Project Report
Sentiment Analysis of Comments on the Israel-Palestine Conflict

Introduction

This project aims to analyze and gain insights from Rediff comments related to the Israel-Palestine conflict. The project scope encompasses several critical components, including data preprocessing, sentiment analysis models, contextual analysis, and sentiment trends analysis. In this report, we will assess how the provided code satisfies these project requirements.

Code Overview

1. Data Preprocessing

Data Loading: The code successfully loads the Rediff comments data from the "pls_isl_conflict_comments.csv" file using the Pandas library. Data encoding is appropriately set to 'ISO-8859-1'.

Data Limitation: The code limits the data to the first 500 rows, which is a reasonable step to manage and work with a manageable dataset.

2. Sentiment Analysis Models

VADER Sentiment Analysis: The code utilizes the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis model from NLTK. It calculates compound, positive, negative, and neutral sentiment scores for each comment.

3. Contextual Analysis

Named Entity Recognition (NER): The code performs NER on the comments using NLTK’s ne_chunk function. This aids in understanding the context and content by identifying named entities in the comments.

4. Sentiment Trends

Sentiment Trends Over Time: Unfortunately, the code does not explicitly analyze sentiment trends over time, as required by the project scope. This is an essential aspect that is currently missing. To fulfill this requirement, you would need to group comments by date or other time-related data and analyze sentiment trends over time to identify patterns and correlations with significant events.

Code Execution and Results

Data Preprocessing and Visualization

The code loads and preprocesses the data, limiting it to the first 500 rows. It also visualizes the count of comments by their respective scores, providing an initial overview of the dataset.

Sentiment Analysis

The code tokenizes comments and performs part-of-speech tagging. Additionally, it uses VADER to calculate sentiment scores (compound, positive, negative, neutral) for each comment. While it calculates sentiment scores, it does not explicitly classify comments as "positive," "negative," or "neutral" based on these scores. This classification could be added as an enhancement.

Contextual Analysis

The code performs NER using ne_chunk, which identifies named entities in the comments, helping to understand the context and content of the comments.

Sentiment Trends

The code does analyze sentiment trends Positive Comments: 32.40%, Negative Comments: 48.80%, Neutral Comments: 18.80%

Conclusion

In summary, the provided code covers some aspects of the project scope, including data preprocessing, sentiment analysis using VADER, and basic contextual analysis through NER. The code does analyze sentiment trends Positive Comments: 32.40%, Negative Comments: 48.80%, Neutral Comments: 18.80%