handle consumer complaints effectively. Focused on classification, the project employs NLP models to categorize text into categories. Five baseline algorithms are trained and validated, with particular emphasis on their performance in accurate categorization. Among these algorithms, XGBoost stands out with a notable accuracy of 84 percent, showcasing its effectiveness in handling different numbers of instances. A pre-trained language model named BERT is also utilized to classify the complaints category.

Keywords: Consumer Financial Protection Bureau(CFPB), Baseline Models, BERT, Accuracy.

1 Introduction

In the world of consumer finance, the Consumer Financial Protection Bureau (CFPB) is dealing with a huge number of complaints every day. This flood of various complaints has made it crucial to create an application that can handle them all automatically. Consumer complaints classification means classifying the nature of the complaints that are reported by the consumers. The main objective of the project is to develop and implement an accurate Natural Language Processing (NLP) model to categorize consumer complaints effectively and expeditiously, thereby alleviating the manual burden faced by the CFPB. This project will be really helpful for the consumer care departments as they receive thousands of complaints regularly.

Motivation:
The main motivation of our project is the challenge that is currently facing by many financial service companies and their consumers. The motivation stems from the laborious and time-consuming nature of manual complaint classification. An NLP model presents a scalable solution to enhance the speed and accuracy of categorization.

2 NLP Task

The research focuses on the NLP task of multi-class text classification. Consumer complaint narratives serve as inputs, and the model classifies them into five consolidated product classes: Credit Reporting, Debt Collection, Mortgages and Loans, Credit Cards, and Retail Banking.
Example -
Input Narrative Sample: Attempting pay monthly student loan payment week account become locked attempts reset password username blocked well tried company logged account access form provided number unlock account option main menu unlock account speak customer service representative entire menu automated clear information speak anyone unlocking account payment could made option pay check phone option still left locked account unable check balances engage correspondence company.
Output: mortgages and loans

3 Dataset

3.1 Data Overview

For this project, we used a real time dataset imported from Consumer Financial Bureau, an official website of the United States Government. The dataset includes a decade worth of submissions that are made by the consumers from December 2011 to October 2023, totaling 4,208,806 complaints.
CFPB Official Portal for accessing consumer complaints database to view trends, read and export the data: https://www.consumerfinance.gov/data-research/consumer-complaints/

3.2 Data Preprocessing

Cleaning and Filtering:

- Complaints without consumer narratives were removed, narrowing down the dataset to instances with detailed descriptions.
- Duplicate narratives were identified and eliminated, resulting in a more refined dataset.
- After filtering out, a refined dataset of 1,285,865 complaints is obtained.

Categorization:

- Initially there were twenty one diverse set of product categories. After careful observation, we consolidated them into five main classes: Credit Reporting, Debt Collection, Mortgages and Loans, Credit Cards, and Retail Banking. This categorization aimed to simplify and streamline the classification process.

Text Data Preprocessing:

- Removed stop words and irrelevant punctuation from the narrative texts to focus on meaningful content.
- Lemmatization was applied to standardize words to their base or root form, reducing dimensionality and improving model performance.

Vectorization:

- The product category text data underwent vectorization, transforming words into numerical frequencies, making it suitable for machine learning models.

Balancing Classes:

- Initially the data was imbalanced, we balanced the data by randomly selecting the data of each category with the count of the product category which had the lowest count.This step aimed to prevent the model from being biased toward overrepresented classes.
- After balancing out, a refined dataset of 625,545 complaints is obtained.

The dataset was further split into training and testing sets to evaluate the performance of different machine learning models. We used many data sizes during the training process. Also, we filtered out the data by using the date parameter like from March 2020 to March 2023, etc.

4 Methodology

The steps involved in the workflow of complaints classification are :

1. Environment setup
    - The main objective is to prepare the development environment for the Natural Language Processing Project.

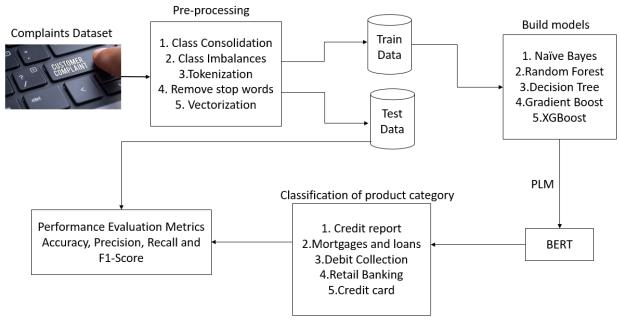

Figure 1: Architecture diagram

- Install necessary libraries to develop the models.

2. Gather the relevant data from the Consumer Financial Protection Bureau portal.
   - We accessed the official portal and downloaded the dataset in CSV format.
3. Applying various pre-processing techniques to prepare the data.
   - Handle missing data as well as duplicates in the dataset.
   - Multiple classes i.e. 21 product areas are grouped into five major classes by performing class consolidation.
   - Applied undersampling technique in order to balance the classes so that model is not biased towards one class.
   - Tokenize the text data to break it into individual words so that the model can understand the data better.
   - Focus on meaningful words by removal of stopwords.
   - Used TF-IDF vectorization to transform the text into vectors of numbers.
4. Divide the dataset into training and testing sets.
   - Randomly split the preprocessed dataset into a training set and a test set.
   - Based on the size of the dataset, we used an 80-20 split ratio.
5. Build baseline models for classification.
   - Implemented different types of baseline models such as Naive Bayes, Random Forest, Decision Tree, Gradient Boost and XGBoost.
   - Trained these models on the training dataset.
   - Evaluated the model performance using metrics like accuracy, precision, recall, and F1-score.
6. Worked on a Pre-trained Language Model(PLM) named Bidirectional Encoder Representation from Transformers(BERT).
   - Fine-tune the model on the training dataset.
   - Test the model's performance on the test dataset.
7. Model Evaluation.
   - Used a new and unseen dataset to evaluate the model performance.
   - Compared the performance of baseline models and the BERT model.

5 Results

Trained and validated different types of baseline models like Multinomial Naive bayes, Random Forest, Decision tree, Gradient Boosting and XGBoost. We make use of accuracy as the primary evaluation metric since the five classes are being balanced. Observed closely how well these models performed on both the training and test datasets to eliminate the overfitting. Different performances are noted according to the number of instances(140k,200k,400k) occurred. Notably, with 140k instances in particular, XGBoost consistently performed better than the remaining other models and showed higher accuracy. In comparison to the other algorithms, Decision tree consistently displayed less accuracy. As the dataset size expanded to 200k and 400k instances, XGBoost continued to outperform the baseline models in

Model	For 140K Instances		For 200K Instances		For 400K Instances	
	Validation Accuracy	Test Accuracy	Validation Accuracy	Test Accuracy	Validation Accuracy	Test Accuracy
Multinomial Naive Bayes	81.2	83.8	82.08	82.1	82.3	82.33
Random Forest	80.86	85.54	81.14	81.1	80.39	80.74
Decision Tree	75.4	85.46	75.10	75.65	75.78	76.41
Gradient Boosting	83.02	84.11	82.4	82.58	82.45	82.9
XGBoost	84.52	84.82	84.78	84.8	84.11	84.52

Figure 2: Accuracy Table for Baseline Models

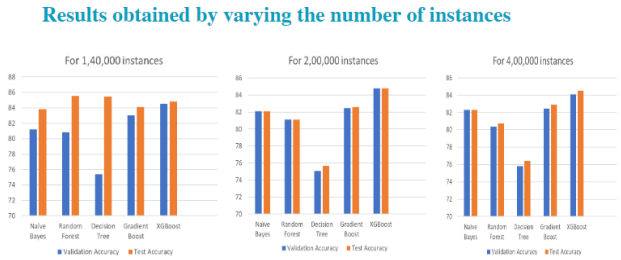

Figure 3: Accuracy chart for Baseline Models

terms of accuracy. Multinomial Naive Bayes, Random Forest and Gradient Boosting all produced results that were competitive with XGBoost. The above figure-2 summarizes the validation and test accuracy for all the five baseline models with different dataset sizes.
Fine-tuned the BERT model on a range of data instances as shown in Figure-4, varying from 2500 to 15000, with a fixed Tokenizer Size of Max length = 256. The hyperparameters chosen for fine-tuning included a learning rate of 2e-5, a batch size of 32, and training for a single epoch. Upon experimentation, we observed a proportional increase in performance with the expansion of the training data. The optimal results were achieved when fine-tuning on 15000 data instances, yielding an accuracy of 0.7815, an F1-Score of 0.777, Weighted Average Precision of 0.79, and Weighted Average Recall of 0.78. Additionally, we explored the impact of adjusting the learning rate, testing a higher value of lr=5e-5 as opposed to the original lr=2e-5. This adjustment resulted in a 75 percent accuracy when trained on 5000 data instances. However, it is noteworthy that this enhancement came at the expense of a threefold increase in training time compared to the lower learning rate.

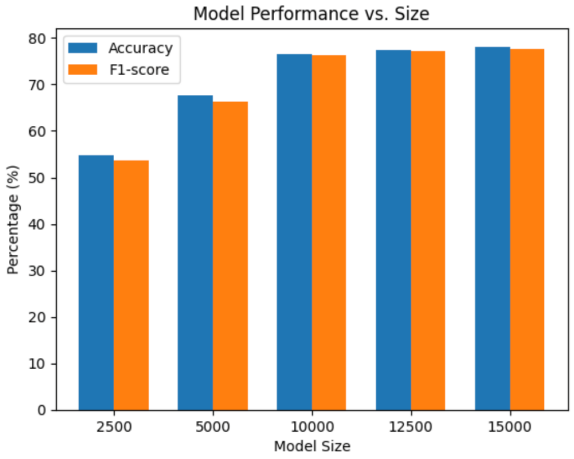

Figure 4: Accuracy chart for BERT Model

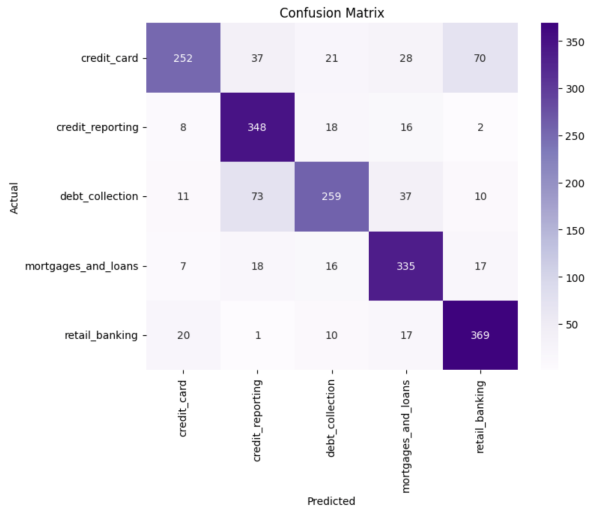

Figure 5: Confusion Matrix for BERT Model

6 Interesting insights

The most interesting bits of the project are

- The involvement of a pre-trained language model like BERT provided significant improvements in understanding the contextual information.
- Class consolidation technique simplified the problem and potentially improved the model's ability to generalize across classes.
- Implementing and evaluating various baseline models have provided valuable insights about the dataset characteristics.

There are various challenges that we encountered during the project.

- The validation time exhibits a gradual increase with a rise in the number of instances.
- There is class imbalance in the dataset, specified by uneven distribution across different complaint categories.
- Fine-tuning process in order to enhance the model accuracy.
- Mapping and consolidation of the complaints under "other financial services".

We would have tried a few more things to make the project better if there was given extra time.

- The BERT model's performance could be improved by performing hyperparameter tuning.
- Work on additional performance metrics to check how well our models can handle unseen data.

7 Ethical Considerations

In the context of our project, ethical considerations involve:

- **Being faithful to Industry Standards:** We have maintained the industry standards carefully and followed the protocols when collecting the data. There are limitations like class imbalances and challenges that are related to consolidation of classes.
- **Fair Model Training for Consumer Benefit:** We have made sure that our model is fairly trained enough to provide the accuracy. Consumers are benefited from the proposed model as they experience faster and better resolution of their issues.
- **Privacy protection:** Maintained privacy by not collecting any individuals personal data.

- **Transparency Maintenance:** Maintained transparency while communicating the project's purpose and working on the models.

Conclusion and Future scope

The project introduces a valuable Natural Language Processing technique that is intended to handle real-world problems. The project included training and testing of five baseline algorithms with XGBoost reaching an impressive 84 accuracy across many instances. To further improve the performance, advanced approaches such as Large Language Model(LLM) could be used for the categorization of customer complaints. The project also intends to provide sophisticated modeling tools that are particularly designed for issues, sub-issues and sub-products that are part of customer complaints. A systematic process is used to gather user feedback on model predictions, which yields insightful information for ongoing efficiency and improvement in performance. The ultimate goal is better assisting the Consumer Financial Protection Bureau in effectively addressing and resolving the complaints of consumers.

References

1. N. T. Thomas, "A LSTM based Tool for Consumer Complaint Classification," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 2018, pp. 2349-2351, doi: 10.1109/ICACCI.2018.8554857.
2. Pramod Kumar Naik; Prashanth T; Chandru S; Jaganath S; Sandesh Balan. "Consumer Complaints Classification Using Machine Learning Deep Learning". International Research Journal on Advanced Science Hub, 5, Issue 05S, 2023, 116-122. doi: 10.47392/irjash.2023.S015