About the project

The project is about creating a text summarizer application that can also generate the nature of text and summary. The application has been deployed using streamlit. There are 2 commonly used methods in text summarization – abstractive and extractive summarization. The technique used is extractive text summarization, where important sentences are identified and extracted from the given text input. Extractive summarization is like a highlighter, while abstractive summarization is like a pen. Abstractive summarization is commonly used in convoluted and unstructured conversations. Since the text input being used here is structured, extractive text summarization is an appropriate method.

Approach taken to implement extractive text summarization

* Create the frequency table for every word in the sentence.
* Create a sentence value dictionary to store the score of every sentence as the sum of frequencies of every word present in it.
* Initialize a variable to store the sum of frequencies of all the sentences.
* Find the average score of all the sentences in the text input
* If a particular sentence has a score greater than 1.2 times the average score of the text input, it is added to the summary.

Sentiment analysis

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon (word) and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. VADER uses a combination of a sentiment lexicon is a list of lexical features (e.g., words) which are generally labeled according to their semantic orientation as either positive or negative. VADER not only talks about the Positivity and Negativity score but also tells us about how positive or negative a sentiment is. The Compound score is a metric that calculates the sum of all the lexicon ratings which have been normalized between -1(most extreme negative) and +1 (most extreme positive).