Chatgpt Tweets Analysis

ChatGPT, a large language model developed by OpenAI. I have been trained on a massive amount of data using deep learning algorithms, allowing me to understand and generate natural language in a way that is similar to how humans communicate. My primary function is to answer questions, engage in conversation, and provide assistance on a wide range of topics.

Data Source:

* Twitter

Method of Data Extraction:

* Snscrpe with the topic of "chatgpt" and the max result is 10000 and saving results as "tweets.json".

Data Preprocessing:

Data preprocessing for our project splitted into two sections one is preprocessing and second is postprocessing.

* Preprocessing includes:
* Fix contractions
* Remove URLs
* Remove non-ASCII characters
* Remove user-handles
* Remove hashtags
* Remove multiple spaces

In preprocessing we prepared data for sentiment analysis, so we do not remove stop words, because stop words carry the sentiment of the sentence, if we remove them our results definitely be affected. So, we refrain from drawing them out.

Sentiment Analysis:

Sentiment analysis is the process of using natural language processing (NLP) and machine learning techniques to automatically identify and extract subjective information from text, such as opinions, emotions, attitudes, and feelings. The goal of sentiment analysis is to determine whether a piece of text expresses a positive, negative, or neutral sentiment.

 TextBlob:

We used Textblob for sentiment analysis. TextBlob is a Python library for processing textual data that includes a sentiment analysis feature. It is built on the Natural Language Toolkit (NLTK) and provides an easy-to-use interface for common NLP tasks, such as part-of-speech tagging, noun phrase extraction, and sentiment analysis. TextBlob's sentiment analysis feature uses a pre-trained Naive Bayes classifier to analyze the polarity of a text, which ranges from -1 (negative) to 1 (positive). In addition to sentiment analysis, TextBlob also supports translation and text classification.

Justification:

Chatgpt has changed the style of content on the internet. You do not have to scroll web pages. You can get exact results with minimal effort and time. In sentiment analysis, we want to analyze consumer views on chatgpt. After the analysis, we can say that people are taking the new technology in positive ways.

Emotion Analysis:

For emotion analysis, we used the pre-trained model EmoRoBERTa model. EmoRoBERTa is a deep learning model for emotion recognition in text. It is an extension of the RoBERTa language model, which is a state-of-the-art pre-trained neural network for natural language processing tasks. EmoRoBERTa is designed to recognize emotions in textual data and has been trained on a large dataset of human-labeled emotion-laden text.

Justification:

Emotion analysis can help to identify the emotional state of an individual or a group of people, such as their mood, tone, attitude, and sentiment. The analysis can be used to measure the intensity of emotions and track how they change over time. Emotion analysis has applications in many areas, including marketing, customer service, healthcare, and social media analysis. For example, companies can use emotion analysis to analyze customer feedback and understand their needs and preferences better. Healthcare professionals can use it to monitor the emotional state of patients and provide better care. By emotion analysis, we try to figure out the emotions of people towards chatgpt. Are they afraid of it, feeling excited, feeling sad, and so on? After the analysis, we can say most of the people are very excited about chatgpt.

Postprocessing:

In postprocessing we removed the followings:

* Remove short words (i.e 3 characters long)
* Remove stop words
* Remove numbers
* Remove special characters

Topic Modeling:

Topic modeling is a technique used in natural language processing to extract underlying topics or themes from a collection of text documents. It is an unsupervised learning approach that allows us to discover latent patterns and structures in a large corpus of text data. Topic modeling can be used to identify the most important topics in a collection of documents, to understand the relationships between different topics, and to discover the underlying structure of the text.

 LDA

Latent Dirichlet Allocation (LDA) is a probabilistic topic modeling technique that is widely used for topic modeling in NLP. LDA assumes that each document in a corpus is a mixture of a small number of topics, and each topic is a distribution over words in the corpus. LDA works by learning the distribution of topics across the documents and the distribution of words across the topics.

Justification:

The intuition for topic modeling is to figure out the topic of discussion by Twitter users. We draw out the top 10 topics in most of the topics there was chatgpt on the top. People keenly talking about chatgpt.

Twitter Bias:

We scrapped the dataset using snscrape. In scrapped results, there were no details about gender, mostly null values for the place and no data for age or other personal information of a Twitter user. But when analyzing the data we opened with the iPhone for Twitter stands at 2nd position as the source of tweets.