Comparative Analysis of Sentiment in Original and Summarized Tweets: Leveraging Transformer Models for Enhanced NLP Insights

Kun Bu and Kandethody Ramachandran

Department of Mathematics and Statistics, University of South Florida, Tampa, FL 33620, USA

Abstract. This paper investigates the sentiments of Twitter users towards the emergent topic of ChatGPT, leveraging advanced techniques in natural language processing (NLP) and sentiment analysis (SA). Our approach uniquely incorporates a dual setting for sentiment analysis: one analyzes the sentiments of original, full-length tweets, while the other first condenses these tweets into succinct summaries before performing sentiment analysis. By employing this dual approach, we are able to offer a comparative analysis of sentiment assessment pre- and post-text summarization, exploring the accuracy and reliability of the summarized sentiments. Central to our methodology is the application of Transformer models, specifically ProphetNet, which facilitates a deeper and more nuanced understanding of the original text. Unlike traditional methods that rely on keyword extraction and aggregation, our approach generates coherent and contextually rich summaries, providing a novel lens for sentiment analysis. This research contributes to the field by presenting a comprehensive study comparing sentiment analysis outcomes between original texts and their summarized counterparts, and examining the effectiveness of different NLP techniques, namely NLTK and the Transformer-based ProphetNet model. The findings offer valuable insights into the dynamics of sentiment analysis in the context of social media and the efficacy of state-of-the-art NLP technologies in processing complex, real-world data.

Keywords: Sentiment Analysis, Natural Language Processing, Text Summarization, Machine Learning, Twitter Data Analysis, ProphetNet, Transformers

1 Introduction

The foundational architecture of ChatGPT is grounded in the GPT (Generative Pre-trained Transformer) model, a pioneering development by OpenAI. Unveiled in June 2018, the inaugural version of GPT featured an extensively trained deep neural network boasting 117 million parameters, utilizing a vast corpus of textual data for its foundational training.

Upon its release, GPT-3 was acclaimed as a significant milestone in the fields of natural language processing and artificial intelligence. Its proficiency in generating coherent, persuasive text across a diverse spectrum of subjects garnered

David C. Wyld et al. (Eds): AIBD, MLSC, ACSTY, NATP, CCCIoT, SVC, SOFE, ITCSS -2024
pp. 49-62, 2024. CS & IT - CSCP 2024 DOI: 10.5121/csit.2024.140404

widespread admiration, sparking enthusiasm over its myriad potential applications. Concurrently, the advent of such an advanced language model stirred a mix of apprehensions. Concerns ranged from the risks of technology misuse, particularly in spreading misinformation and fabricating news, to broader societal and ethical issues. These included the potential ramifications for employment and the economy, as well as the pressing matters of bias, privacy, and governance in AI.

In response to these dialogues, this study leverages Twitter data to conduct a sentiment analysis of public discourse surrounding ChatGPT. The innovative aspect of this research lies in fine-tuning the transformer model for text summarization, coupled with the enhancement of model performance through an ensemble approach. This methodology condenses tweets into concise statements. Sentiment analysis is then employed on both the original and summarized texts, serving as a validation mechanism to compare the efficacy of the summarization. This is achieved through the utilization of both fine-tuned and ensemble models. A baseline is established by labeling sentiments in the original tweets, with classifications applied to the sentiments of the summarized content. The results are quantitatively presented through ROC (Receiver Operating Characteristic) plots, showcasing the AUC (Area Under the ROC Curve) to illustrate the accuracy of the sentiment analysis.

2 State-of-the-Art

Sentiment analysis, tracing its origins to the early 2000s, has undergone significant evolution, propelled by technological advancements and the burgeoning availability of extensive labeled datasets. A pivotal development was the release of the "Polarity Datasets" by Bo Pang and Lillian Lee in 2004 [1], offering publicly accessible movie reviews annotated with positive or negative sentiments. Medhat et al. (2014) [2] conceptualize sentiment analysis as a classification endeavor, delineating three primary levels: document-level, sentence-level, and aspect-level. Our paper endeavors to utilize text summarization techniques, transforming document-level sentiment analysis (SA) into a sentence-level analysis, thereby facilitating a comparative assessment of SA classifications.

At the sentence level, SA primarily involves classifying sentiments expressed within individual sentences. The preliminary phase entails discerning subjective from objective sentences. For subjective content, the objective is to identify whether the sentiment is positive or negative. Wilson et al. [3] note that sentiment expressions may not always be subjective, yet it is recognized that there is no inherent dichotomy in classifying sentiments at document and sentence levels, as sentences can function as succinct documents [4].

In recent years, deep neural network (DNN) models have been increasingly applied in SA tasks, yielding promising outcomes. Basiri et al. (2021) [5] introduced the ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model, marking a notable advancement in the domain. Furthermore, transformer models like Google's BERT (Bidirectional Encoder Representations from Transformers) [6], introduced in 2018, have significantly revolutionized sentiment analysis. BERT employs attention mechanisms to grasp contextual relationships between words, enhancing sentiment detection accuracy. Microsoft ProphetNet [9] differs from traditional transformer models by focusing on n-gram prediction during training. While traditional transformer models like BERT or GPT predict the next word in a sequence one at a time, ProphetNet is designed to predict several future tokens simultaneously. Yadav and Vishwakarma (2020) [7] categorize traditional SA approaches into three types: (1) Lexicon-based methods, involving a precompiled sentiment dictionary to ascertain word-level sentiment; (2) Machine learning strategies, using handcrafted features to train classifiers for sentiment labeling; and (3) Deep learning techniques, deploying complex neural networks to extract abstract semantic features for SA [10].

Pre-trained models like BERT and ProphetNet, initially developed on extensive datasets, are fine-tuned for specific SA tasks. This paper extends this innovation by incorporating ProphetNet, an advanced transformer model, into both text summarization and sentiment analysis. ProphetNet stands out with its unique n-gram prediction mechanism, enhancing the depth and context of language processing. Unlike traditional models that predict the next word sequentially, ProphetNet predicts several future tokens concurrently, thereby offering a more comprehensive understanding of textual nuances.

The utilization of ProphetNet, in conjunction with established models, represents a significant leap in our approach to understanding user sentiments on platforms like Twitter. By deploying ProphetNet in text summarization, we aim to generate concise yet contextually rich summaries, thereby providing a more robust foundation for subsequent sentiment analysis. The model's effectiveness will be evaluated using ROC curves and AUC metrics, offering insights into the nuances of sentiment analysis post-summarization.

3 Method

3.1 Exploratory Data Analysis

The dataset utilized in this study comprises a meticulously curated collection of tweets, primarily identified through the utilization of hashtags such as #chatgpt, #gpt3, and #gpt4. These tweets predominantly revolve around the discourse on the ChatGPT language model, encompassing a spectrum of themes from user experiences and insights to queries and assistance requests pertaining to ChatGPT. Additionally, the dataset is enriched with tweets that feature links to ChatGPT-related articles, websites, and diverse forms of media including images, videos, and more.

In essence, the foundational dataset, aggregating tweets under the chatgpt hashtag, offers a comprehensive snapshot of the ongoing digital conversations about ChatGPT. Encompassing 74,783 unique tweets, the database is a rich repository of various attributes such as user_name, text, user_location, and more. For a focused analysis, this study narrows down to a specific subset of tweets emanating from verified Twitter accounts based in California. This subset, amounting to 97 tweets, facilitates an in-depth exploration of sentiments and perspectives about the newly introduced ChatGPT tool within a distinct demographic. Figure 1 delineates the demographic profile of these California-based Twitter users, providing valuable insights into the regional engagement with ChatGPT.

After the data cleaning, sentiment analysis is employed as a dual-phase validation method to critically assess the performance of text summarization, employing both fine-tuning and ensemble models. This procedure is systematically depicted in Figure 2, which illustrates the research workflow. Initially, the workflow involves the extraction of tweets specifically from verified Twitter users in California, a subset gleaned from the larger dataset. Subsequently, two parallel processes are initiated. The first process involves condensing each tweet from its original full-text form into a single, succinct sentence – an exercise in text summarization. The second process, concurrently executed, embarks on the preliminary round of sentiment analysis. This initial sentiment analysis categorizes each tweet – based on the original, unabridged content – into one of three sentiment classes: positive, negative, or neutral. Following these initial steps, the study progresses to a second round of sentiment analysis. This phase is applied to the previously summarized text, with the objective of assigning sentiment classifications, once again, into positive, negative, or neutral categories. This two-tiered approach to sentiment analysis allows for an intricate comparison between the sentiments derived from the original tweets and their summarized counterparts. The culmination of this research involves a critical juxtaposition of the sentiment analysis results obtained from both rounds. This comparative analysis is anchored on the premise of treating the initial round of sentiment analysis (based on full-text tweets) as the actual sentiment, against which the sentiments inferred from the summarized texts (second round) are evaluated as predicted outcomes. The insights gleaned from this comparison form the basis of the study's conclusions, offering a nuanced understanding of the efficacy and implications of text summarization in the realm of sentiment analysis.

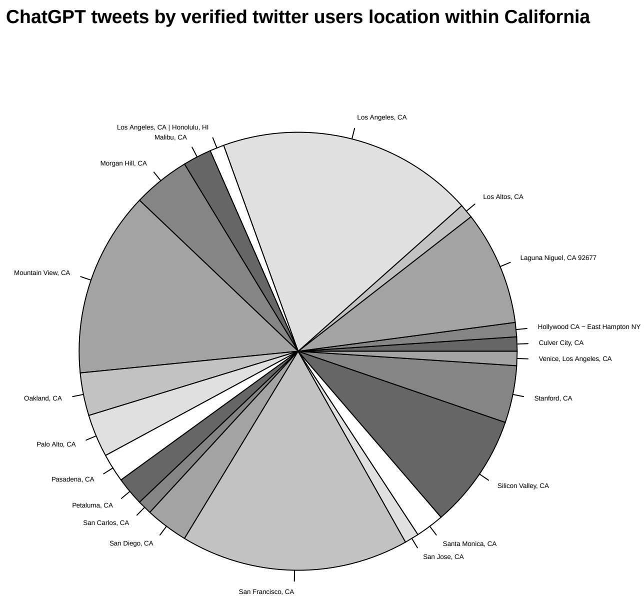

Fig. 1. Demographics of ChatGPT tweets by verified twitter users located within the State of California

3.2 Text Summarization

Text Summarization is a process that condenses a document or an article into a brief yet comprehensive version, encapsulating all crucial information. In our study, we implement this technique to transform original, full-length tweets into concise, single-sentence summaries. This is accomplished through the use of the advanced text summarization model ProphetNet, developed by Microsoft. Our primary objective is to compare the results of sentiment analysis performed both before and after the summarization process.

ProphetNet is a cutting-edge sequence-to-sequence model engineered by Microsoft Research Asia. It introduces an innovative self-supervised learning objective known as future n-gram prediction. Built upon the Transformer architecture, ProphetNet is adept at handling a range of natural language processing (NLP) tasks, including but not limited to text summarization, translation, and question-answering. One of its core strengths lies in its capability for bidirectional and auto-regressive (AR) generation (Sutskever, 2023)[10], rendering it exceptionally

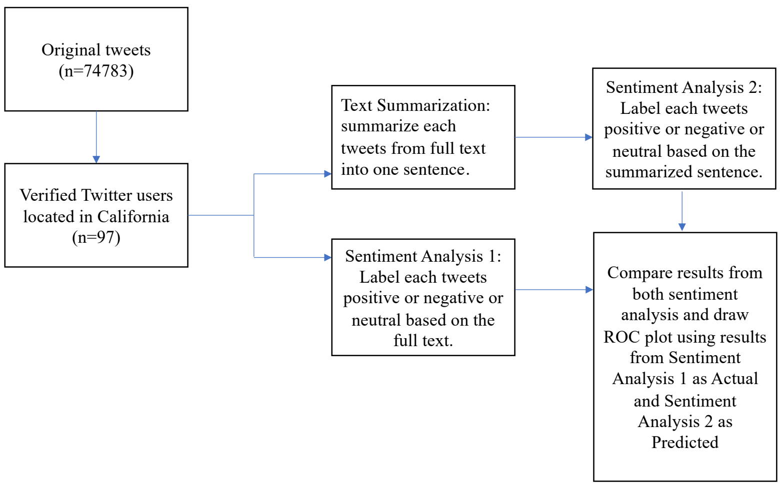

Fig. 2. A detailed flowchart for sentiment analysis workflow in Twitter data

versatile for sequence-to-sequence applications. Distinguishing itself from BART (Bidirectional and Auto-Regressive Transformers), ProphetNet integrates a novel pre-training approach termed "Masked Sequence-to-Sequence Pre-training." This method enhances traditional sequence-to-sequence models by optimizing for n-step ahead prediction, as opposed to the conventional one-step ahead approach. Such future n-gram prediction provides additional guidance, fostering a forward-looking model strategy and mitigating the risk of overfitting to strong local correlations (Qi, 2021)[9]. In our methodology, The ProphetNet model was then utilized to distill the full text of tweets into succinct summaries. These summaries were subsequently analyzed for sentiment using NLTK and Transformer models. Figure 3 below illustrates the architectural framework of the ProphetNet model, elucidating its operational mechanics.

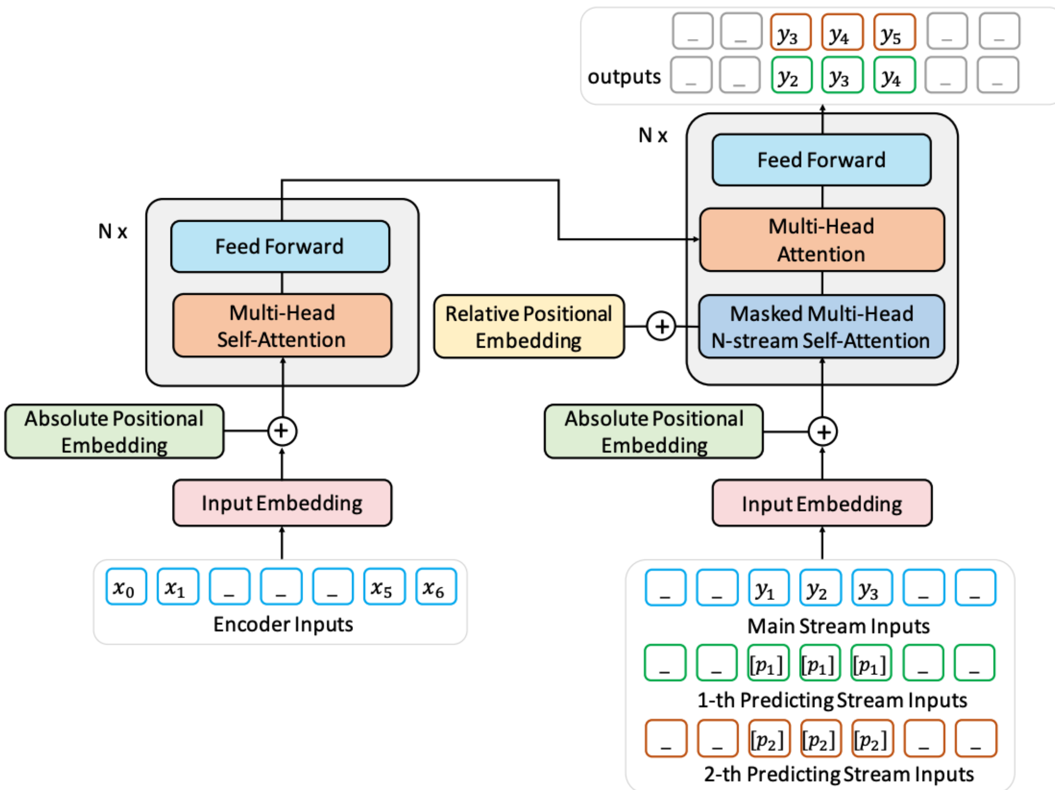

Fig. 3. The architecture of ProphetNet.(Qi, 2021)[9]

In Figure 3, the architecture of ProphetNet shows the travail example with just 2-grams, where the modeling goal is

$$p(y_t, y_{t+1}|y_{<t}, x) \tag{1}$$

for each step, where the $x = (x_1, x_2)$ is the given source sequence, and $p(y_t, y_{t+1}|y_{<t}, x)$ is the optimized next single token at each time step t, where $y_{t:t+n-1}$ denotes the next continuous n future tokens. The left part of this Figure shows the encoder which is the original Transformer encoder. The right part presents the decoder of the proposed n-stream self-attention. ProphetNet decoder predicts n future tokens simultaneously as:

$$p(y_t|y_{<t}, x), \dots, p(y_{t+n-1}|y_{<t}, x) = \textbf{Decoder}(y_{<t}, \mathcal{H}_{enc}) \tag{2}$$

where for $n = 2$, the future token becomes:

$$p(y_t|y_{<t}, x), p(y_{t+1}|y_{<t}, x) = \mathbf{Decoder}(y_{<t}, \mathcal{H}_{enc}) \quad (3)$$

3.3 Sentiment Analysis

Furthermore, we implement sentiment analysis utilizing two distinct methodologies: the lexicon-based approach, leveraging the robust Natural Language Toolkit (NLTK), and the state-of-the-art BERT models from the Transformer suite. The NLTK method involves attributing sentiment scores to individual words, aggregating them to derive an overall sentiment score for the text. This approach is one of two primary methods in automated sentiment extraction. The lexicon-based approach, as described by Turney (2002) [11], calculates a document's sentiment orientation based on the semantic orientation of words or phrases within it. Conversely, the text classification method, outlined by Pang, Lee, and Vaithyanathan (2002) [12], involves constructing classifiers from labeled text instances, aligning with supervised classification paradigms. This latter method can also be framed as a statistical or machine-learning approach.

Our research predominantly follows the lexicon-based strategy, employing dictionaries of words annotated with semantic orientations or polarities (Taboada, 2011) [13]. Liu (2010) [14] elucidates that lexicon-based methods generally utilize the predominant orientation of opinion words within a sentence to determine its overall sentiment. Therefore, if positive or negative opinions dominate, the sentence is classified accordingly. In instances where positive and negative opinion words are equally present, the sentence's sentiment is predicted based on the average orientation of effective opinions or the sentiment of the preceding opinion sentence.

When applied to sentiment analysis, BERT can classify text as positive, negative, or neutral with high accuracy. This classification is not just based on specific sentiment words but also on how these words interact and form meaning in context. For example, BERT can distinguish between "I am happy" and "I am not happy," understanding the negation in the second sentence and classify "not happy" as a negative sentiment. Compared to earlier models, BERT's bidirectional approach and deeper contextual understanding allow for more accurate sentiment analysis, especially in cases where context drastically changes the sentiment conveyed by specific words or phrases. This motivated us to apply BERT in our sentiment analysis.

This dual-method approach, combining the nuanced capabilities of NLTK with the advanced contextual understanding of BERT, provides a comprehensive sentiment analysis framework. Our methodology is designed to accurately capture and reflect the sentiment embedded in textual data, facilitating a deeper understanding of public opinion dynamics.

4 Results

This section is divided into two major phases namely Transformer Based Sentiment Analysis and NLTK Based Sentiment Analysis, each part offers a comparative analysis of sentiment assessment pre- and post-text summarization.

4.1 Transformer Based Sentiment Analysis

This dual-bar chart comparison of sentiment analysis results using a transformer model on full-text data in Figure 4 shows the left chart displays the count of sentiments categorized as neutral, positive, and negative, with neutral sentiments appearing most frequently, followed by positive sentiments, and a relatively small count of negative sentiments. The right chart represents the same sentiment data but expressed as a percentage of the total sentiments analyzed. Here, neutral sentiments constitute the majority, occupying slightly over 50% of the dataset. Positive sentiments account for nearly half of the remaining distribution, while negative sentiments make up a minimal portion, indicating a lesser prevalence of negative sentiment in the analyzed tweets.

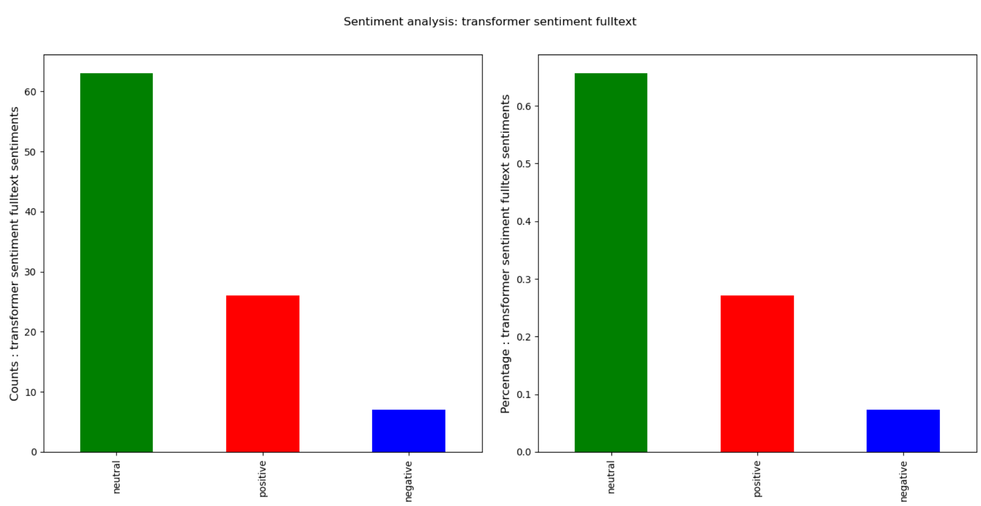

Fig. 4. Sentiment analysis for full text using Transformer

Figure 5 the results of sentiment analysis performed on summarized texts using a transformer model. On the left, we see the absolute count of sentiments categorized as neutral, positive, and negative. The neutral category predominates with a significantly higher count, followed by a moderate number of positive sentiments, and a relatively low count of negative sentiments. On the right, the same sentiment data are represented as a proportion of the total, providing a percentage breakdown of each sentiment category. Here again, the neutral sentiment constitutes the majority, exceeding 60% of the total, with positive sentiments representing a sizable minority, and negative sentiments appearing as the least common, forming a small fraction of the dataset.

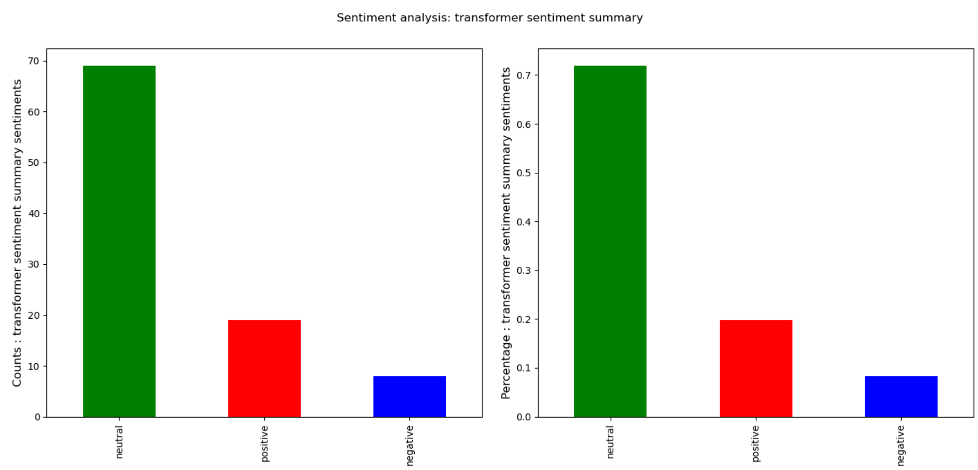

Fig. 5. Sentiment analysis for summarized text using Transformer.

4.2 NLTK Based Sentiment Analysis

In the NLTK method, the frequency and percentage of the results in sentiment analysis based on the whole vitrified California user's tweets shows that more than 55% of twitter users hold a positive attitude of ChatGPT, about 35% neutral and more than 10% negative. The distribution of the sentiment analysis for the summarized text results is nearly 70% neutral,approximately 25% positive and less than 5% negative.The results show that compared with full text, summarized text tends to be more neutral than positive sentiment.

4.3 Validation

The analysis indicates an equivalent proportion of negative sentiment across both full texts and their summarized counterparts when processed by the NLTK model.

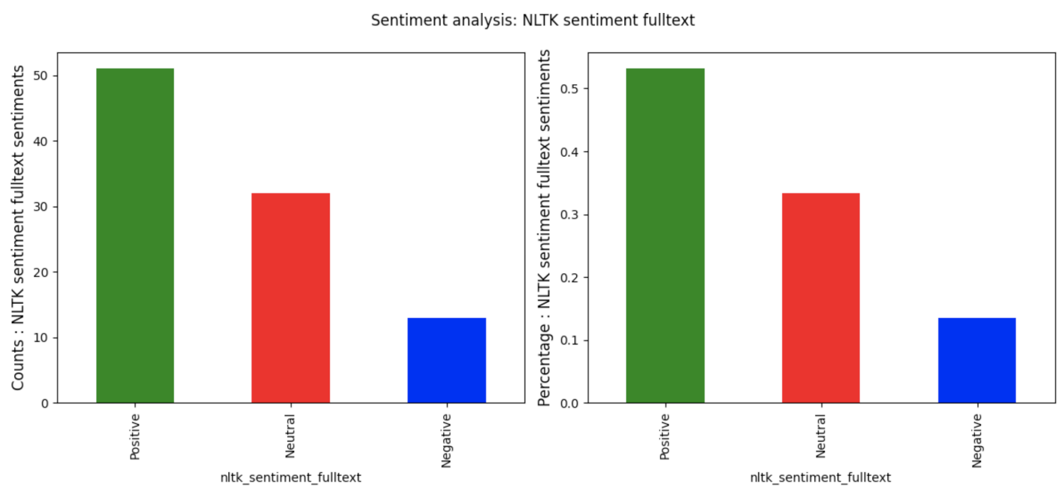

Fig. 6. Sentiment analysis for full text using NLTK.

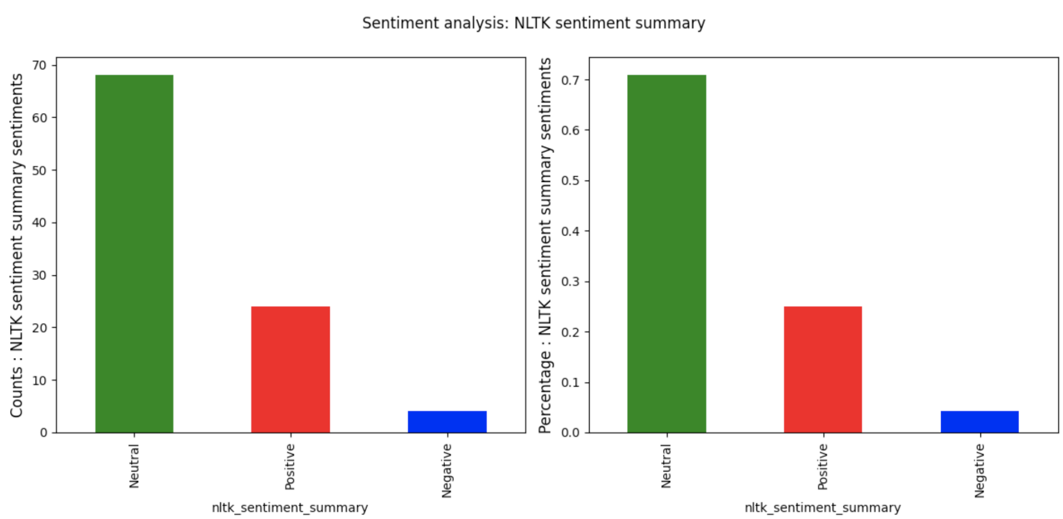

Fig. 7. Sentiment analysis for summarized text using NLTK.

However, the summarized texts exhibit a higher inclination towards neutrality compared to the full texts. This trend towards neutrality in summaries can be attributed to the primary objective of summarization, which is to distill the core content of texts into a succinct and impartial synopsis. Summaries endeavor to eschew subjective bias, striving instead to encapsulate the key points in an even-handed fashion. In contrast, the Transformer model demonstrates a more consistent sentiment distribution between full texts and their summaries, effectively mitigating any discrepancy in sentiment that arises from the summarization process.

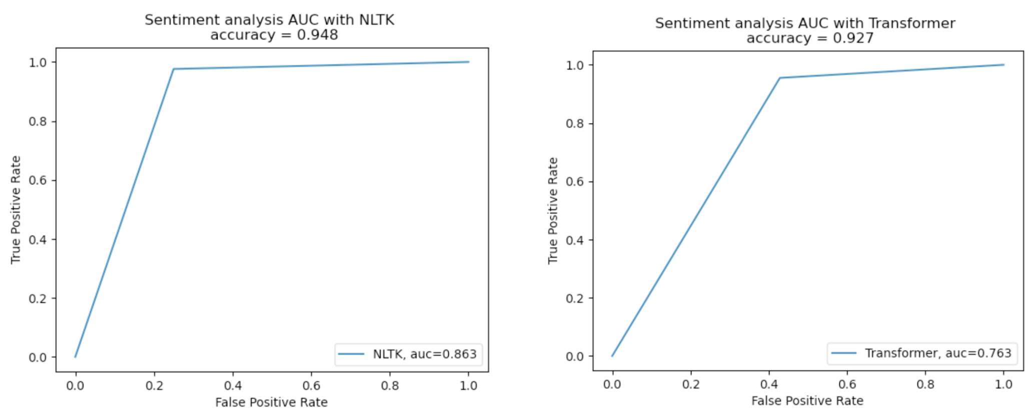

Fig. 8. ROC for NLTK sentiment prediction vs. ROC for Transformer sentiment prediction.

In the validation process, we plot the ROC for both NLTK sentiment analysis results and Transformer sentiment results by combining neutral sentiment with positive sentiment as one class, and negative sentiment as another class. Then we use the summarized sentiment analysis results as predicted and full text sentiment analysis results as labels since full text tweets contain more information than summarized sentences. The comparison shows in Figure 8, where the NLTK method provides an accuracy of 0.948 and AUC value of 0.863, while the Transformer method has an accuracy of 0.927 and AUC of 0.763.

5 Conclusion

This research paper presents an analysis of Twitter users' opinions towards ChatGPT, employing various sentiment analysis approaches. The findings indicate that the majority of users exhibit positive or neutral attitudes towards ChatGPT. To conduct the analysis, we utilized the latest text summarization technique to condense full tweets into concise sentences. Subsequently, we employed both NLTK and Transformer models to analyze sentiment in the original full-text tweets as well as the summarized versions.

The results demonstrate the effectiveness of both sentiment analysis approaches in accurately classifying positive and negative tweets. Moreover, the analysis results align consistently before and after summarization when utilizing the Transformer model. However, the NLTK approach tends to predict a higher proportion of neutral sentiments in the summarized tweets. This can be attributed to the current objective of summarization, which aims to provide an objective and concise representation of the main points, avoiding bias and subjective interpretations. Neutral summaries focus on presenting essential information in a balanced manner. Furthermore, during the validation process, the Receiver Operating Characteristic (ROC) curves illustrate that the predicted outputs of text summarization achieve over 90% accuracy in both NLTK and Transformer models, highlighting the significance of our research approach.

In conclusion, this study employs state-of-the-art natural language processing (NLP) methods to explore the sentiments of Twitter users towards ChatGPT. It sheds light on the challenge of accurately differentiating between neutral and positive sentiments in sentiment analysis.

References

1. Pang, Bo, and Lillian Lee. "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts." arXiv preprint cs/0409058 (2004).
2. Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams engineering journal 5.4 (2014): 1093-1113.
3. Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann. "Recognizing contextual polarity in phrase-level sentiment analysis." Proceedings of human language technology conference and conference on empirical methods in natural language processing. 2005.
4. Liu, Bing. "Sentiment analysis and opinion mining." Synthesis lectures on human language technologies 5.1 (2022): 1-167.
5. Basiri, Mohammad Ehsan, et al. "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis." Future Generation Computer Systems 115 (2021): 279-294.
6. Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
7. Yadav, Ashima, and Dinesh Kumar Vishwakarma. "Sentiment analysis using deep learning architectures: a review." Artificial Intelligence Review 53.6 (2020): 4335-4385.
8. Chan, J.YL., Bea, K.T., Leow, S.M.H. et al. State of the art: a review of sentiment analysis based on sequential transfer learning. Artif Intell Rev 56, 749–780 (2023).
9. Qi, Weizhen, et al. "Prophetnet: Predicting future n-gram for sequence-to-sequence pre-training." arXiv preprint arXiv:2001.04063 (2020).

10. Chan, J.YL., Bea, K.T., Leow, S.M.H. et al. State of the art: a review of sentiment analysis based on sequential transfer learning. Artif Intell Rev 56, 749–780 (2023).
11. Turney, Peter D. "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews." arXiv preprint cs/0212032 (2002).
12. Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques." arXiv preprint cs/0205070 (2002).
13. Taboada, Maite, et al. "Lexicon-based methods for sentiment analysis." Computational linguistics 37.2 (2011): 267-307.
14. Liu, Bing. "Sentiment analysis and subjectivity." Handbook of natural language processing 2.2010 (2010): 627-666.

Authors

Kun Bu received Master of Arts in Statistics from University of South Florida in 2020. Currently, she is pursuing her Ph.D. in Statistics from the University of South Florida. Her research interests include machine learning, statistical reinforcement learning (self-learning feedback mechanism) and natural language processing.

Dr. Kandethody Ramachandran received Ph.D. in Applied Mathematics from Brown University in 1988. Currently, he is a Professor of Mathematics and Statistics at the University of South Florida. His research interests are concentrated in the areas of applied probability and statistics. His research publications span a variety of areas such as control of heavy traffic queues, stochastic delay systems, machine learning methods applied to game theory, finance, cyber security, and other areas, software reliability problems, applications of statistical methods to microarray data analysis, and streaming data analysis.