**Project Report**

**Sentiment Analysis using Reddit Api**

Mustafa Faysal – 25231

Moeen Haider – 24505

**Objective**

The objective of a sentiment analysis report based on submissions/comments from Reddit API is to gain insights into the overall sentiment of the Reddit community towards a particular topic or set of topics. This can be useful for businesses or organizations to understand how their brand or products are being perceived by the Reddit community, or for researchers to study public opinion on a variety of topics.

**Methodology**

To achieve the objective, we followed the following methodology:

*Identification of Relevant Subreddits:* We identified popular or most used subreddits which could be relevant to the user and added them to the dropdown list in the GUI.

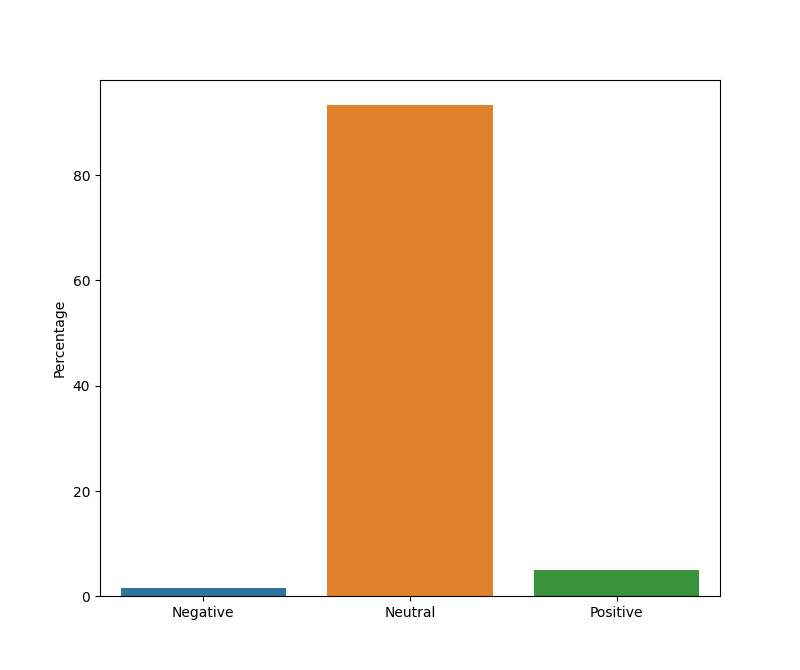
*Retrieval of Comments using PRAW:* We used PRAW, the Python Reddit API Wrapper, to retrieve submissions/comments from the relevant subreddits based on user input. We did not limit the number of submissions received so that we can get an insight of all the subreddit ‘HOT’ submissions.

*Sentiment Analysis using Vader Model:* We used the Vader (Valence Aware Dictionary for Sentiment Reasoning) model, a rule-based sentiment analysis tool, to perform sentiment analysis on the comments/submissions. The Vader model assigns a sentiment score to each comment based on the presence of positive, negative, and neutral words in the text.

*Visualization of Results:* We visualized the results of the sentiment analysis using a bar graph to provide an easy-to-understand summary of the sentiment of the retrieved comments/submissions.

**Limitations**

1. *Lack of Context:* The Vader model is limited in its ability to interpret the context in which words are used. For example, it does not recognize sarcasm or irony in a statement, which might not result in the most accurate analysis.
2. *Limited to English:* The Vader model is trained on English language text data, so it may not be as accurate when analyzing text in other languages.
3. *Limited to Social Media:* The Vader model was specifically designed for sentiment analysis of social media posts, such as tweets and Facebook updates. It may not be as effective when analyzing sentiment in news articles or academic papers.
4. *Dependence on Lexicons:* The Vader model relies heavily on lexicons to identify sentiment, which may not capture the full range of emotions or sentiment expressed in text.
5. *Over-reliance on Intensity:* The Vader model places a heavy emphasis on the intensity of sentimental words, which may not always accurately reflect the overall sentiment.

**Results :**

**Future Direction :**

In future we have streamlined a set of goals, in which we will be using multiple models such as the Vader. Which will give the user a wide variety of datasets for them to understand and analyze. Furthermore twitter and facebook Api can also be embedded to get data from different social platforms.