

SUMMARY EXTRACTION ALONG WITH SENTIMENT ANALYSIS

Group-2

Members:

- ❖ Pooja Rani Nayak
- ❖ Mrudulaiswarya Subha Vyakaranam
- ❖ Venkata Sandeep Pesala
- ❖ Sagar Mahadev Panchal
- ❖ Mani Aditya Tellakula
- ❖ Prabhu Kumar Sambar
- ❖ Sandeep Kumar Balaga

MENTOR

KARTHIK

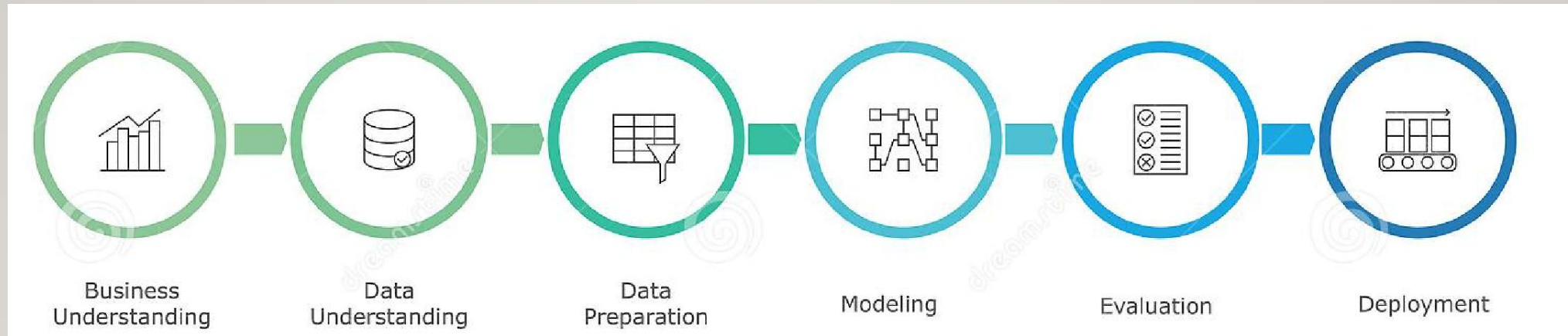
BUSINESS PROBLEM

To extract Summary of E-books along with sentiment analysis.

Objective

To extract E-books of your choice and extract summary ,categorize summary as positive , negative or neutral. Build a NLP model to achieve the said objective with accepted accuracy of 75% and above.

PROJECT FLOW



Text Pre Processing

Stop Words:

The words which are generally filtered out before processing a natural language are called stop words. These are actually the most common words in any language (like articles, prepositions, pronouns, conjunctions, etc) and does not add much information to the text.

Word Frequency:

Term frequency is commonly used in Text Mining, Machine Learning, and Information Retrieval tasks. As documents can have different lengths, it's possible that a term would appear more frequently in longer documents versus shorter ones



Tokenization :

Tokenization is the process of tokenizing or splitting a string, text into a list of tokens. One can think of token as parts like a word is a token in a sentence, and a sentence is a token in a paragraph.

Word Tokenization:

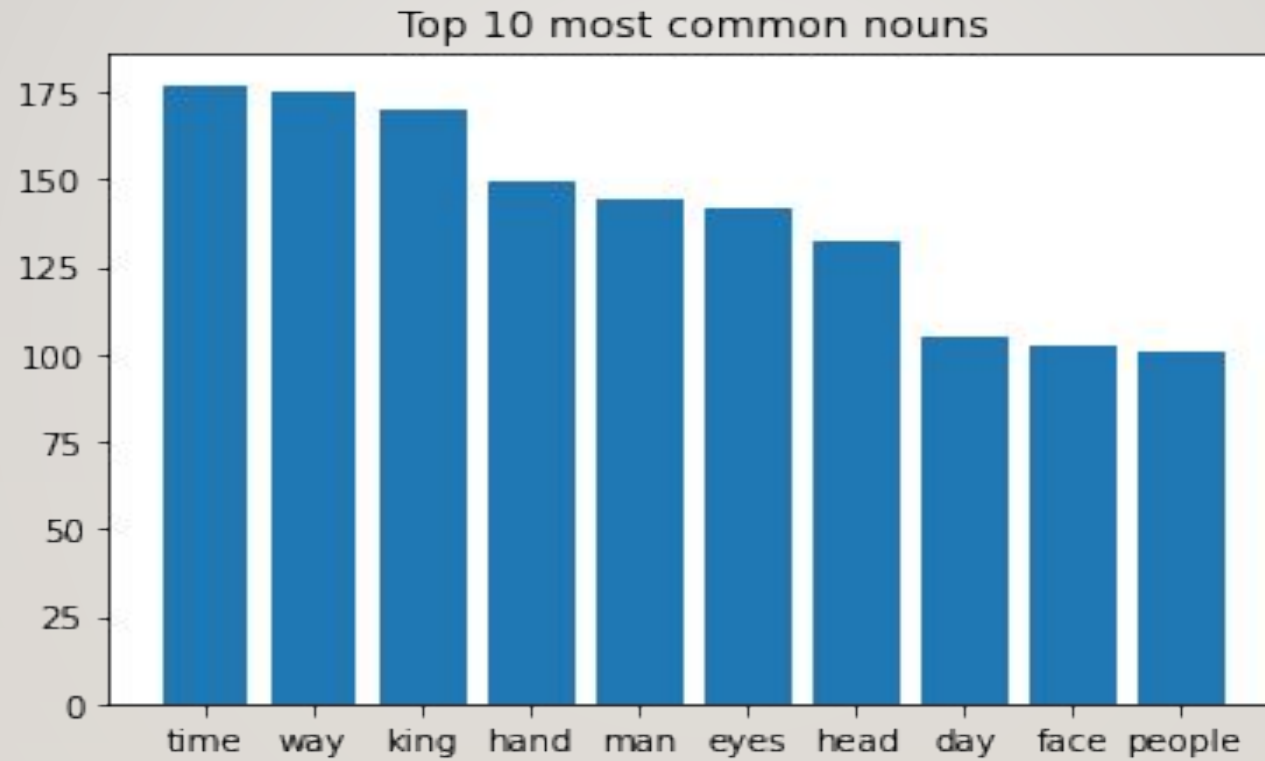
Tokenization is essentially splitting a phrase, sentence, paragraph, or an entire text document into smaller units, such as individual words or terms. Each of these smaller units are called tokens. Check out the below image to visualize this definition: The tokens could be words, numbers or punctuation marks.

Sentence Tokenization:

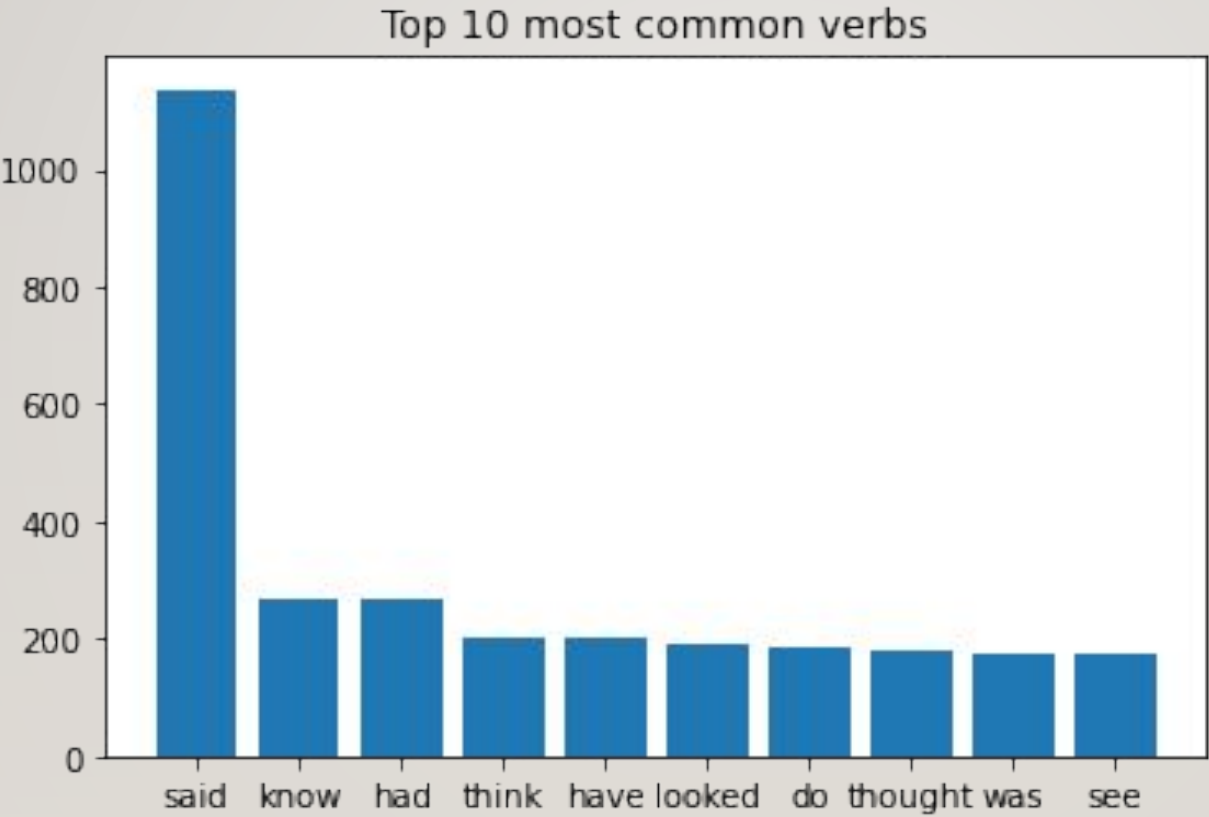
Sentence tokenization is the process of splitting text into individual sentences. For literature, journalism, and formal documents the tokenization algorithms built in to spacy perform well, since the tokenizer is trained on a corpus of formal English text.



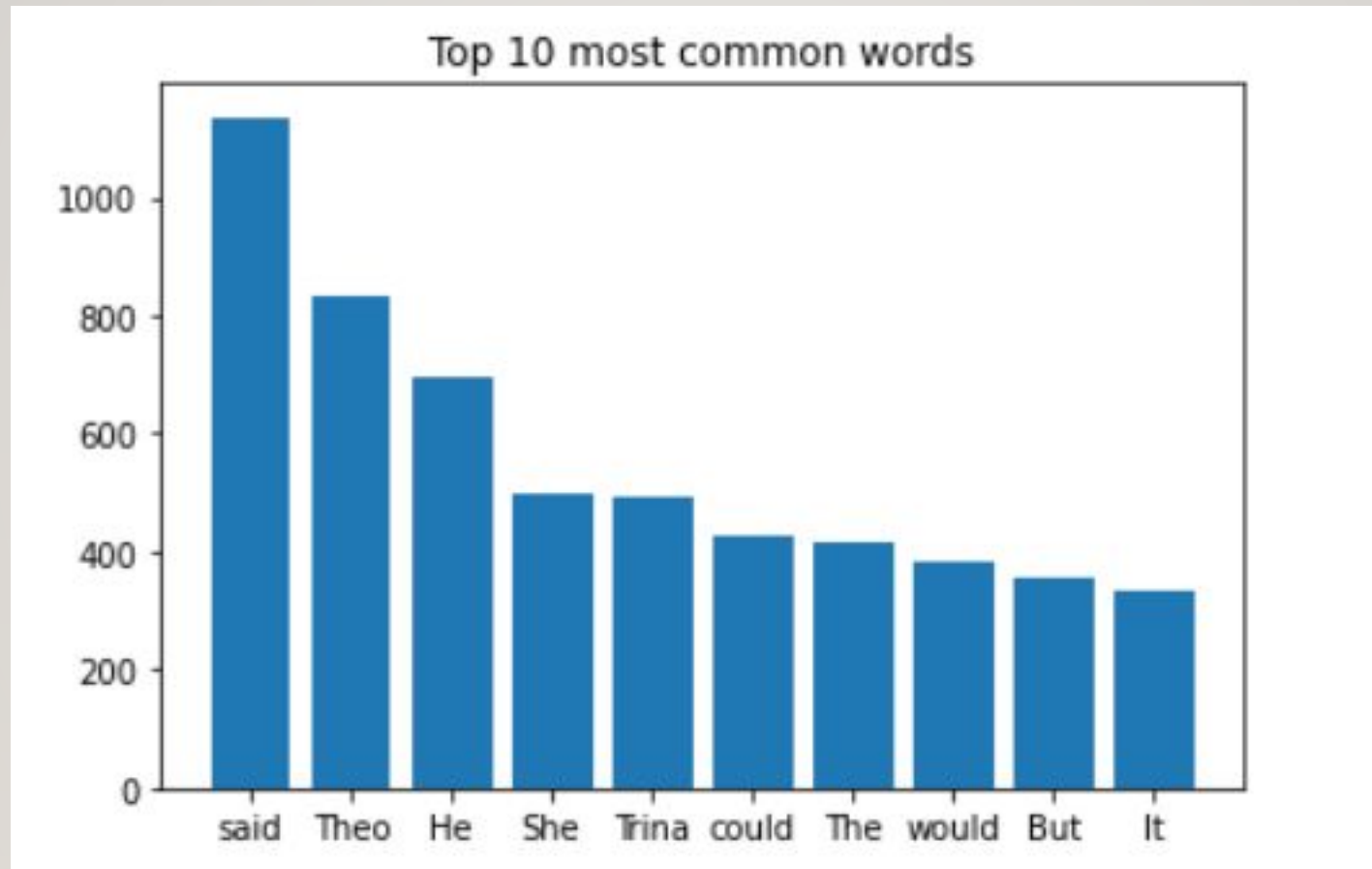
Top 10 Most Common Nouns:



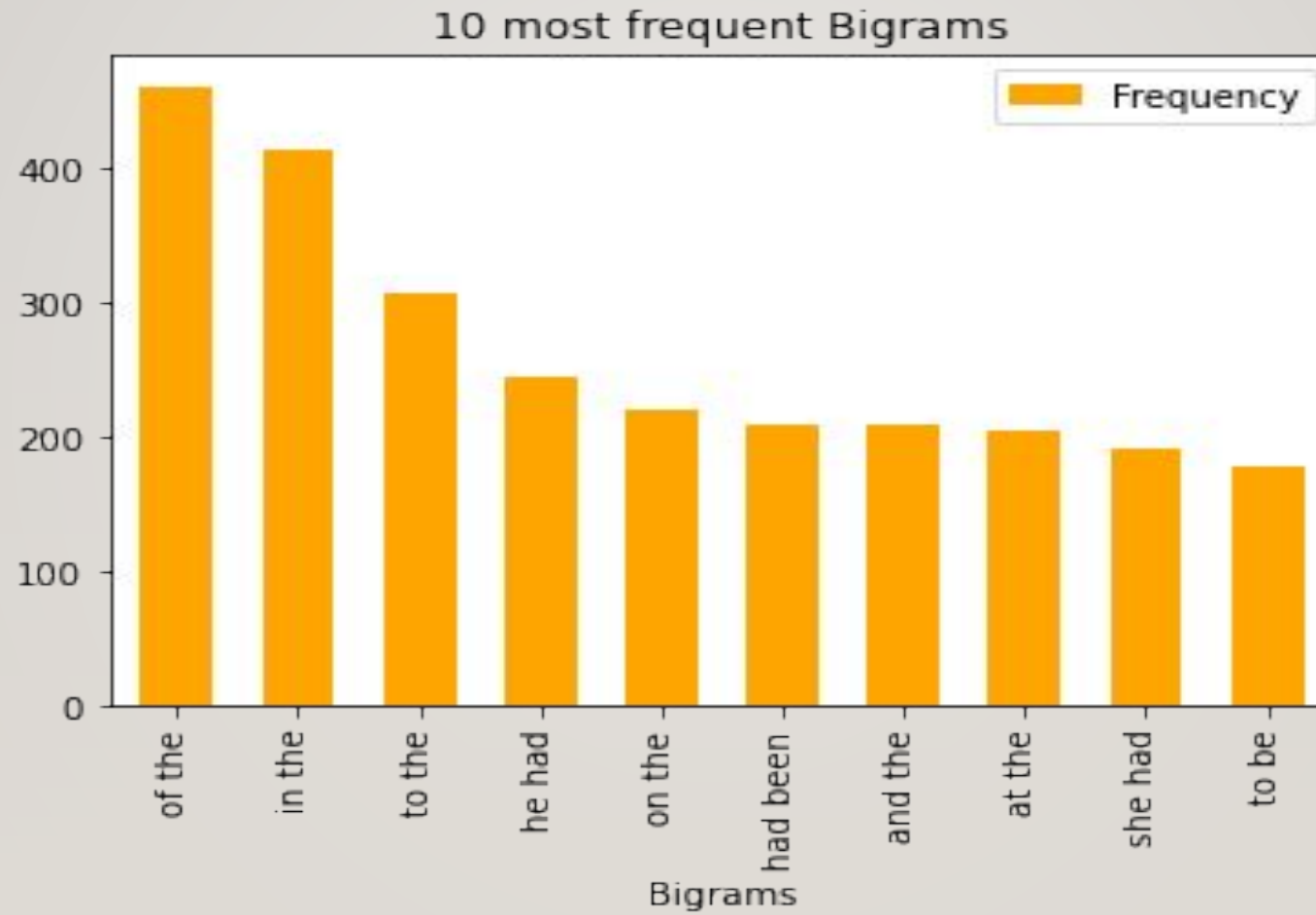
Top 10 Most Common Verbs:



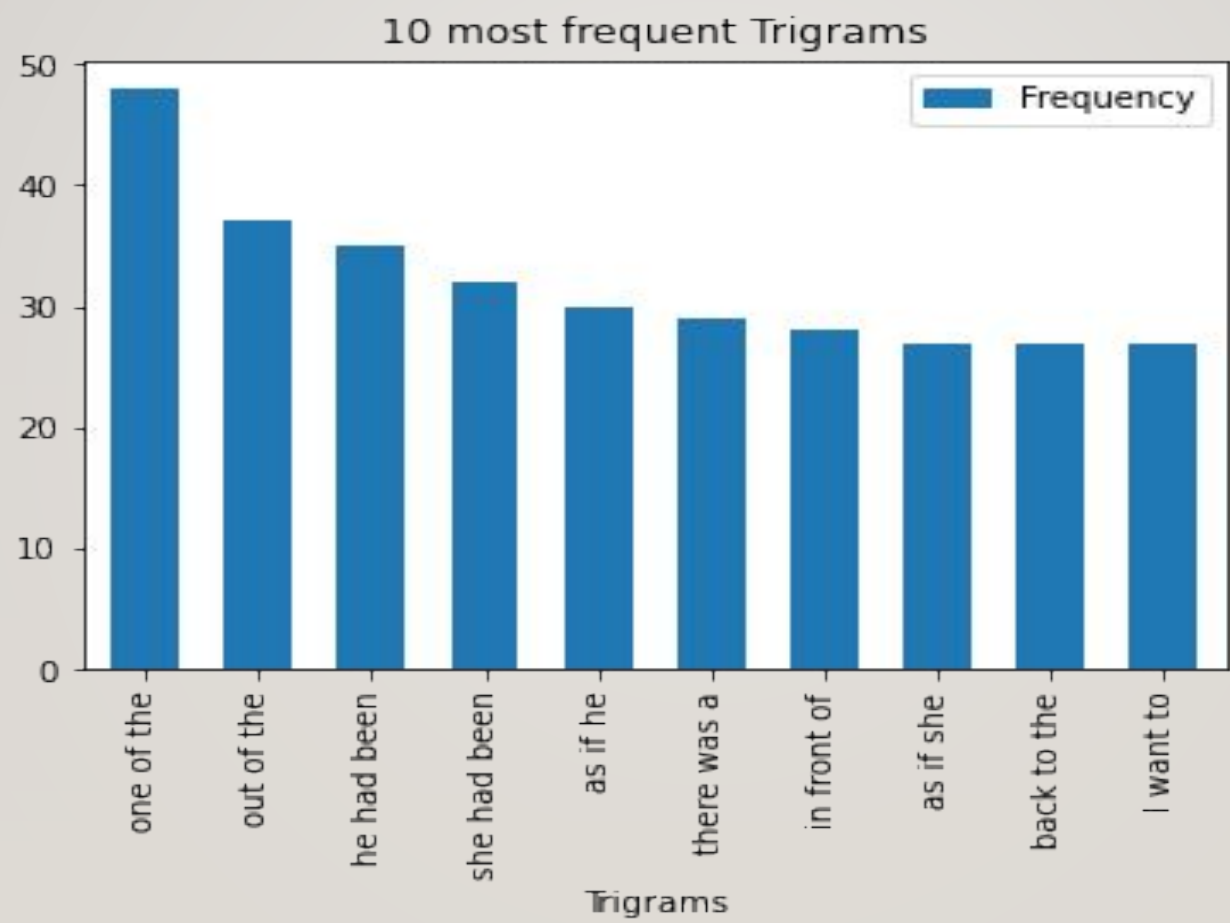
Top 10 Most Common Words:



10 Most Frequent Bigrams:

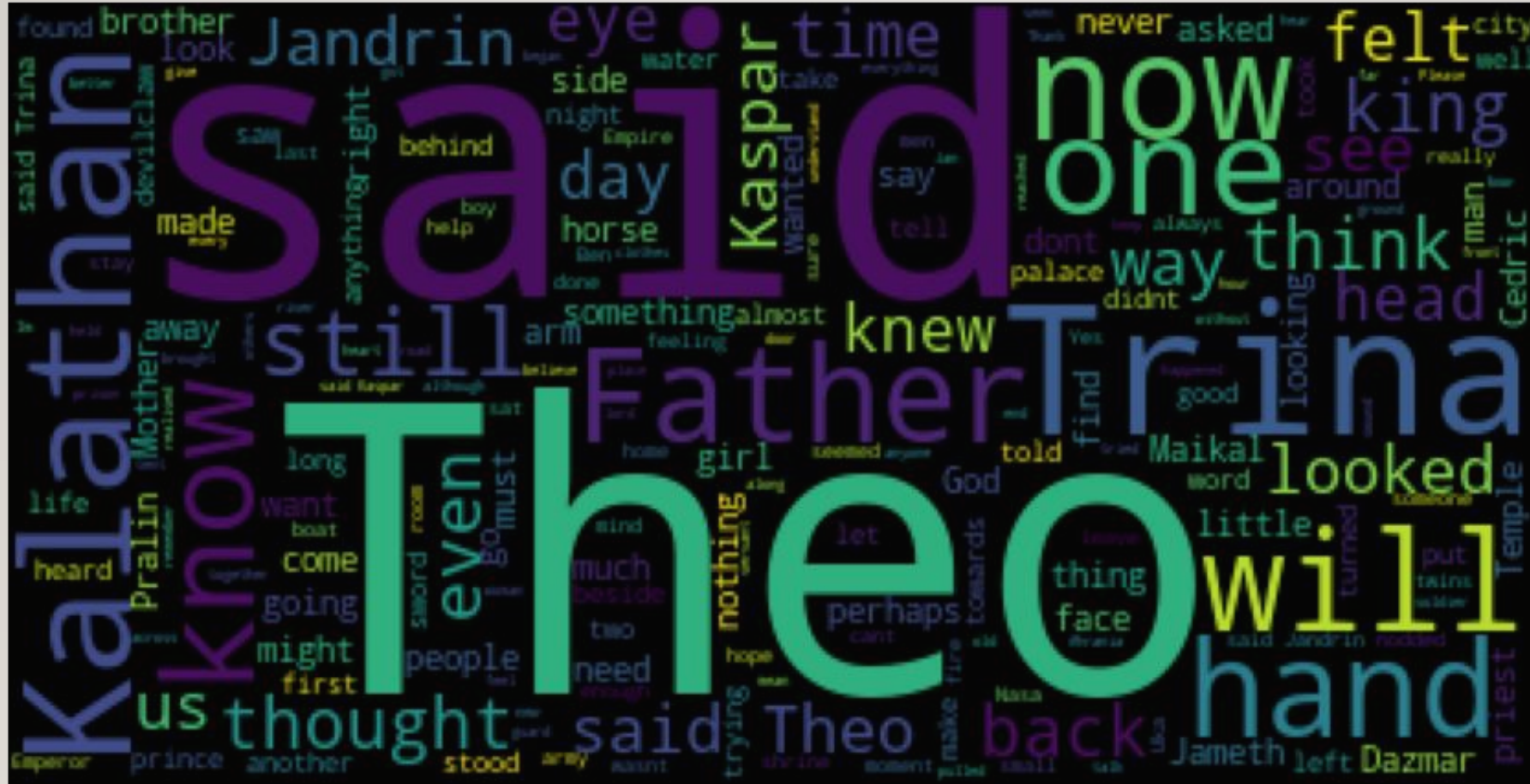


10 Most Frequent Trigrams:



Word Cloud:

Words frequented in text



Sentiment Analysis

Vader:

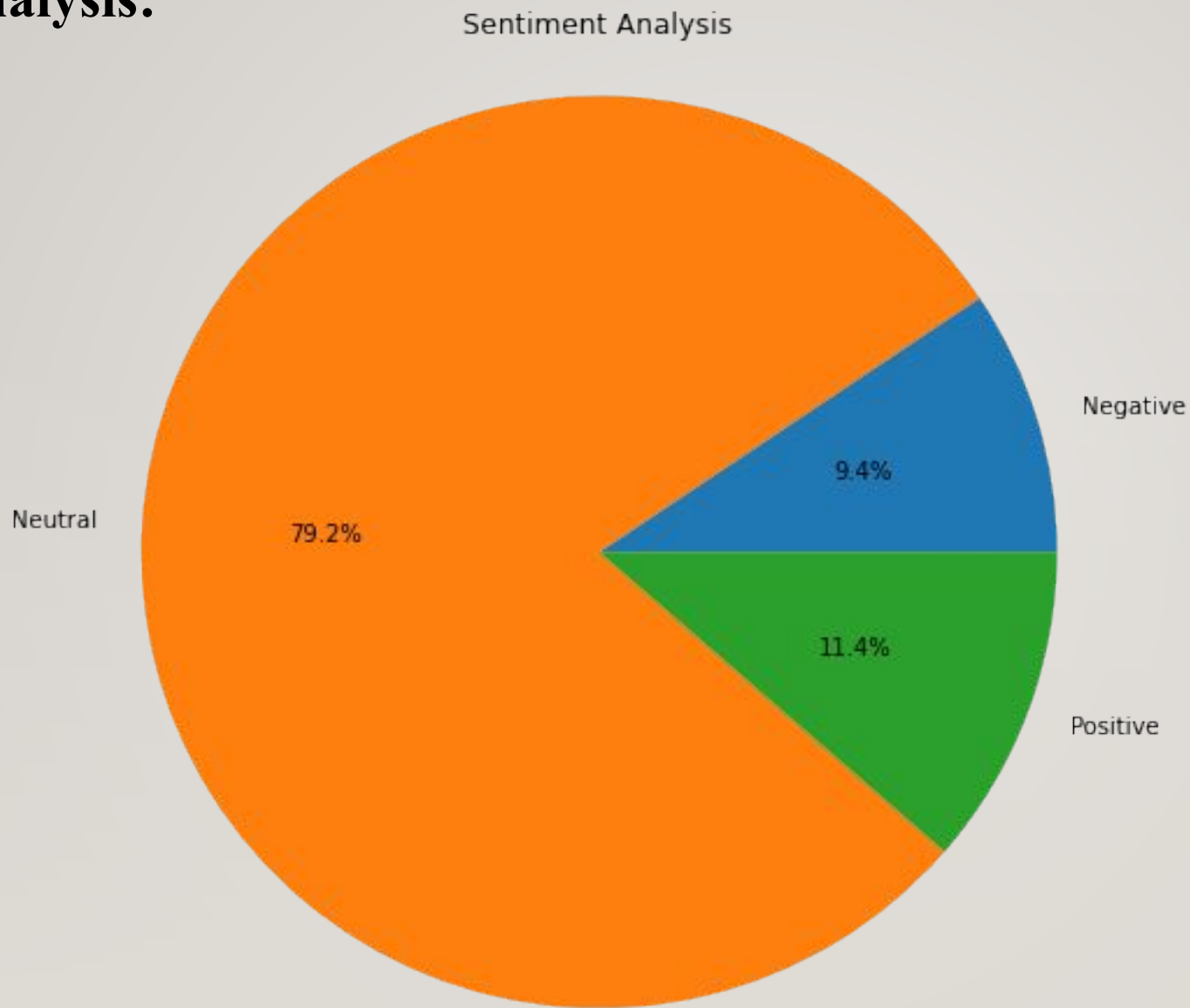
VADER (Valence Aware Dictionary for Sentiment Reasoning) is a model used for text sentiment analysis that is sensitive to both polarity (positive/negative) and intensity (strength) of emotion. It is available in the NLTK package and can be applied directly to unlabeled text data.

Polarity:

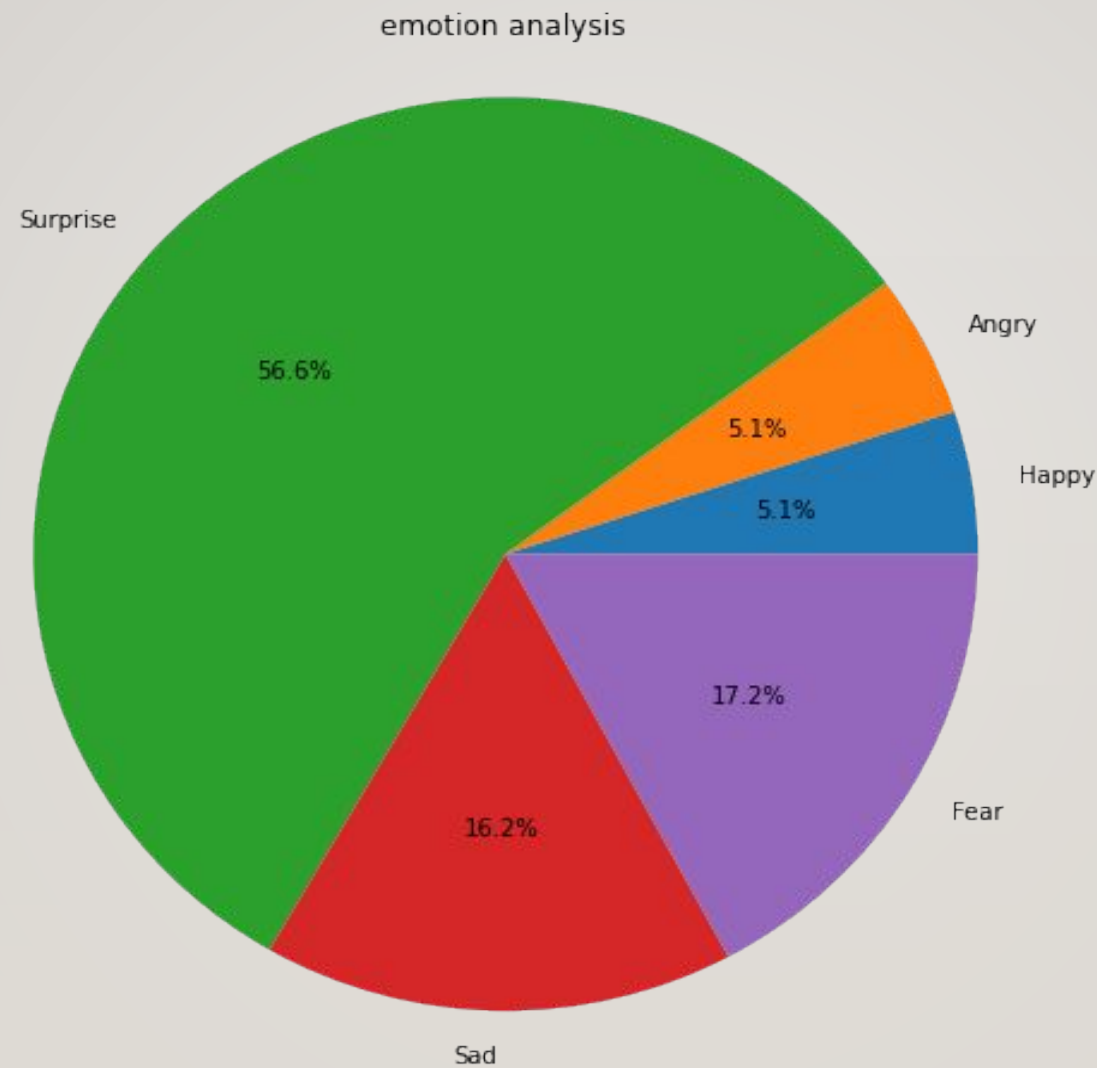
Polarity analysis takes into account the amount of positive or negative terms that appear in a given sentence. It is useful to some extent, since it does a good job of structuring data sets



Sentiment Analysis:

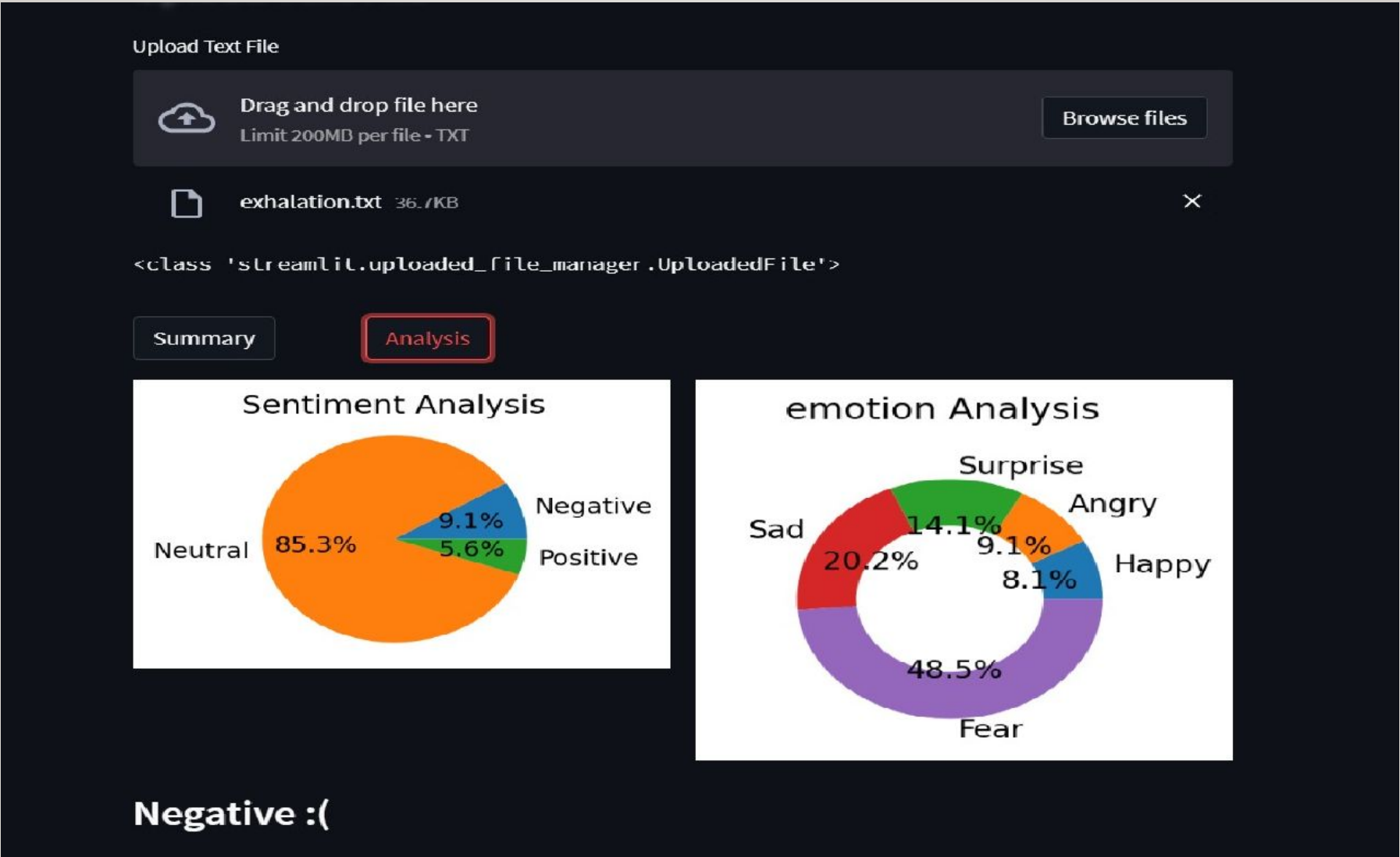


Emotion Analysis:



Model Deployment

Using Stream lit:



Upload Text File



Drag and drop file here

Limit 200MB per file • TXT

Browse files



exhalation.txt 36.7KB



```
<class 'streamlit.uploaded_file_manager.UploadedFile'>
```

Summary

Analysis

The amount of air that I draw from each day's new pair of lungs is exactly as much as seeps out through the joints of my limbs and the seams of my casing, exactly as much as I am adding to the atmosphere around me; all I am doing is converting air at high pressure to air at low. While this perhaps does not constitute air sharing in the strictest sense, there is camaraderie derived from the awareness that all our air comes from the same source, for the dispensers are but the exposed terminals of pipes extending from the reservoir of air deep underground, the great lung of the world, the source of all our nourishment. Air can neither be created nor destroyed; the total amount of air in the universe remains constant, and if air were all that we needed to live, we would never die. It is possible that this hypothetical universe has the same or higher air pressure as ours, but suppose that it had a much lower air pressure than ours, perhaps even a true vacuum? When the pressure everywhere in the universe is the same, all air will be motionless, and useless; one day we will be



THE LEADER WHO HAD NO TITLE.pdf 1.0MB

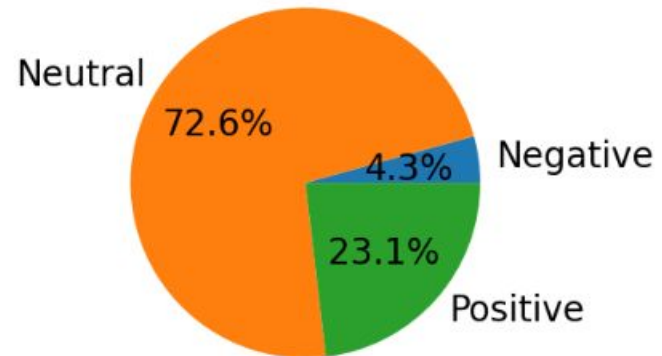


<class 'streamlit.uploaded_file_manager.UploadedFile'>

Summary

Analysis

Sentiment Analysis



Emotion Analysis



Positive :)



Drag and drop file here

Limit 200MB per file • PDF

Browse files



THE LEADER WHO HAD NO TITLE.pdf 1.0MB



```
<class 'streamlit.uploaded_file_manager.UploadedFile'>
```

Summary

Analysis

“Victims fall in love with excuses like, ‘I’m not the owner or the manager here, so I can’t do much to improve things,’ or ‘I’m not smart enough to lead change in the face of challenging circumstances,’ or ‘I don’t have enough time to do all the things I know I could do to make our company great,’ or ‘I’ve tried to be the best that I can be, but none of these ideas have ever worked.’ It was not only a way that any person alive today, within any organization—a business, a community, an entire nation — could use their natural power to inspire others to become all they were meant to be, it was also a way for each of us to realize our best potential and create value for the world around us. Employment was just a vehicle to pass my time instead of an excellent chance to shine a light on other people and a way to use my days to build a better organization—and in so doing—a better world. By the way, did you know that if by your shining example of personal leadership you inspire ten people each day to play at their absolute best, by the end of only four weeks you’ve positively

CHALLENGES FACED

- Extracting text from PDF.
- Long waiting period for detecting new and changed data.
- Installations of various versions of software for executing.
- Finding Keywords.
- More processing is required
- It is difficult to find out if the data is relevant.
- Searching for info and indexing is a challenge

**THANK
YOU**