

---

**THE GEORGE WASHINGTON UNIVERSITY**

---

WASHINGTON, DC

DATS 6312 : NATURAL LANGUAGE PROCESSING

INDIVIDUAL PROJECT REPORT

ON

**CUSTOMER COMPLAINT  
ANALYSIS AND  
PREDICTION SYSTEM**

Leveraging NLP To Enhance Customer Experience

BY:

MANOJ PADALA

# Overview of this project

## 1. **Sentiment Analysis of Complaints:**

- Analyzes the sentiment of customer complaints.
- Helps companies prioritize and escalate complaints faster based on sentiment.

## 2. **Complaint Summarization:**

- Summarizes complaints concisely, especially when they are lengthy or written in poor English.
- Assists customer service teams in understanding complaints more effectively.

## 3. **Resolution Prediction:**

- Predicts the likely resolution for complaints based on historical data.
- Provides customers with insights into possible outcomes, improving transparency and customer satisfaction.

## 4. **Complaint Classification:**

- Classifies complaints into specific categories.
- Enables companies to route issues to the appropriate teams, ensuring quicker resolution.

Data Used:

Consumer Complaint Database by CFPB

Rows: 2199541

Size of Data: 4.05GB

Dates: 01/01/2018 to 11/11/2024

Below are the topic which I have worked on

1. Summarization
2. Company Response to customer prediction.

## 1. Summarization:

**Preprocessing:** I preprocessed the data by removing white spaces, newlines, converting the text to lowercase, eliminating punctuation, and removing legal references and certain XXXX patterns.

Since the dataset was large and summarizing all the data would take a considerable amount of time, I filtered it down to 67,000 rows. The maximum length for summaries was set to 120 words, and the minimum length was set to 20 words.

### Summarization Models

#### 1. T5-Small:

- The T5-Small model leverages the "text-to-text" framework, where all NLP tasks are converted into a text generation problem. For summarization, the input is a long complaint narrative, and the model generates a concise summary.
- In this project, T5-Small was initially used to summarize 67,000 complaint narratives. While it achieved basic summarization, it struggled with longer and more complex inputs, resulting in a ROUGE score of 0.21. This limitation led to the exploration of a more advanced model.

**Time taken to summarise:** 9 hours

#### Evaluation Metrics:

- ROUGE Score: 0.21

Since the score was very bad, I switched to the facebook/bart-large-cnn model, which took approximately 13 hours to summarize the data.

#### 2. BART-Large-CNN:

- BART is a denoising autoencoder that excels at text generation and summarization. For summarization tasks, BART fine-tunes on large-scale datasets like CNN/Daily Mail, enabling it to produce coherent and concise outputs.
- By using BART-Large-CNN, the summarization improved significantly, achieving an average ROUGE score of 0.71. This model effectively handled lengthy and intricate complaint narratives, making it ideal for this use case.

### **Evaluation Metrics for BART model:**

- Average ROUGE Score: 0.71
  - ROUGE-1: 0.7361
  - ROUGE-2: 0.7020
  - ROUGE-L: 0.6813

**Time taken:** 13hours

I have saved the model locally.

## **2. Company Response to customer prediction.**

### **Model1: Finetuned Bert Base Uncased Model with Summary**

Since we limit the input to 512 tokens for the model, important tokens that appear at the end of a narrative may be missed, potentially leading to incorrect predictions. To address this issue, we summarized the narratives and trained the model on the summarized data.

**Number of records used for the Data: 67k**

### **Distribution of Dataset and its labels**

<b>Company response to consumer.</b>	<b>Count</b>
Closed with explanation	25500
Closed with non-monetary relief	25500
Closed with monetary relief	12703
Closed	1378
Untimely response	2111

For predicting the "Company Response to Consumer," I have used the columns "Company," "Issue," "Product," and "Summary" to fine-tune our model.

I have applied an 80-20 train-test split for fine-tuning and used class weights to handle the imbalanced dataset.

Model used : "bert-base-uncased"

Epochs: 4

Optimizer: AdamW

The "bert-base-uncased" model is a pretrained transformer model that uses a masked language modeling (MLM) objective. In MLM, 15% of the words in a sentence are randomly masked, and the model predicts the missing words, allowing it to learn a bidirectional representation of the sentence. This is particularly useful in your case as it can capture the full context of a narrative, even when important tokens are located at the end.

Additionally, the model employs next sentence prediction (NSP), where it learns whether two concatenated sentences are logically consecutive. This pretraining approach enhances the model's ability to understand context and relationships between different parts of the narrative, which is valuable when working with customer complaint resolutions.

### **Case-Specific Implementation:**

- The narratives were either summarized (using BART) or directly processed for classification.
- Input features like "Company," "Issue," "Product," and "Narrative/Summary" were utilized.
- The model's bidirectional nature allowed it to capture context effectively, even for lengthy inputs.

### **Evaluation metrics:**

Imbalance Handling: Addressed using class weights.

Validation F1-Score: 0.82

Validation Accuracy: 0.84

Validation Loss: 0.454

### **Model2: Finetuned Bert Base Uncase Model with Processed Narrative with large data:**

We aim to model the same model using a larger dataset with processed narratives and compare the results. I have filtered the dataset from 2 million records to 222,437 by reducing the number of records for the values "Closed with explanation" and "Closed with non-monetary relief" to 100,000 each in the "Company Response to Consumer" column.

### **Company Response to Consumer Breakdown:**

- Closed with explanation: 100,000
- Closed with non-monetary relief: 100,000
- Closed with monetary relief: 12,703
- Untimely response: 2,111
- Closed: 1,378

**Imbalance Handling:** Addressed using class weights.

**Optimizer Used:** AdamW

**Train-Test Split:** 80:20

The data was preprocessed by removing white spaces, newlines, converting the text to lowercase, eliminating punctuation, and removing legal references and certain XXXX patterns.

**Features Used:** "Company," "Issue," "Product," and "Complaint Narrative"

**Validation Metrics:**

- F1-Score: 0.81
- Accuracy: 0.77
- Loss: 0.67

The model has been saved locally.

Now, we can predict 4 outcomes:

**Outcome 1:**

- **Model 1:** Fine-tuned BERT Base Uncased Model with Summary
- **Inputs:** Company, Product, Issue, Summary (generated by our finetuned summarization model)

**Outcome 2:**

- **Model 1:** Fine-tuned BERT Base Uncased Model with Summary
- **Inputs:** Company, Product, Issue, Customer Complaint Narrative

**Model 2:**

- **Model 2:** Fine-tuned BERT Base Uncased Model with Processed Narrative and Large Data
- **Inputs:** Company, Product, Issue, Summary (generated by our finetuned summarization model)

**Model 2:**

- **Model 2:** Fine-tuned BERT Base Uncased Model with Processed Narrative and Large Data
- **Inputs:** Company, Product, Issue, Customer Complaint Narrative

Deployed an application using python library “streamlit”.

Results:

Since our model was not trained on the data before 2018

Let’s test our application with data before 2018

Test Case1:

**2765408**

Date CFPB received the complaint  
12/27/2017

Consumer's state  
NJ

Consumer's zip  
085XX

Submitted via  
Web

Did consumer dispute the response?  
N/A

Product

Credit reporting, credit repair services, or other personal  
consumer reports

Sub-product: Other personal consumer report

Issue

Problem with a company's investigation into an existing issue

Sub-issue: Investigation took more than 30 days

Consumer consent to publish narrative

☒ Consent provided

Consumer complaint narrative

Hi, Company Name : Paypal Dispute transaction date : XX/XX/XXXX in the amount of {\$500.00} ( Fraud transaction ) Dispute filed : XX/XX/XXXX Dispute ID : XXXX Dispute description : Unauthorised transaction Initial decision from Paypal received on XX/XX/XXXX was to deny my claim After that I appealed to Paypal fraud specialist on XX/XX/XXXX, and I got this message : " Hello XXXX XXXX, We recently received your appeal for an unauthorized case that was denied. We 'll get back to you with our findings within 9 days. " Its been more than 75 days and I have n't received any update or Credit of {\$500.00} During this time, I have contacted them twice and both times they have told me that the case is still under review and will be expedited. Nothing is happening. Paypal should immediately credit {\$500.00} to my account. Thanks XXXX XXXX

**Company information**

Date complaint sent to company  
12/27/2017

Company name  
Paypal Holdings, Inc

Timely response?

☒ Yes

Company response to consumer

Closed with monetary relief

Company public response

N/A

The above data is from CFPB on date 12/27/2017

Output from our app

# Customer Complaint Sentiment and Summary Analyzer

Enter your complaint text below to analyze its sentiment and generate a summary.

Enter the Consumer Complaint Narrative:

appealed to Paypal fraud specialist on XX/XX/XXXX, and I got this message : " Hello XXXX XXXX, We recently received your appeal for an unauthorized case that was denied. We 'll get back to you with our findings within 9 days. " Its been more than 75 days and I have n't received any update or Credit of {\$500.00} During this time, I have contacted them twice and both times they have told me that the case is still under review and will be expedited. Nothing is happening. Paypal should immediately credit {\$500.00} to my account. Thanks XXXX XXXX

Analyze

## Sentiment Analysis

Sentiment: Negative

## Summary

**Generated Summary:** Initial decision from Paypal received on XX/XX/XXXX was to deny my claim. After that I appealed to Paypal fraud specialist on XX/XX/XXXX, and I got this message : " Hello XXXX XXXX, We recently received your appeal for an unauthorized case that was denied. We 'll get back to you with our findings within 9 days " Its been more than 75 days and I have n't received any update or Credit of {\$500.00}

## Analysis 1

**Input Text:** Complaint Narrative: Hi, Company Name : Paypal Dispute transaction date : XX/XX/XXXX in the amount of {500.00} ( Fraud transaction ) Dispute filed : XX/XX/XXXX Dispute ID : XXXX Dispute description : Unauthorised transaction Initial decision from Paypal received on XX/XX/XXXX was to deny my claim After that I appealed to Paypal fraud specialist on XX/XX/XXXX, and I got this message : " Hello XXXX XXXX, We recently received your appeal for an unauthorized case that was denied. We 'll get back to you with our findings within 9 days. " Its been more than 75 days and I have n't received any update or Credit of {500.00} During this time, I have contacted them twice and both times they have told me that the case is still under review and will be expedited. Nothing is happening. Paypal should immediately credit {500.00} to my account. Thanks XXXX XXXX | Company: Paypal Holdings, Inc | Product: Credit reporting, credit repair services, or other personal consumer reports | Issue: Problem with a company's investigation into an existing issue

**Model 1 Prediction:** Closed with monetary relief (Probabilities: [0. 0.17 0.76 0.07 0. ])

**Model 2 Prediction:** Closed (Probabilities: [0. 0.2 0.01 0.8 0. ])

## Analysis 2

**Input Text:** Initial decision from Paypal received on XX/XX/XXXX was to deny my claim. After that I appealed to Paypal fraud specialist on XX/XX/XXXX, and I got this message : " Hello XXXX XXXX, We recently received your appeal for an unauthorized case that was denied. We 'll get back to you with our findings within 9 days " Its been more than 75 days and I have n't received any update or Credit of {\$500.00}

**Model 1 Prediction:** Closed with monetary relief (Probabilities: [0. 0.23 0.6 0.17 0. ])

**Model 2 Prediction:** Closed with non-monetary relief (Probabilities: [0. 0.54 0.04 0.42 0. ])

2 of our outcomes predicted the right outcome

Our model 1 which was fine-tuned with the summary has predicted the correct outcome



## Testcase2:

12/26/2017

Consumer's state

GA

Consumer's zip

31419

Submitted via

Web

Did consumer dispute the response?

N/A

Mortgage

Sub-product: Conventional home mortgage

Issue

Applying for a mortgage or refinancing an existing mortgage

Consumer consent to publish narrative

☒ Consent provided

Consumer complaint narrative

RE : LOAN DEPOT I received a flyer showing an attractive interest rate for a mortgage refinance so I called for further information. I answered the representative 's questions ( XXXX XXXX NMLS # XXXX ). And, I told him I would only be interested in a fixed rate 30-year conventional mortgage with no points or closing costs. My credit rating is more than 800 and debt to asset ratio low. He quoted me a rate very little less than what I have but more than the flyer advertised. Furthermore, there were fees exceeding (\$3000.00). I made it clear that I was NOT INTERESTED in his loan and told him to destroy my information and to stop any and all processes to apply for any loan and/or refinance. He said he would but he did not. Today I received a large stack of papers with my personal information ( much wrong ), and an application for a variable rate loan with even more fees that were disclosed in the phone call. I called Loan Depot today and told them to STOP THIS PROCESS, DESTROY MY INFORMATION, AND TO NOT SHARE ANY INFORMATION WITH ANYONE AT ANY TIME. The lady I spoke with said she would honor my requests. I feel betrayed, lied to, and used. My personal information has been used against my wishes and leaves me open for more abuse in the future. Loan Depot should be stopped from their deceptive practices and from abusing other people 's personal information. Thank you.

## Company information

Date complaint sent to company

12/26/2017

Company name

LD Holdings Group, LLC

Timely response?

☒ Yes

Company response to consumer

Closed with non-monetary relief

Company public response

Company believes it acted appropriately as authorized by contract or law

## Output

# Summary Analyzer

Enter your complaint text below to analyze its sentiment and generate a summary.

Enter the Consumer Complaint Narrative:

LOAN DEPOT I received a flyer showing an attractive interest rate for a mortgage refinance so I called for further information. I answered the representative 's questions ( XXXX XXXX NMLS # XXXX ). And, I told him I would only be interested in a fixed rate 30-year conventional mortgage with no points or closing costs. My credit rating is more than 800 and debt to asset ratio low. He quoted me a rate very little less than what I have but more than the flyer advertised. Furthermore, there were fees exceeding (\$3000.00). I made it clear that I was NOT INTERESTED in his loan and told him to

Analyze

## Sentiment Analysis ↔

Sentiment: Neutral

## Summary

**Generated Summary:** "I feel betrayed, lied to, and used. My personal information has been used against my wishes and leaves me open for more abuse in the future. Loan Depot should be stopped from their deceptive practices"

**Input Text:** Complaint Narrative: LOAN DEPOT I received a flyer showing an attractive interest rate for a mortgage refinance so I called for further information. I answered the representative 's questions ( XXXX XXXX NMLS # XXXX ). And, I told him I would only be interested in a fixed rate 30-year conventional mortgage with no points or closing costs. My credit rating is more than 800 and debt to asset ratio low. He quoted me a rate very little less than what I have but more than the flyer advertised. Furthermore, there were fees exceeding {\$3000.00}. I made it clear that I was NOT INTERESTED in his loan and told him to destroy my information and to stop any and all processes to apply for any loan and/or refinance. He said he would but he did not. Today I received a large stack of papers with my personal information ( much wrong ), and an application for a variable rate loan with even more fees that were disclosed in the phone call. I called Loan Depot today and told them to STOP THIS PROCESS, DESTROY MY INFORMATION, AND TO NOT SHARE ANY INFORMATION WITH ANYONE AT ANY TIME. The lady I spoke with said she would honor my requests. I feel betrayed, lied to, and used. My personal information has been used against my wishes and leaves me open for more abuse in the future. Loan Depot should be stopped from their deceptive practices and from abusing other people 's personal information. Thank you. | Company: LD Holdings Group, LLC | Product: Mortgage | Issue: Applying for a mortgage or refinancing an existing mortgage

**Model 1 Prediction:** Closed with non-monetary relief (Probabilities: [0. 0.87 0.11 0.02 0. ])

**Model 2 Prediction:** Closed with non-monetary relief (Probabilities: [0. 0.96 0.02 0.02 0. ])

## Analysis 2 ↩

**Input Text:** "I feel betrayed, lied to, and used. My personal information has been used against my wishes and leaves me open for more abuse in the future. Loan Depot should be stopped from their deceptive practices"

**Model 1 Prediction:** Closed with non-monetary relief (Probabilities: [0. 0.99 0. 0.01 0. ])

**Model 2 Prediction:** Closed with non-monetary relief (Probabilities: [0. 0.99 0. 0.01 0. ])

4 of our outcomes predicted the right outcome

## Conclusion:

From the attached images, we can observe that the model effectively summarizes complaints with lengthy narratives, enabling companies to gain a clearer understanding. Additionally, the classification model provides resolution predictions to assist in addressing these complaints.