Customer Complaint Analysis and Prediction System

Leveraging NLP to Enhance Customer Experience

Group - 7:

Bhumika Mallikarjun Horapet

Manoj Padala

Swathi Murali

Swathi K R

Data

Consumer Complaint Database
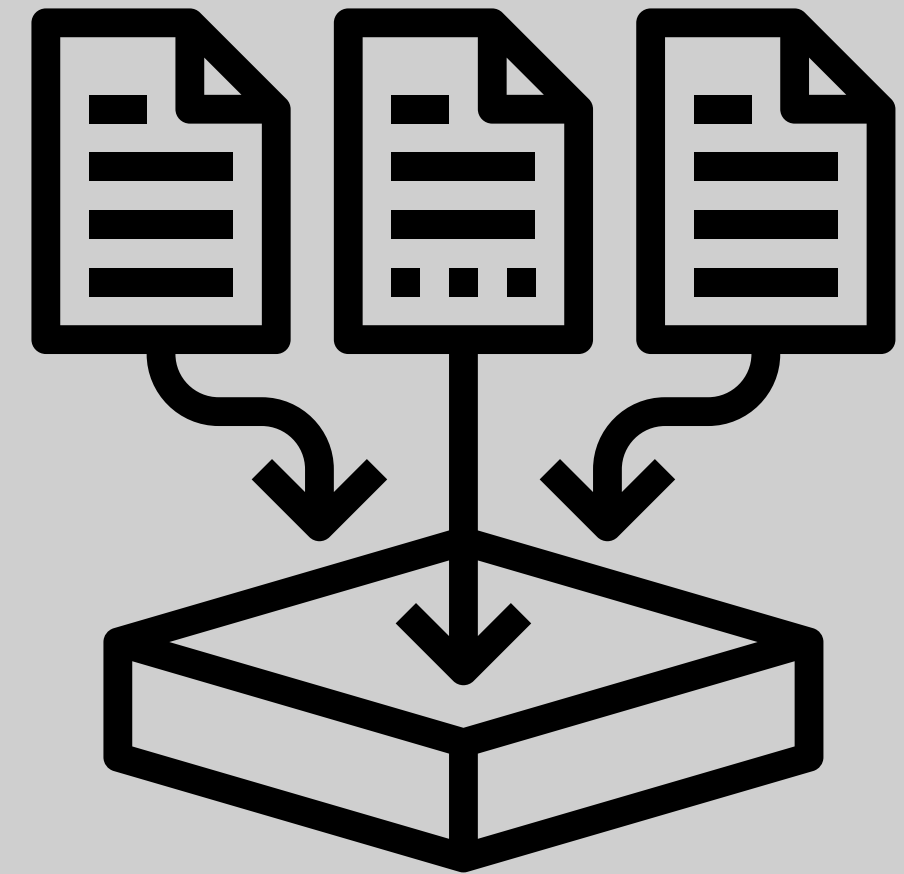
Data Link: https://www.consumerfinance.gov/data-research/consumer-complaints/

Total number of rows: 2199541

Size of the data: 4.05GB

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

GW

Problem Statement:

- Predict the likely outcome of customer complaints.
- Summarize lengthy or unclear narratives for better understanding by the service team.
- Classify issues into categories to streamline escalation to the appropriate team.
- Analyze sentiment to prioritize complaints based on urgency.

Key Benefits:

- Customer will get know predicted response from company
- Improve satisfaction through better complaint handling.
- Improve understanding of complaint from the customer (even if the language is unclear).

Pre-Processing

GW

1 Removed the stop words and Lemmatized

2 Removed punctuations, remove spaces and lowered the text

3 Removed the XXX patterns

4 Removed whitespaces , newlines.

Summarization of Complains

GW

Summarized to check if the classification of resolution gets improved with summary generated forn complaint resolution

Data Filtered to 70k records from 2M records.

Summarization of Complains

Models used:

- T5- small
  time taken to summarize: 9hrs
  Avg Rouge Score: 0.23

- facebook/bart-large-cnn
  time taken to summarize: 13hrs
  Avg Rouge Score: 0.71

COMPLAINT CLASSIFICATION

CHALLENGES

GW

UNSTRUCTURED

MULTIPLE FORMATS

ACCURATE CATEGORIZATION

SCALABILITY

SOLUTION APPROACH

MODEL TESTING

EVALUATION

MODEL SELECTION

FINE TUNING

MODELS TESTED

Model	Accuracy	Macro Avg F1-Score	Weighted Avg F1-Score	Top-3 Accuracy	Validation Loss
DistilBERT	58.25%	0.16	0.34	0.58	2.7234
RoBERTa	56.19%	0.17	0.32	0.56	2.8800
ALBERT	37.91%	0.06	0.16	0.37	3.5996
XLM-RoBERTa	30.00%	0.15	0.30	52.5%	3.0853
DistilBERT	67.71%	0.19	0.38	0.67	2.8119

FUTURE ENHANCEMENTS

- **Model Ensemble**
- **Advanced Class Imbalance Handling**

GW

Company Response Classification

GW

Models used:

- Bert Base Uncased
  Rows used: 67k

Columns used: Company, Issue, Category and Summary,
Test train split: 80:20
Epochs: 4
Handled Imbalance issue using Class weights
Validation F1- Score: 0.82
Validation Accuarcy: 0.84

Company Response Classification

Model used:

- Bert Base Uncased
  Rows used: 230k

Columns used: Company, Issue, Category and Complaint
Labels: Closed, Closed with explaination, Closed with monetary relief, Closed with non monetary relief, Untimely response
Test train split: 80:20
Epochs: 4
Handled Imbalance issue using Class weights
Validation F1- Score: 0.81
Validation Accuracy: 0.77

Company Response Classification

GW

Model used:

- Roberta-base
  Rows used: 67k

Columns used: Company, Issue, Category and Summary,
Test train split: 80:20
Epochs: 4
Handled Imbalance issue using Class weights
Validation F1- Score: 0.82
Validation Accuarcy: 0.80

Sentiment Analysis for Consumer Complaints

GW

Model Comparison:

1. Model Implementation
    - Data - 67k customer complaints data
    - Four transformer models were implemented:
        - BERT (bert-base-uncased)
        - RoBERTa (roberta-base)
        - DistilBERT (distilbert-base-uncased)
        - T5 (t5-base)
    - Models were fine-tuned on consumer complaint data.
    - Sentiment labels: Negative (0), Neutral (1), Positive (2)
2. Evaluation Metrics
    - Classification report generated for each model
    - Accuracy and F1 score used as primary comparison metric

Results and Best Model Selection

GW

Model	Accuracy	Precision	Recall	F1-Score
BERT (base-uncased)	0.9028	0.84	0.82	0.83
RoBERTa (base)	0.8920	0.88	0.80	0.83
DistilBERT (base-uncased)	0.8942	0.84	0.81	0.82
T5 (base)	0.8418	0.81	0.68	0.72

- BERT outperformed other models with 90.28% accuracy
- DistilBERT closely followed with 89.42% accuracy
- T5 showed the lowest performance at 84.18%
- Best Model: BERT (base-uncased)

Fine-tuned BERT Model for Sentiment Analysis

GW

- Model: BERT (bert-base-uncased)
- Data: Consumer complaints with 'Processed Narrative' column
- Task: 3-class sentiment classification (Negative, Neutral, Positive)
- Approach: 5-fold stratified cross-validation

Results and Performance

Class	Precision	Recall	F1-Score	Support
0 (Negative)	0.97	0.97	0.97	31,258
1 (Neutral)	0.94	0.92	0.93	3,342
2 (Positive)	0.97	0.97	0.97	33,048

- Accuracy: 0.97
- Macro Avg: 0.96
- Weighted Avg: 0.97

DEMO

Thank You!