ChatGPT Sentiment Analysis
Final Technical Report

Team 1: Mitch Breeden, Rock Chan, Brian Monter, Eliza Punnoose

Professor: Maryam Heidari

AIT 526-002: Natural Language Processing

George Mason University

07/10/2023

Abstract

In the beginning days of ChatGPT's popularity, there were a lot of unknowns about what it could do, how it could help individuals, and where it was going from here. Throughout this paper, we will consider various questions, the two main being; What approach is best for determining sentiment for a tweet? And what is the initial sentiment surrounding ChatGPT on twitter based on the tweets in the specified range? The two approaches that were used to determine sentiment were the Pattern approach and the Vader approach. While both seemed to agree that the initial sentiment surrounding ChatGPT fell somewhere between neutral and positive, it may surprise you which one we determined was the superior approach. After the analysis was done and understood, it was determined that Pattern was the more applicable approach in our case, although Vader is specifically designed for social media text. Based on the beginning days of ChatGPT's popularity, the initial sentiment shows it is useful, helpful, and here to stay.

Introduction

ChatGPT is an advanced language model capable of understanding and generating text in a conversational context. It can answer questions, provide explanations, offer suggestions, and engage in interactive conversations. However, opinions about ChatGPT vary among different individuals and communities. There is certainly positive feedback on its capabilities and applications including education, research, creative writing, and general knowledge sharing. However, with the positives also comes the negatives; there are concerns about misinformation, ethical considerations, as well as its limitations.

Sentiment analysis, also known as opinion mining, is the process of determining whether a piece of text is positive, negative, or neutral. In our project, we will use sentiment analysis to gauge the human perception about ChatGPT to identify areas of concern and address them accordingly. The dataset that we have chosen contains roughly 100,000 tweets over a period of 3 days, which tells us how relevant ChatGPT is in today's landscape. The problem we are considering is interesting because of the capabilities of language models present in today's age. The trending word in the AI world is ChatGPT, and any positive advancement in it will help the user experience.

Related Work

Taecharungroj, 2023 is a study analyzing tweets about ChatGPT, an AI chatbot, in the first month after its launch. The researchers used the Latent Dirichlet Allocation (LDA) topic modeling algorithm to identify the functional domains and potential impacts of ChatGPT. LDA is a generative probabilistic model in which the basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words (Latent Dirichlet Allocation). This article conducted extensive study on ChatGPT's functionality and user experience. It lacks sentiment analysis to determine how people are responding, including how many people like and dislike ChatGPT. The analysis determined that ChatGPT can potentially impact humans and technological advancement in both positive and negative ways. The author discussed four potential important issues that will need to be addressed: the evolution of jobs, a new technological landscape, the quest for artificial intelligence, and the progress-ethics conundrum.

ANIL. 2023, March 25 is a sentiment analysis from Kaggle. The project's objective is to learn what Twitter users think and prefer about ChatGPT, one of the most popular apps right now. The project uses a variety of visualization graphs to demonstrate the conceptual foundation of ChatGPT. In our project, we'll aim to make this project longer and look into various sentiment analysis approaches to enrich it. We are going to predict the sentiment level of the tweets in the dataset.

Objectives

As stated above, our main objective is to understand the human perception of ChatGPT by sentiment analysis. We will:

- Identify opinions
- Sort these opinions into relevant categories
- Identify the negative and neutral opinions
- Determine the best sentiment analysis approach

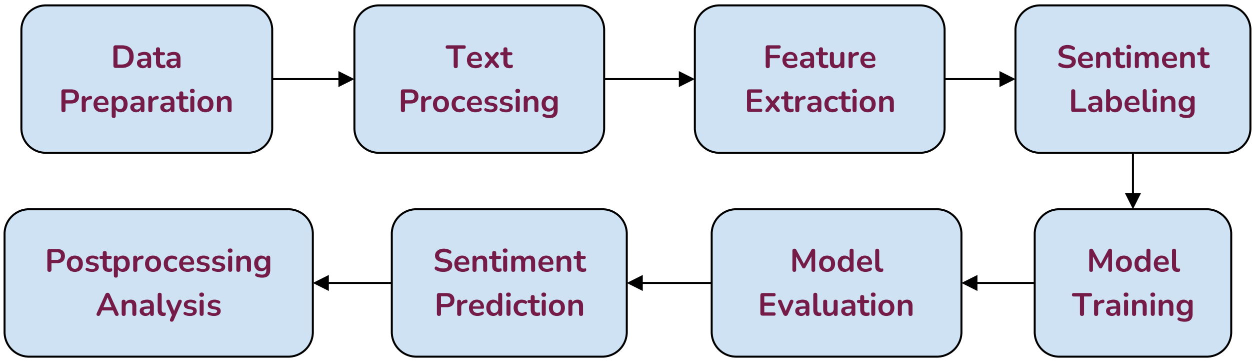
Selected Dataset

The dataset we have chosen is around 100,000 tweets in English and only contains the tweet itself; usernames, tags, and links are deleted or masked. Here are the variables that the data contains:

- ID: unique tweet id
- Date: date the tweet was sent
- Username: username of the person who tweeted (masked for this data and non-real ids are generated)
- Tweet: the content of the tweet (tags and links deleted)
- ReplyCount: number of replies to tweets
- RetweetCount: number of retweets to tweets
- LikeCount: number of likes to tweets
- QuotesCount: number of quotes to tweets

Proposed System

We will be attempting to go through the below-mentioned steps for our analysis.



During these steps we will use NLP to tokenize words and phrases, word frequencies, n-grams, etc. We are using a labeled dataset of tweets with sentiment labels (positive, negative, and neutral). The sentiment of tweets produced by ChatGPT may therefore be predicted using a machine learning model such as Naive Bayes which will allow us to predict the sentiment whereas other solutions focus on the backend of ChatGPT itself.

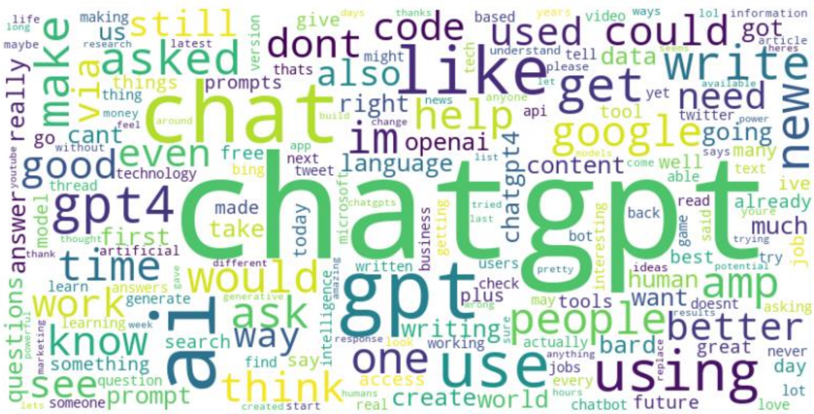
The System

The entirety of the system was built using Python. Our system is built to perform sentiment analysis on text data, specifically tweets containing the word ChatGPT using various natural language processing (NLP) libraries and tools. We imported various packages such as pandas, nltk, string, numpy, emoji, textblob, wordcloud, preprocessor, and matplotlib in order to further manipulate and visualize the data, as well as perform sentiment analysis on the data. We used a transformer-based sentiment analysis model by importing the AutoTokenizer and AutoModelForSequenceClassificaion from the transformers library, which sets up a pre-trained model used for sentiment analysis.

From there, we proceeded to understand and clean the data before doing any further analysis on it. We pulled various statistics about the dataset such as column count, row count, data types, missing values, and duplicates. In addition, we divided up the “timestamp” column into separate “Date”, “Hour”, and “Min'' columns which allowed for easier analysis. Next, we removed any emojis that were in the tweets and filtered out any tweets that were not in English. Then we defined a function to preprocess the tweets that would be preprocessed in a new column as the cleaned version of the tweets. The preprocessing steps included lowercase conversion, punctuation removal, handling of dots and spaces, removal of words starting with certain characters, hashtag processing, character repetition reduction, and stop word removal.

Experimental Results and Analysis

Top Words for Positive Sentiment



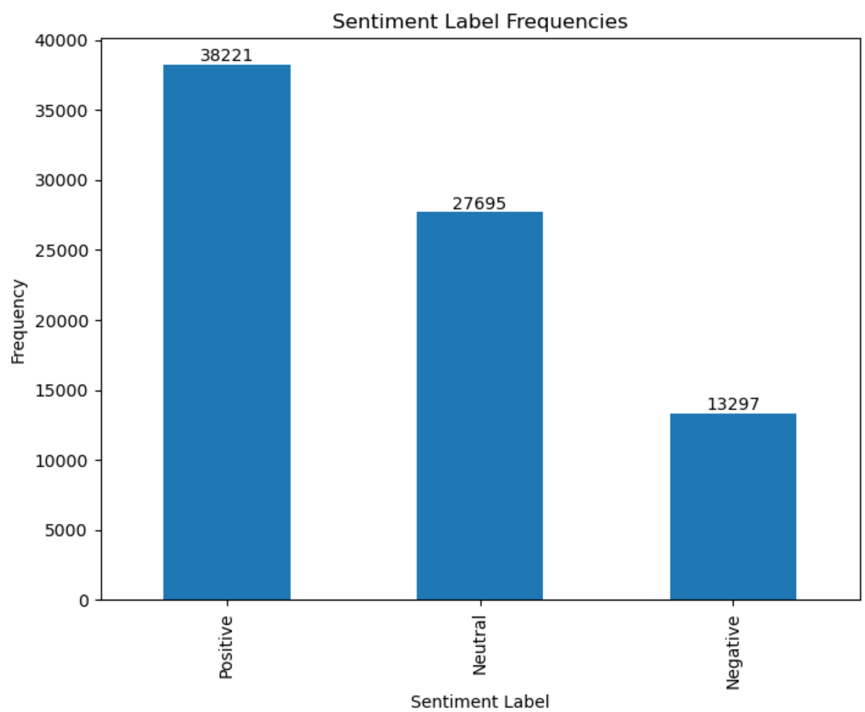
Top Words for All Sentiment

The first visualizations that were created were four different word clouds. One for all sentiment, one for positive sentiment, one for neutral sentiment, and one for negative sentiment. By looking at the first word cloud, top words for all sentiment, it is noticeable that the top words shed a positive and useful light on the use of ChatGPT. Some of the words that give this sentiment include like, help, better, good, and use. The top words for positive sentiment word cloud are very similar to the top words for all sentiment.

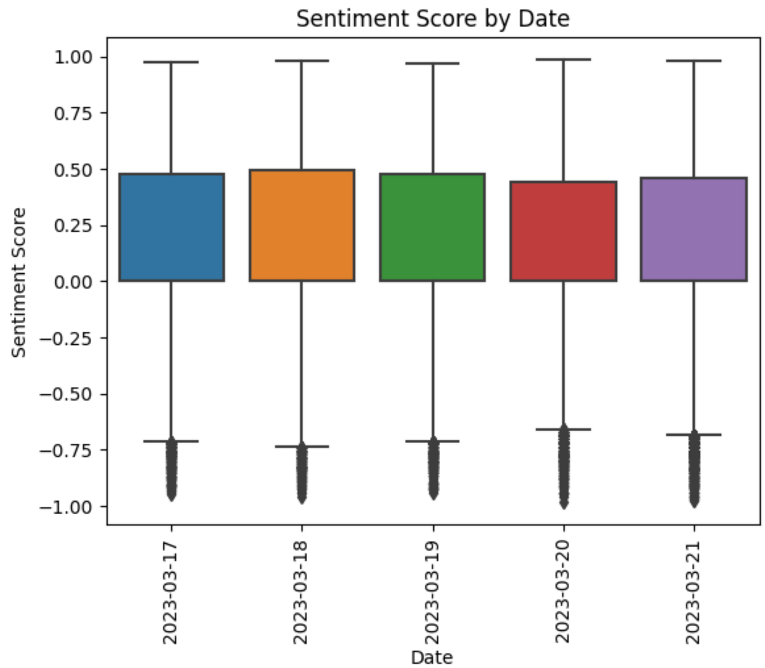
Top Words for Neutral Sentiment

Top Words for Negative Sentiment

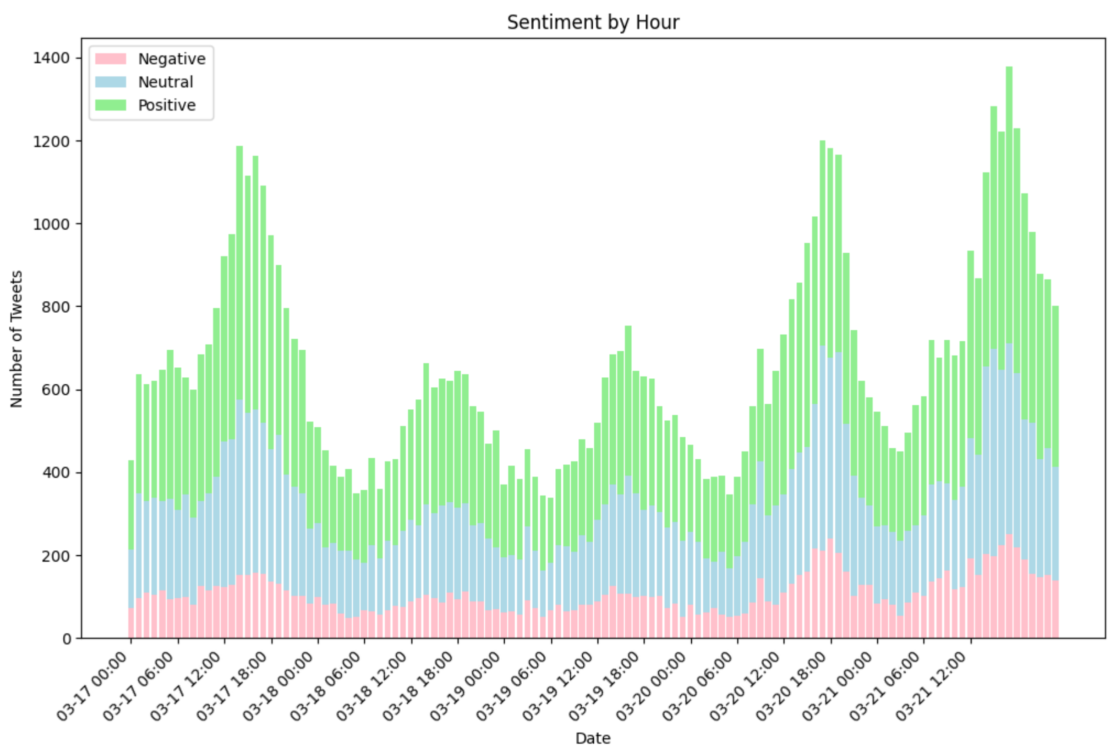
When looking at the last two word clouds, the top words for neutral sentiment and the top words for negative sentiment, you see many of the same words but a few stand out. Some of these words include chat, asked, bad, people, think, and write. These words indicate that users had a neutral or negative experience when asking it to perform some specific tasks such as writing.



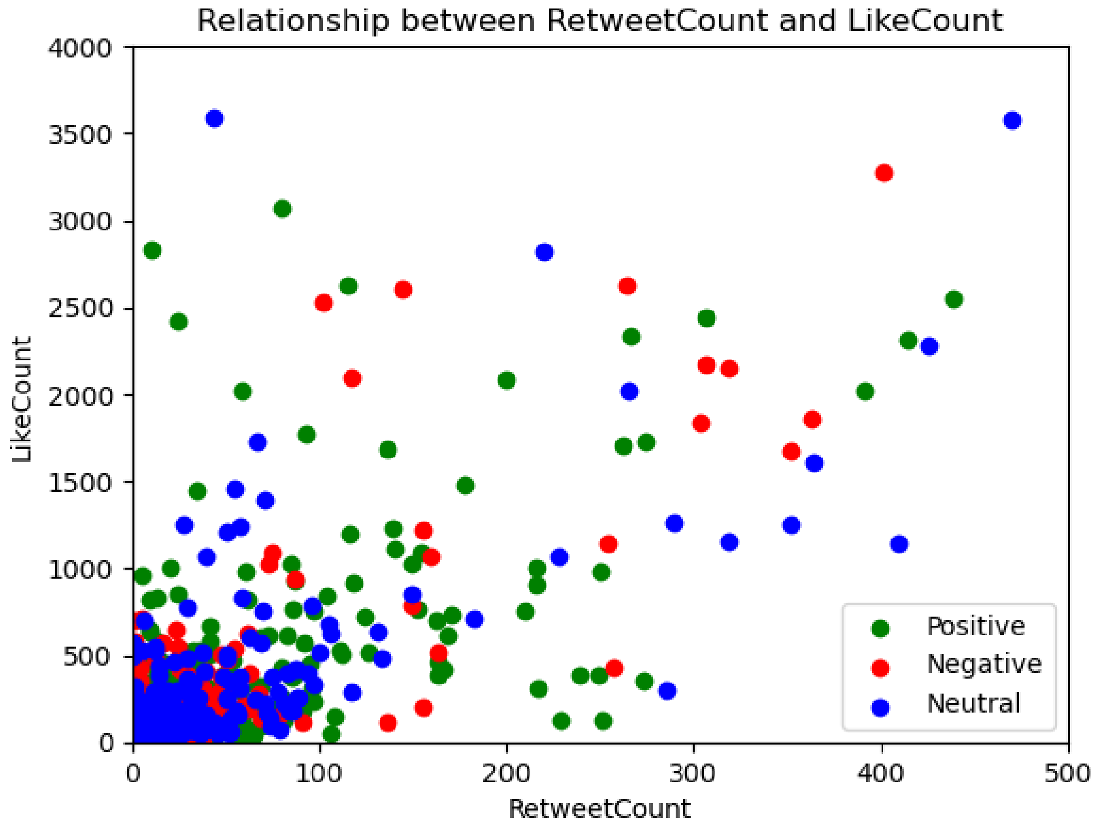
From there, it was important to see how often each sentiment label is coming about in the tweets. A column bar plot was created to show the frequency of each sentiment label in the data. As you can see, 38,221 labels were identified as positive, 27,695 were identified as neutral, and 13,297 were identified as negative. This again indicated that users tend to have a more positive experience compared to a negative one with ChatGPT.



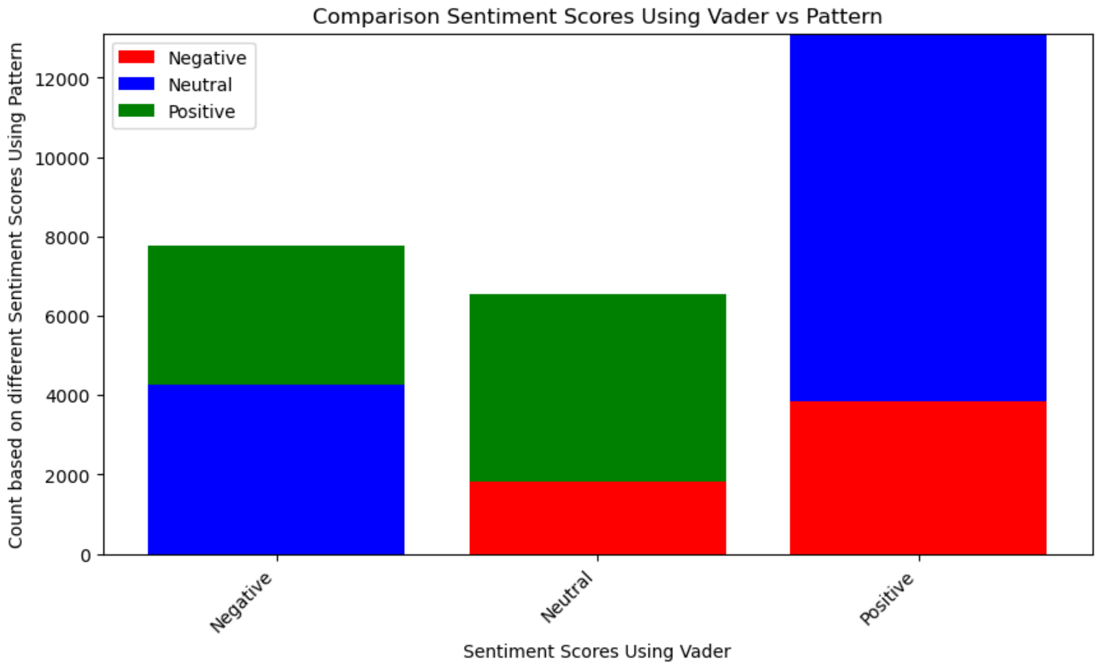
Next, we looked at sentiment by the hour and by the day. The first graph is a box graph of the sentiment score by the date. The sentiment score is on a scale from -1 to 1 and 50% mean is 0 to 0.5. The range of the sentiment score is from -0.75 to 1.



When we looked at the sentiment by the hour, some interesting insights could be seen. First, less people seem to be on twitter over the weekend (March 18th and 19th in this case). We also noticed that more people tend to be on twitter and sharing their opinions around 12:00pm each weekday, beware to all the workplace managers out there. There still seems to be a clear difference between the amount of positive or neutral tweets compared to the negative tweets.

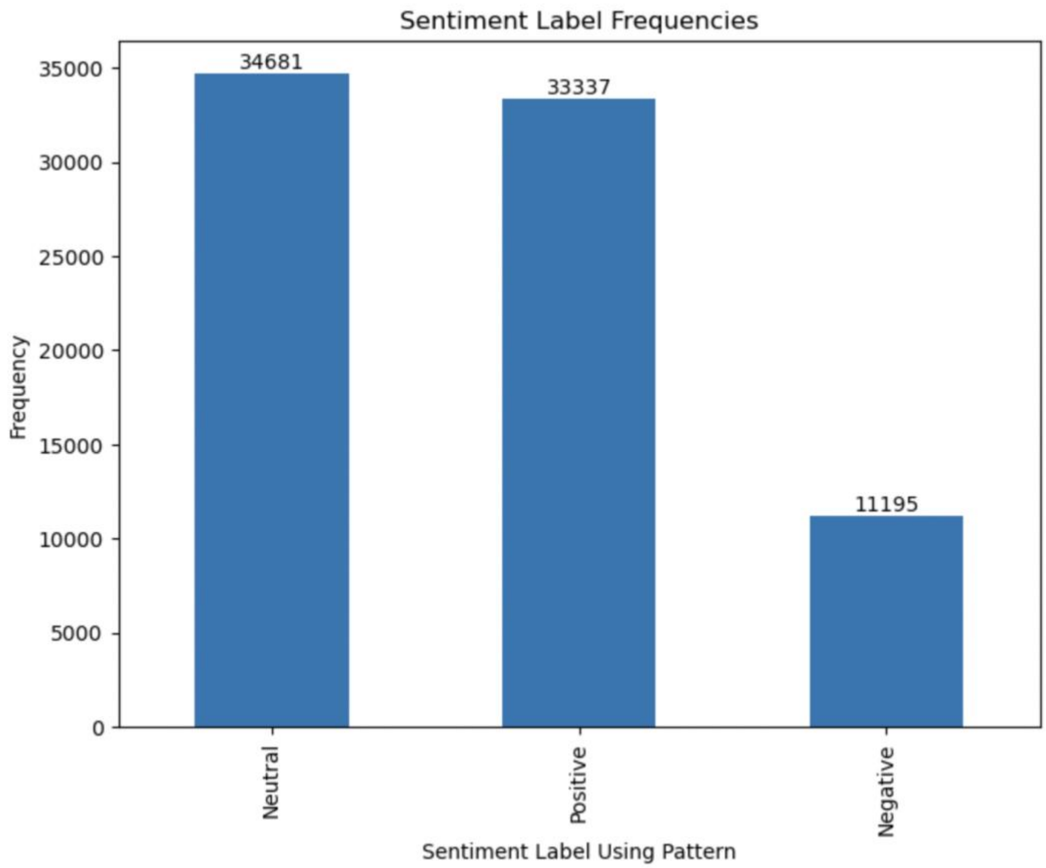


The number of tweets related to each sentiment may not tell the whole story. In addition to the above visualizations, the next visualization that was considered was the relationship between Retweet Count and Like Count, and what sentiment each tweet was. There is no clear pattern between how many retweets a tweet gets versus how many likes it may get, the points in the graph are scrambled. What can be seen is that the most liked tweet and the most retweeted tweet were both identified as neutral. Additionally, there seems to be more green (positive) points that were liked and retweeted more than the negative sentiment tweets.



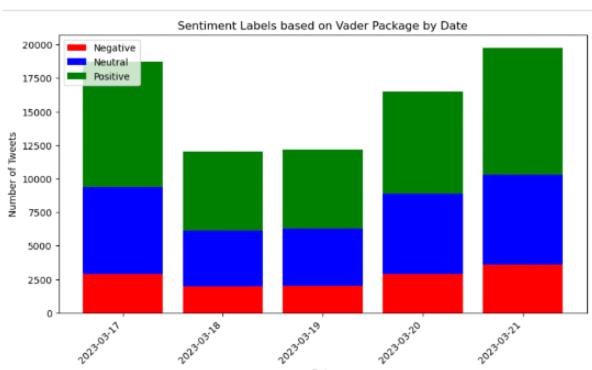
We then looked at a bar chart that compared the Pattern approach to the Vader approach for sentiment analysis. Pattern is a text data mining tool that has a text processing component. On the other hand, Vader's approach is specifically attuned to handle social media text (it can even decode the sentiment of emojis, although we took those out). This graph takes only those tweets where the sentiments identified by both Vader and Pattern mismatched. The figure shows for each sentiment identified by Vader, the count of sentiments that were redistributed with the pattern package. For reference, the following table shows that from the mismatched data, tweets which were marked as "Negative" by Vader 7,764 tweets were marked differently by Pattern and how they were distributed between Neutral and Positive.

Sentiment_Pattern_Score	Negative	Neutral	Positive
Sentiment			
Negative	0	4264	3500
Neutral	1805	0	4732
Positive	3857	9259	0



The frequency distribution shows that most of the tweets which were marked as positive or negative using VADER were marked neutral using Pattern. Manually checking the classification, the sentiment identification by Pattern was more accurate.

Comparing the mismatches based on date:



```
Sentiment
Negative    13297
Neutral     27695
Positive    38221
```

```
Sentiment_Pattern_Score
Negative    11195
Neutral     34681
Positive    33337
```

A total of ~27k mismatches between the 2 packages were identified.

Conclusions

The tweets surrounding ChatGPT between March 18th, 2023 and March 21st, 2023, specifically ones with ChatGPT within the tweet, were generally considered to be more positive or neutral in terms of the sentiment of the tweet. While there was a good amount of negative sentiment around ChatGPT within that time frame too, most of the sentiment leaned the other way. After comparing two approaches to identify sentiment, Pattern and Vader, we concluded that Pattern was a much superior approach.

References

Taecharungroj, V. (2023, February 16). “What Can ChatGPT Do?” Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. Big Data and Cognitive Computing; MDPI. https://doi.org/10.3390/bdcc7010035

ANIL. (2023, March 25). Twitter Sentiment Analysis for ChatGPT. Kaggle. https://www.kaggle.com/code/sanlian/twitter-sentiment-analysis-for-chatgpt

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003, January). *Latent dirichlet allocation - Journal of Machine Learning Research*. Latent Dirichlet Allocation. https://jmlr.org/papers/volume3/blei03a/blei03a.pdf

GeeksforGeeks. (2023, April 6). *What is sentiment analysis?*. GeeksforGeeks. https://www.geeksforgeeks.org/what-is-sentiment-analysis/

Sahagian, G. (2021, April 13). *What NLP Library you should use for your sentimental analysis project*. Medium. https://medium.com/geekculture/what-nlp-library-you-should-use-for-your-sentimental-analysis-project-bef6b357a6db