Enhanced Customer Experience: Product Complaints Identification

A PROJECT REPORT

Submitted by

SRAVYA KAITHA

JAGANNATH PATRO ALLUPATI

NARSI REDDY SANIVARAPU

Under the guidance of

Professor Mr. Khaled Sayed

For the Course DSCI-6011-02

DEEP LEARNING

UNIVERSITY OF NEW HAVEN

WEST HAVEN, CONNECTICUT

SPRING 2024

TABLE OF CONTENTS

TITLE	PAGE NO
ABSTRACT	03
INTRODUCTION	04
PROBLEM STATEMENT	05
AIMS AND OBJECTIVES	05
STATEMENT OF VALUE	05
LITERATURE REVIEW	06
TECHNICAL IMPLEMENTATION	08
RESULTS AND ANALYSIS	12
CONCLUSION	14
REFERENCES	15

ABSTRACT

Enhancing customer experience through effective product complaint identification is critical for business success. The paper explores the use of natural language processing (NLP) approaches for successful complaint classification. Continuous improvement offers greater accuracy and consumer satisfaction.

Integrating NLP into complaint management systems allows for proactive responses to consumer problems, which fosters loyalty and confidence. This project implements two models (Bi-LSTM and DistilBERT from Hugging Face). From the analysis conducted, Bi-LSTM outperforms DistilBERT as it achieves an average F1 score of 0.87 compared to 0.72 of the DistilBERT.

Keywords— product complaint, natural language processing

INTRODUCTION

In today's highly competitive business environment, providing outstanding customer satisfaction is critical to the success and longevity of any enterprise. One critical component of doing this is the proper management and resolution of consumer concerns (Preotiuc-Pietro et al., 2019). However, many firms have found this to be a considerable issue because of the vast volume of consumer feedback that must be carefully reviewed. Natural Language Processing (NLP) is a critical tool for organizations to assess consumer feedback and ensure their pleasure. Organizations may use NLP approaches to properly identify client complaints, allowing them to respond quickly and effectively to each issue reported by their consumers.

Customer complaints provide insightful feedback that may help businesses find areas for development, alleviate pain spots, and improve the overall customer experience. Prompt and effective complaint resolution may result in enhanced customer loyalty, higher brand perception, and, ultimately, better business results. This project aims to use natural language processing (NLP) to categorize customer complaints properly (Wang et al., 2023). Using techniques such as text preparation, sentiment analysis, and machine learning algorithms, the system can automatically determine the nature of each complaint and put it into appropriate groupings. This enables firms to streamline complaint handling. Automatically classifying complaints helps businesses send them to the right teams or departments, resulting in a faster and more effective resolution process. This can also improve customer responsiveness since businesses and organizations may demonstrate their dedication to customer satisfaction and establish better client relationships (Wang et al., 2023). By automating the complaint classification process, organizations may better direct their resources, such as customer care professionals, to the most important concerns.

In general, the use of NLP to identify customer complaints marks a paradigm leap in customer relationship management. It provides organizations with a strong tool for delivering exceptional levels of service excellence and establishing long-term client loyalty. Businesses can leverage the full potential of NLP to improve the customer experience and prosper in today's competitive economy by integrating strategically and innovating relentlessly.

PROBLEM STATEMENT

Customer feedback is instrumental in determining the trajectory of businesses across diverse sectors. Continuously, enterprises endeavour to grasp their clientele's desires and inclinations, and among the tools that have risen to prominence is Natural Language Processing (NLP). Through harnessing NLP's capabilities, businesses can extract valuable insights from customer feedback, empowering them to base decisions on data and elevate the overall customer journey.

AIMS AND OBJECTIVES

This project aims to develop an NLP model capable of accurately discerning product complaints from diverse customer feedback sources. The objectives include:

- To enhance customer experience through the effective identification and resolution of complaints.
- To streamline the process of complaint analysis, enabling businesses to respond promptly and appropriately to customer concerns.
- To improve overall customer satisfaction and loyalty by facilitating proactive engagement and swift resolution of product-related issues.

STATEMENT OF VALUE

This project is worthwhile because it addresses the vital need for businesses to handle and respond to consumer complaints efficiently. By automating the process of identifying complaints using NLP, businesses may streamline their customer service operations, reduce response times, and ultimately increase overall customer satisfaction and retention.

LITERATURE REVIEW

Mukherjee (2022) discusses the utilization of Natural Language Processing (NLP) techniques to analyze consumer complaints in the financial sector. Financial products, crucial for achieving various financial goals, sometimes fail to meet expectations, leading to customer crises. The Consumer Complaint Database contains grievances from customers nationwide. The author proposes a TF-IDF-based Information Retrieval model to expedite and streamline complaint classification. This model aims to empower computers to comprehend document contents, including contextual nuances, to extract insights and categorize complaints efficiently. Using NLP, the model seeks to enhance the speed and accuracy of complaint analysis, ultimately improving customer satisfaction and regulatory compliance in the financial sector.

Singh et al. (2020) explore the application of Neural Machine Translation (NMT) in identifying customer complaints from product reviews, focusing on low-resource scenarios where labeled training data is scarce. Traditionally, supervised learning approaches have been employed for complaint recognition on platforms like Twitter, but they demand extensive labeled data, which can be cost-prohibitive. To overcome this challenge, the study suggests utilizing NMT models to translate reviews from low-resource languages like Hindi into English, a language for which labeled data is more readily available. The performance of downstream complaint identification tasks is then evaluated using these translated English reviews. By leveraging NMT, the research aims to facilitate complaint recognition in diverse linguistic contexts, enabling organizations, multinationals, and online retailers to better understand and address customer grievances without relying on expensive labeled datasets.

Singh et al. (2021) look into the significance of addressing customer complaints swiftly and efficiently in today's competitive business landscape. It emphasizes the role of complaints in discerning customer needs, which in turn guides companies in their research and development endeavors. However, automating the identification of complaints from vast online content presents challenges. To tackle this, the authors propose a multitask learning approach that simultaneously handles complaint identification and sentiment classification. They leverage weak supervision to annotate the corpus with sentiment labels and incorporate commonsense knowledge features through AffectiveSpace. Experimental results showcase the efficacy of their framework, achieving high accuracy in both complaint identification and sentiment classification tasks. The study underscores the interdependence between sentiment analysis and complaint classification, highlighting the mutual benefits of learning these tasks concurrently.

Bozyiğit et al. (2022) focus on categorizing customer complaints related to packaged food products in the Turkish language using machine learning (ML) approaches. While previous research has largely concentrated on English, this paper aims to address this gap by developing accurate classification models for Turkish complaints. Various ML algorithms employing TF-IDF and word2vec feature representation techniques were implemented to classify complaints. Experimental results highlight XGBoost with TF-IDF weighting as the most effective method, achieving an 86% F-measure score. Interestingly, word2vec-based classifiers exhibit lower performance compared to TF-IDF. Furthermore, feature selection using the Chi-Square (CH2) method enhances prediction performance across all TF-IDF-based ML algorithms, with XGBoost achieving an 88% F-measure score. Overall, the study contributes to enhancing customer satisfaction in the food industry by providing insights into complaint categorization using ML techniques tailored to the Turkish language.

Tahsin et al. (2023) explore the integration of Natural Language Processing (NLP) and Federated Learning (FL) for analyzing consumer complaints, focusing on laptop reviews. Traditional methods of analyzing customer feedback are often time-consuming and require manual effort. To overcome these challenges, the study applies NLP techniques to categorize and analyze user reviews from a laptop feedback dataset. Federated learning is employed to ensure client privacy and handle uneven data distribution. The FedAvg algorithm integrates local model learning into a central model after each federated training round. Three BERT variants (BERT, DistilBERT, and RoBERTa) are experimented with for classification tasks, with RoBERTa yielding the best results (71.59% accuracy). The paper proposes a comprehensive system for generating product summaries and customer discontent reports based on graphical representations and NLP-assisted classification. The best-performing model variant is deployed in a web application, demonstrating the successful execution of the proposed system. Overall, the study offers insights into leveraging NLP and FL for effective consumer complaint analysis in the context of laptop feedback.

TECHNICAL IMPLEMENTATION

This section covers the main component of the project, comprehensively describing all the steps needed to develop and train a model to help with customer complaints.

Dataset Description

The dataset used in this project contains 162,421 customer complaints. It was acquired from https://github.com/halpert3/complaint-content-classification-nlp/blob/main/project_data/complaints_processed.csv.

Exploratory Data Analysis (EDA)

This is a very important process of any data analysis as it ensures that a proper understanding of the data is in place before the actual analysis. The first step is to load the data and visualize its preview to confirm that the loading process works perfectly. This preview of the first few records is shown in the figure below.

	Unnamed: 0	product	narrative
0	0	credit_card	purchase order day shipping amount receive pro...
1	1	credit_card	forwarded message date tue subject please inve...
2	2	retail_banking	forwarded message cc sent friday pdt subject f...
3	3	credit_reporting	payment history missing credit report speciali...
4	4	credit_reporting	payment history missing credit report made mis...
5	5	credit_reporting	payment history missing credit report made mis...
6	6	credit_reporting	va date complaint experian credit bureau invol...
7	7	credit_reporting	account reported abbreviated name full name se...
8	8	credit_reporting	account reported abbreviated name full name se...
9	9	credit_reporting	usdoexxxx account reported abbreviated name fu...
10	10	mortgages_and_loans	beginning mortgage held mb financial mb mortga...
11	11	credit_card	called request new york state covid relief pla...
12	12	credit_card	capital one secured credit account opened name...
13	13	credit_reporting	collection account acct opened balance account...
14	14	debt_collection	collection agency system inc mn phone sent let...

Figure 1:Data preview.

The dataframe has a column called "Unnamed: 0," which is not necessary for the analysis; hence, we dropped it, and the narrative column is also renamed to tweet. The resulting dataframe is shown in the figure below.

	product	tweet
0	credit_card	purchase order day shipping amount receive pro...
1	credit_card	forwarded message date tue subject please inve...
2	retail_banking	forwarded message cc sent friday pdt subject f...
3	credit_reporting	payment history missing credit report speciali...
4	credit_reporting	payment history missing credit report made mis...

We then need to understand some basic information about the dataset, such as its size, the size of each product category, and the number of missing values.

```
* Size of dataframe: (162421, 2)

* Datatype of columns are:
 product    object
tweet      object
dtype: object

* Count of different product categories:
 credit_reporting       91179
debt_collection        23150
mortgages_and_loans    18990
credit_card            15566
retail_banking         13536
Name: product, dtype: int64

* Number of NaNs among tweets are: 10
```

Figure 2: Basic information of the data.

Including missing values in a data analytics project always leads to wrong analysis. We, therefore, drop them and only proceed with records that have all the values.

We can also visualize product categories in a chart to help us understand the numbers behind each category. The figure below shows this visualization.

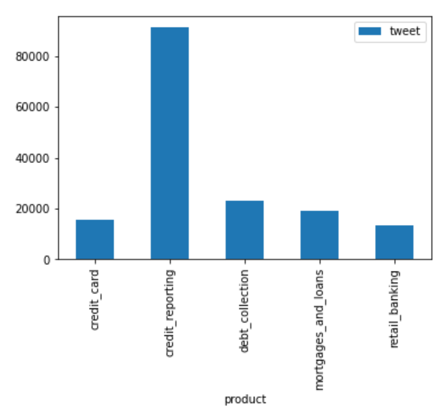

Figure 3: Product categories distribution.

Splitting the dataset

The dataset is then split into train and test sets that will be used during model training and evaluation.

Balancing the dataset

From figure 3 above, it is evident that the dataset is imbalanced, with the "credit_reporting" product category having a significantly higher number of records than the other categories, and this would lead to inaccuracies in the model if the data is used this way. We therefore balance the train dataset using a method called random oversampling.

Text processing

Working with text data, such as our dataset, requires thorough processing of the texts to make sure the models achieve higher accuracy. We are working directly with raw texts, which is a limitation of most models. One way of text processing is using different NLP techniques from NLTK, such as stopwords removal, lemmatization, and other normal text processing techniques, such as removing punctuations and other unwanted characters. The dataset was subjected to some of these techniques to make sure we dealt with clean data.

Model Building

Bi-LSTM

This model is trained on the training data and evaluated on the test data. Its architecture is shown in the figure below.

```python
# Model Training
model = Sequential()
model.add(Embedding(max_words,
                    embedding_dim,
                    input_length=max_sequence_length))

# Bidirectional LSTM
model.add(Bidirectional(LSTM(16, return_sequences=True, dropout=0.4, recurrent_dropout=0)))

model.add(GlobalMaxPool1D())

model.add(Dense(5,activation='softmax'))

model.summary()
```

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, 170, 32)           80000
_________________________________________________________________
bidirectional (Bidirectional (None, 170, 32)           6272
_________________________________________________________________
global_max_pooling1d (Global (None, 32)                0
_________________________________________________________________
dense (Dense)                (None, 5)                 165
=================================================================
Total params: 86,437
Trainable params: 86,437
Non-trainable params: 0
_________________________________________________________________
```

Figure 4: Bi-LSTM model architecture.

The data is then passed to the network, and the model is trained using ten epochs. The training progress is shown in the figure below.

```
Epoch 1/20
3291/3291 [==============================] - 56s 16ms/step - loss: 1.1929 - accuracy: 0.6375 - val_loss: 0.1532 - val_accuracy: 0.8557
Epoch 2/20
3291/3291 [==============================] - 51s 15ms/step - loss: 0.5989 - accuracy: 0.8487 - val_loss: 0.1430 - val_accuracy: 0.8645
Epoch 3/20
3291/3291 [==============================] - 51s 15ms/step - loss: 0.5591 - accuracy: 0.8583 - val_loss: 0.1401 - val_accuracy: 0.8681
Epoch 4/20
3291/3291 [==============================] - 51s 15ms/step - loss: 0.5508 - accuracy: 0.8592 - val_loss: 0.1391 - val_accuracy: 0.8692
Epoch 5/20
3291/3291 [==============================] - 51s 15ms/step - loss: 0.5346 - accuracy: 0.8635 - val_loss: 0.1406 - val_accuracy: 0.8668
Epoch 6/20
3291/3291 [==============================] - 51s 15ms/step - loss: 0.5225 - accuracy: 0.8678 - val_loss: 0.1356 - val_accuracy: 0.8720
Epoch 7/20
3291/3291 [==============================] - 51s 15ms/step - loss: 0.5214 - accuracy: 0.8666 - val_loss: 0.1364 - val_accuracy: 0.8714
Epoch 8/20
3291/3291 [==============================] - 50s 15ms/step - loss: 0.5166 - accuracy: 0.8690 - val_loss: 0.1352 - val_accuracy: 0.8727
Epoch 9/20
3291/3291 [==============================] - 50s 15ms/step - loss: 0.5165 - accuracy: 0.8682 - val_loss: 0.1398 - val_accuracy: 0.8648
Epoch 10/20
3291/3291 [==============================] - 50s 15ms/step - loss: 0.5084 - accuracy: 0.8692 - val_loss: 0.1344 - val_accuracy: 0.8732
Epoch 11/20
3291/3291 [==============================] - 50s 15ms/step - loss: 0.5023 - accuracy: 0.8719 - val_loss: 0.1335 - val_accuracy: 0.8727
Epoch 12/20
3291/3291 [==============================] - 50s 15ms/step - loss: 0.5007 - accuracy: 0.8726 - val_loss: 0.1335 - val_accuracy: 0.8742
Epoch 13/20
3291/3291 [==============================] - 50s 15ms/step - loss: 0.5032 - accuracy: 0.8714 - val_loss: 0.1331 - val_accuracy: 0.8741
Epoch 14/20
3291/3291 [==============================] - 51s 15ms/step - loss: 0.4963 - accuracy: 0.8738 - val_loss: 0.1315 - val_accuracy: 0.8754
Epoch 15/20
3291/3291 [==============================] - 51s 15ms/step - loss: 0.4951 - accuracy: 0.8737 - val_loss: 0.1339 - val_accuracy: 0.8751
Epoch 16/20
3291/3291 [==============================] - 51s 16ms/step - loss: 0.4972 - accuracy: 0.8730 - val_loss: 0.1365 - val_accuracy: 0.8717
Epoch 17/20
3291/3291 [==============================] - 51s 16ms/step - loss: 0.4914 - accuracy: 0.8747 - val_loss: 0.1349 - val_accuracy: 0.8747
Epoch 18/20
3291/3291 [==============================] - 51s 15ms/step - loss: 0.4895 - accuracy: 0.8756 - val_loss: 0.1374 - val_accuracy: 0.8705
Epoch 19/20
3291/3291 [==============================] - 51s 15ms/step - loss: 0.4802 - accuracy: 0.8784 - val_loss: 0.1349 - val_accuracy: 0.8734
Epoch 20/20
3291/3291 [==============================] - 51s 15ms/step - loss: 0.4881 - accuracy: 0.8754 - val_loss: 0.1349 - val_accuracy: 0.8737
```

Figure 5: Bi-LSTM model training process.

<u>DistilBERT Model (from Hugging Face)</u>

The second model trained on the dataset is DistilBERT from Hugging Face. The encoding process of the model is shown in the figure below.

```
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
Batches:   0%|          | 0/1 [00:00<?, ?it/s]
```

Figure 6: DistilBERT encoding process.

A logistic regression model is then applied to the output of the DistilBERT and trained on the train data.

RESULTS AND ANALYSIS

This section comprehensively analyzes the results obtained by the models.

Bi-LSTM

The figure below shows the classification report for the model.

	precision	recall	f1-score	support
0	0.78	0.78	0.78	3159
1	0.93	0.92	0.92	18217
2	0.82	0.76	0.79	4648
3	0.85	0.83	0.84	3802
4	0.81	0.90	0.85	2657
micro avg	0.88	0.87	0.87	32483
macro avg	0.84	0.84	0.84	32483
weighted avg	0.88	0.87	0.87	32483
samples avg	0.87	0.87	0.87	32483

Figure 7: Bi-LSTM classification report.

The Bi-LSTM model excels in identifying customer complaints, with an average F1 Score of 0.87 across classes. It shows a particularly strong ability to classify common complaints (Class 1) while maintaining consistent performance across less frequent categories.

DistilBERT Model

The figure below shows the classification report for the model.

	precision	recall	f1-score	support
0	0.63	0.64	0.63	439
1	0.65	0.66	0.66	269
2	0.87	0.89	0.88	2236
3	0.73	0.73	0.73	422
4	0.72	0.65	0.68	634
accuracy			0.79	4000
macro avg	0.72	0.71	0.72	4000
weighted avg	0.79	0.79	0.79	4000

Figure 8: DistilBERT with logistic regression classification report.

The model is generally performing well, particularly in class 2, which has the most instances. However, there is a noticeable variation in performance across different classes, with classes 0 and 1 having lower precision and recall. The relatively balanced F1 scores suggest that the model does not heavily favor precision over recall or vice versa for any class. The average f1-score for the model is 0.72.

Model Comparison

The Bi-LSTM model has a much higher average F1-score than the DistilBERT model. This suggests that the Bi-LSTM is more successful overall in classifying customer complaints into various categories. The Bi-LSTM's strength is its ability to analyze sequential data while also capturing the context of language, which is critical for effectively comprehending and categorizing customer complaints based on textual content. Given these factors, the Bi-LSTM model is the better choice for this task, as it demonstrates greater capabilities in terms of both high accuracy and performance consistency across a variety of complaint kinds.

CONCLUSION:

Using Natural Language Processing (NLP) to identify product complaints improves the customer experience dramatically. NLP approaches simplify the processing of massive volumes of consumer input, allowing firms to handle issues quickly and accurately. Businesses may use sentiment analysis, topic modeling, and machine learning classifiers to identify complaints and prioritize resolution efforts properly. The continuous modification of NLP models guarantees that complaint identification and customer satisfaction continue to increase. This project leverages the use of two models (Bi-LSTM and DistilBERT from Hugging Face). From the analysis conducted, Bi-LSTM outperforms DistilBERT as it achieves a higher average f1 score of 0.87 compared to 0.72 of DistilBERT.

REFERENCES:

- Bozyiğit, F., Doğan, O., & Kilinç, D. (2022). Categorization of Customer Complaints in Food Industry Using Machine Learning Approaches. *Journal of Intelligent Systems: Theory and Applications*, *5*(1), 85–91. https://doi.org/10.38016/jista.954098
- Mukherjee, S. (2022). Identification of Nature of Complaint from Consumer Complaint Database using NLP. *International Journal of Science and Research*. https://doi.org/10.21275/SR23302130625
- Preotiuc-Pietro, D., Gaman, M., & Aletras, N. (2019, June 10). *Automatically Identifying Complaints in Social Media*. ArXiv.org. https://doi.org/10.48550/arXiv.1906.03890
- Singh, A., Saha, S., Hasanuzzaman, Md., & Dey, K. (2021). Multitask Learning for Complaint Identification and Sentiment Analysis. *Cognitive Computation*, *14*. https://doi.org/10.1007/s12559-021-09844-7
- Singh, R., Haque, R., Hasanuzzaman, M., & Way, A. (2020). *Identifying Complaints from Product Reviews in Low-resource Scenarios via Neural Machine Translation* (pp. 108–116). https://aclanthology.org/2020.icon-main.14.pdf
- Tahsin, M. U., Hossain, S., & Rahman, R. M. (2023). Combining Natural Language Processing and Federated Learning for Consumer Complaint Analysis: A Case Study on Laptops. *SN Computer Science*, *4*(5). https://doi.org/10.1007/s42979-023-01989-6
- Wang, J., Lai, J.-Y., & Lin, Y.-H. (2023). Social media analytics for mining customer complaints to explore product opportunities. *Computers & Industrial Engineering*, *178*, 109104. https://doi.org/10.1016/j.cie.2023.109104.