SENTIMENTAL ANALYSIS CODE AND INSIGHTS ON THE PLOTS

The code is designed to perform sentiment analysis on the Reuters C50 text corpus, where each document is associated with a specific author. It begins by loading necessary libraries for text mining, sentiment analysis, and data visualization. The documents are read from the corpus, and each text is linked to its author. The texts are then preprocessed by converting them to lowercase, removing punctuation, numbers, whitespace, and common stop words. Next, the code computes sentiment scores for various emotions (like joy, trust, anger, etc.) for each document using the NRC lexicon. These sentiment scores are then aggregated by author to calculate the average positive and negative sentiments. The code visualizes these results by creating bar charts that display the net sentiment (positive minus negative) for each author. Finally, it also calculates and visualizes the average sentiment scores across all texts in the corpus, showing how different sentiment categories (like anger, joy, etc.) are distributed among the authors.

A screenshot of a graph

Description automatically generated

Net Sentiment by Author:

The bar chart shows the positive and negative sentiment (positive sentiment – net sentiment) for each of the author in the Reuters C50 corpus. Authors like Therese Poletti, Pierre Tran, and Jim Gilchrist have high positive net sentiments, showing that their articles or documents mostly tend to have positive emotions. On the other side, authors like Karl Penhaul, and Benjamin KangLim portray negative sentiments, suggesting their articles and writings include negative emotions. Most of the authors have a positive net sentiment, indicating that overall, the corpus leans slightly towards positive emotions.

A chart with green and yellow bars

Description automatically generated

Same thing as above.

The Ranking helps show the emotional tone differences among authors in the Reuters corpus, potentially useful for understanding their writing style or target audience.

A graph with different colored bars

Description automatically generated

Trust is the most dominant sentiment across the Reuters corpus, suggesting that the content often evokes reliability or confidence.

Anticipation and Fear are also prominent, indicating that the texts might frequently discuss upcoming events or concerns.

Negative sentiments like Anger, Sadness, and Disgust are present but less dominant, suggesting that while the corpus contains some negative content, it is not overwhelming.

The sentiment analysis reveals an extensive range of emotional tones among the different authors in the Reuters C50 corpus. Some authors largely convey positive emotions, which can reflect optimistic or encouraging topics, while others have a more negative tone, maybe dealing with more critical or challenging subjects. The frequency of **Trust** and **Anticipation** across the corpus suggests that the texts could aim to build confidence and highlight certain prospects, which could be typical of news reporting styles. To Understand these sentiment distributions could provide deeper insights into the nature of the content and the perspectives of different authors within this corpus.