Sentiment Analysis of Amazon Customer Reviews using Transformers

A PROJECT REPORT

Submitted by

Keerthi Balabhadruni

Ajay Kumar Reddy Punuru

UNIVERSITY OF NEW HAVEN

WEST HAVEN, CONNECTICUT

SPRING 2024

TABLE OF CONTENTS

ABSTRACT

In this project, we undertake sentiment analysis of Amazon customer reviews using transformer-based models. The primary objective is to classify reviews into positive, negative, or neutral sentiments, providing businesses with insights to enhance products and customer satisfaction. We begin by collecting and preprocessing the Amazon Fine Food Reviews dataset from Kaggle. Leveraging tools like NLTK and the Transformers library, we analyze the sentiment using VADER and the Roberta Pretrained Model. By combining results from these models, we conduct a comprehensive evaluation based on accuracy, precision, recall, and F1 score metrics. Visualizations including bar charts and confusion matrices are employed to represent model performance. Our analysis aids businesses in understanding customer preferences, improving product quality, and making data-driven decisions. Through this project, we aim to empower businesses with actionable insights to enhance their competitive advantage in the market.

Introduction:

In today's highly competitive e-commerce landscape, understanding customer sentiments is paramount for businesses striving to stay ahead. Sentiment analysis, a cornerstone of natural language processing, offers a powerful tool to decipher customer opinions and attitudes towards products and services. In this project, we focus on analyzing the sentiment of Amazon customer reviews using advanced transformer-based models. By delving into the vast pool of reviews, we aim to extract valuable insights that can drive strategic decision-making and foster improvements in product offerings and customer experiences.

The exponential growth of online reviews on platforms like Amazon underscores the need for sophisticated sentiment analysis techniques. As Amazon continues to serve as a leading marketplace for a diverse array of products, mining insights from customer feedback becomes increasingly indispensable. Our project endeavors to harness the potential of transformer-based models, such as the Roberta Pretrained Model, to effectively categorize reviews into positive, negative, or neutral sentiments. By leveraging state-of-the-art natural language processing tools, we strive to provide businesses with actionable intelligence that can inform product development, marketing strategies, and customer engagement initiatives.

Through a meticulous approach encompassing data collection, preprocessing, and sentiment analysis, we seek to uncover nuanced patterns within Amazon customer reviews. By juxtaposing traditional sentiment analysis techniques like VADER with cutting-edge transformer models, we aim to showcase the efficacy and robustness of advanced NLP methodologies. Ultimately, our project aspires to equip businesses with the insights needed to not only comprehend customer sentiments but also to proactively address concerns, refine offerings, and foster enduring customer relationships in the ever-evolving landscape of e-commerce.

Datasets:

Dataset Overview:

The dataset utilized in this project is sourced from Kaggle and comprises Amazon Fine Food Reviews. It encompasses a comprehensive collection of customer reviews spanning various food products available on the Amazon platform. This dataset offers a rich repository of textual data, including reviews, ratings, timestamps, and helpfulness votes, providing a holistic view of customer sentiments and preferences.

The Amazon Fine Food Reviews dataset consists of structured data fields such as ID, Product ID, User ID, Profile Name, Helpfulness Numerator, Helpfulness Denominator, Score, Time, Summary, and Text. These attributes offer valuable insights into the characteristics of each review, enabling detailed analysis and sentiment classification.

Dataset Usage:

Throughout the project, the Amazon Fine Food Reviews dataset serves as the primary source of data for sentiment analysis tasks. Leveraging natural language processing techniques, we pre-process and analyze the textual content of customer reviews to classify sentiments as positive, negative, or neutral. By harnessing the power of transformer-based models and sentiment analysis libraries such as VADER, we aim to extract actionable insights from the dataset to aid businesses in understanding customer feedback and enhancing product offerings.

Data Pre-processing:

Data pre-processing is a crucial step in preparing the Amazon Fine Food Reviews dataset for sentiment analysis. This process involves several key steps to clean, normalize, and transform the raw data into a format suitable for analysis:

1. **Removing Irrelevant Columns:** Initially, we identify and remove any columns from the dataset that are not relevant to our sentiment analysis task. This streamlines the dataset and focuses our analysis on the essential attributes.

2. **Handling Missing Values:** Next, we address any missing or null values in the dataset. Depending on the extent of missing data, we may choose to impute missing values using techniques such as mean imputation or drop rows/columns with missing data.
3. **Text Cleaning:** The textual content of customer reviews often contains noise, such as HTML tags, punctuation, and special characters. We employ text cleaning techniques, including removing HTML tags, converting text to lowercase, and removing punctuation and special characters, to ensure consistency and readability.
4. **Tokenization:** We tokenize the cleaned text, splitting it into individual words or tokens. This step facilitates further analysis by breaking down the text into its constituent parts.
5. **Stopword Removal:** Stopwords, common words that do not carry significant meaning, are removed from the tokenized text to focus on content words that convey sentiment.
6. **Lemmatization or Stemming:** To further normalize the text, we apply lemmatization or stemming techniques to reduce words to their base or root form. This helps in reducing the dimensionality of the data and improving the accuracy of sentiment analysis.
7. **Encoding Labels:** Finally, we encode the sentiment labels (positive, negative, neutral) into numerical values for model training and evaluation purposes. This transformation ensures compatibility with machine learning algorithms.

Models:

In this project, we leverage state-of-the-art transformer-based models for sentiment analysis of Amazon customer reviews. These models represent the cutting edge of natural language processing (NLP) technology and offer powerful capabilities for understanding and classifying textual data.

1. **Roberta Pretrained Model:** The Roberta Pretrained Model is a variant of the Transformer architecture, specifically designed for NLP tasks. Trained on a large corpus of text data, including social media posts, news articles, and web pages, Roberta excels at capturing contextual information and semantic relationships within text. We utilize the Roberta model to perform sentiment analysis on Amazon customer reviews, leveraging its sophisticated understanding of language semantics to classify reviews into positive, negative, or neutral sentiments.

2. **VADER (Valence Aware Dictionary and sEntiment Reasoner):** In addition to transformer-based models, we employ VADER, a lexicon and rule-based sentiment analysis tool specifically designed for social media text. VADER utilizes a pre-defined lexicon of words with associated sentiment scores to analyze the sentiment of textual data. By leveraging VADER's extensive lexicon and rule-based approach, we complement the analysis performed by the transformer models, providing a comprehensive evaluation of sentiment across Amazon customer reviews.

Evaluation Metrics:

In assessing the performance of our sentiment analysis models on Amazon customer reviews, we utilize a range of evaluation metrics to quantify their effectiveness in classifying sentiments accurately. These metrics provide valuable insights into the models' strengths and weaknesses, guiding further refinement and optimization efforts. The key evaluation metrics employed in our analysis include:

1. **Accuracy:** Accuracy measures the proportion of correctly classified reviews out of the total number of reviews. It provides a general overview of the model's performance but may not be sufficient when dealing with imbalanced datasets.
2. **Precision:** Precision measures the proportion of correctly classified positive or negative reviews out of all reviews classified as positive or negative by the model. It focuses on the correctness of positive or negative predictions, minimizing false positives.
3. **Recall:** Recall measures the proportion of correctly classified positive or negative reviews out of all actual positive or negative reviews in the dataset. It focuses on capturing as many positive or negative reviews as possible, minimizing false negatives.
4. **F1 Score:** The F1 score is the harmonic mean of precision and recall, providing a balanced measure of a model's performance. It takes into account both false positives and false negatives, making it a robust metric for evaluating classification performance.
5. **ROC-AUC Curve:** The Receiver Operating Characteristic (ROC) curve plots the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings. The Area Under the ROC Curve (AUC) quantifies the overall performance of the model in distinguishing between positive and negative sentiments, with higher values indicating better performance.

Proposed Methodology:

Our methodology for sentiment analysis of Amazon customer reviews encompasses a systematic approach aimed at extracting meaningful insights from textual data. The key steps involved in our proposed methodology are as follows:

1. **Data Collection and Preprocessing:** We begin by collecting the Amazon Fine Food Reviews dataset from Kaggle and preprocessing it to ensure data quality and consistency. This involves removing irrelevant columns, handling missing values, and cleaning the textual content of reviews by removing HTML tags, punctuation, and stopwords.
2. **Model Selection and Training:** We select and train transformer-based models such as Roberta for sentiment analysis tasks. These models are pre-trained on vast amounts of text data and fine-tuned on specific NLP tasks, making them well-suited for classifying sentiments in customer reviews. Additionally, we utilize rule-based approaches like VADER for comparative analysis.
3. **Sentiment Analysis:** With trained models in place, we perform sentiment analysis on the preprocessed Amazon customer reviews. The models classify each review as positive, negative, or neutral based on the textual content, leveraging their understanding of language semantics and sentiment cues.
4. **Evaluation and Model Comparison:** We evaluate the performance of the sentiment analysis models using a range of evaluation metrics, including accuracy, precision, recall, F1 score, and ROC-AUC curve. This allows us to compare the effectiveness of different models and identify the most suitable approach for sentiment classification.
5. **Visualization and Interpretation:** To facilitate interpretation and decision-making, we visualize the results of sentiment analysis using bar charts, confusion matrices, and ROC curves. These visualizations provide insights into the distribution of sentiments across reviews and the models' classification performance.
6. **Final Report and Analysis:** Finally, we compile the findings of our sentiment analysis into a comprehensive evaluation report. This report includes a detailed analysis of model performance, comparison of transformer-based models and rule-based approaches, and recommendations for businesses based on the insights gleaned from customer reviews.

Results:

After applying our proposed methodology for sentiment analysis of Amazon customer reviews, we obtained insightful results that shed light on customer sentiments towards various products. Here are the key findings from our analysis:

1. **Model Performance:** The transformer-based models, particularly Roberta, demonstrated high accuracy and robust performance in classifying sentiments across Amazon customer reviews. These models effectively captured the nuances of language semantics and context, leading to accurate sentiment classification.
2. **Comparison with Rule-based Approaches:** When compared with rule-based approaches like VADER, transformer-based models consistently outperformed in terms of accuracy, precision, recall, and F1 score. This highlights the superiority of deep learning-based approaches in sentiment analysis tasks.
3. **Sentiment Distribution:** We observed a diverse distribution of sentiments across Amazon customer reviews, with a significant proportion classified as positive, followed by neutral and negative sentiments. This distribution varied across product categories, reflecting the diversity of customer experiences and preferences.
4. **Impact on Business Decisions:** The insights gleaned from sentiment analysis have significant implications for business decision-making. By understanding customer sentiments, businesses can identify areas for product improvement, enhance marketing strategies, and foster stronger customer relationships.
5. **Visualization of Results:** Visualizations such as bar charts, confusion matrices, and ROC curves provided a clear overview of model performance and sentiment distribution. These visual aids facilitated interpretation and communication of the analysis results to stakeholders.

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Run Markdown

```
In [25]: !python -m nltk.downloader vader_lexicon

/usr/lib/python3.10/runpy.py:126: RuntimeWarning: 'nltk.downloader' fou
nd in sys.modules after import of package 'nltk', but prior to executio
n of 'nltk.downloader'; this may result in unpredictable behaviour
  warn(RuntimeWarning(msg))
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...

In [26]: from nltk.sentiment import SentimentIntensityAnalyzer
         from tqdm.notebook import tqdm

         sia = SentimentIntensityAnalyzer()

In [27]: sia.polarity_scores('I am so happy!')
Out[27]: {'neg': 0.0, 'neu': 0.318, 'pos': 0.682, 'compound': 0.6468}

In [28]: sia.polarity_scores('This is the worst thing ever.')
Out[28]: {'neg': 0.451, 'neu': 0.549, 'pos': 0.0, 'compound': -0.6249}

In [29]: sia.polarity_scores(example)
Out[29]: {'neg': 0.22, 'neu': 0.78, 'pos': 0.0, 'compound': -0.5448}

In [30]: # Run the polarity score on the entire dataset
         res = {}
         for i, row in tqdm(df.iterrows(), total=len(df)):
             text = row['Text']
             myid = row['Id']
             res[myid] = sia.polarity_scores(text)
```

```
In [48]: from transformers import pipeline

         sent_pipeline = pipeline("sentiment-analysis")

No model was supplied, defaulted to distilbert/distilbert-base-uncased-
finetuned-sst-2-english and revision af0f99b (https://huggingface.co/di
stilbert/distilbert-base-uncased-finetuned-sst-2-english).
Using a pipeline without specifying a model name and revision in produc
tion is not recommended.

config.json:    0%|          | 0.00/629 [00:00<?, ?B/s]

model.safetensors:    0%|          | 0.00/268M [00:00<?, ?B/s]

tokenizer_config.json:    0%|          | 0.00/48.0 [00:00<?, ?B/s]

vocab.txt:    0%|          | 0.00/232k [00:00<?, ?B/s]

In [49]: sent_pipeline('I love sentiment analysis!')
Out[49]: [{'label': 'POSITIVE', 'score': 0.9997853636741638}]

In [50]: sent_pipeline('Make sure to like and subscribe!')
Out[50]: [{'label': 'POSITIVE', 'score': 0.9991742968559265}]

In [51]: sent_pipeline('booo')
Out[51]: [{'label': 'NEGATIVE', 'score': 0.9936267137527466}]

In [ ]: # The End
```

Conclusion:

In conclusion, our sentiment analysis of Amazon customer reviews using transformer-based models and rule-based approaches has provided valuable insights into customer sentiments towards various products. Through rigorous data preprocessing, model training, and evaluation, we have successfully classified sentiments as positive, negative, or neutral, enabling businesses to better understand customer preferences and experiences.

Our analysis revealed the effectiveness of transformer-based models, particularly Roberta, in accurately classifying sentiments across a diverse range of reviews. These models outperformed traditional rule-based approaches like VADER, demonstrating the superiority of deep learning-based techniques in capturing complex language semantics.

By visualizing the results and comparing model performance using various evaluation metrics, we have equipped businesses with actionable insights to improve product offerings, enhance marketing strategies, and optimize customer engagement. The diverse distribution of sentiments across reviews underscores the importance of tailored approaches to address customer feedback effectively.

Moving forward, businesses can leverage the findings of our sentiment analysis to make informed decisions, prioritize areas for improvement, and enhance overall customer satisfaction. By embracing advanced NLP techniques and leveraging the power of data-driven insights, businesses can stay competitive in the dynamic landscape of e-commerce.

In essence, our sentiment analysis serves as a valuable tool for businesses to gain deeper insights into customer sentiments, drive strategic decision-making, and ultimately foster long-term success in the marketplace.

References

1. [Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.
2. Hugging Face. (n.d.). *Transformers Documentation*. Retrieved from https://huggingface.co/transformers/
3. Kaggle. (n.d.). *Amazon Fine Food Reviews*. Retrieved from https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews
4. Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543.
5. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is All You Need. *Advances in Neural Information Processing Systems (NeurIPS)*, 5998-6008.