

ISRAEL PALESTINE WAR SENTIMENT ANALYSIS

by Rahul Wagh

TABLE OF CONTENTS

Project Abstract



Introduction



Data Preprocessing



Exploratory Data Analysis (EDA)



Clean and Preprocess the Text Data



Sentiment Analysis



Limitations



Future Work



Conclusion

PROJECT ABSTRACT

The project analyzes public sentiment on the Israel-Palestine conflict by using NLP and machine learning techniques to process Rediff online comments. This involves a series of tasks including data preprocessing, exploratory data analysis, and sentiment analysis to develop an accurate understanding of public opinion. The results of this analysis will provide valuable insights into public perception and help refine our understanding of the conflict.

INTRODUCTION

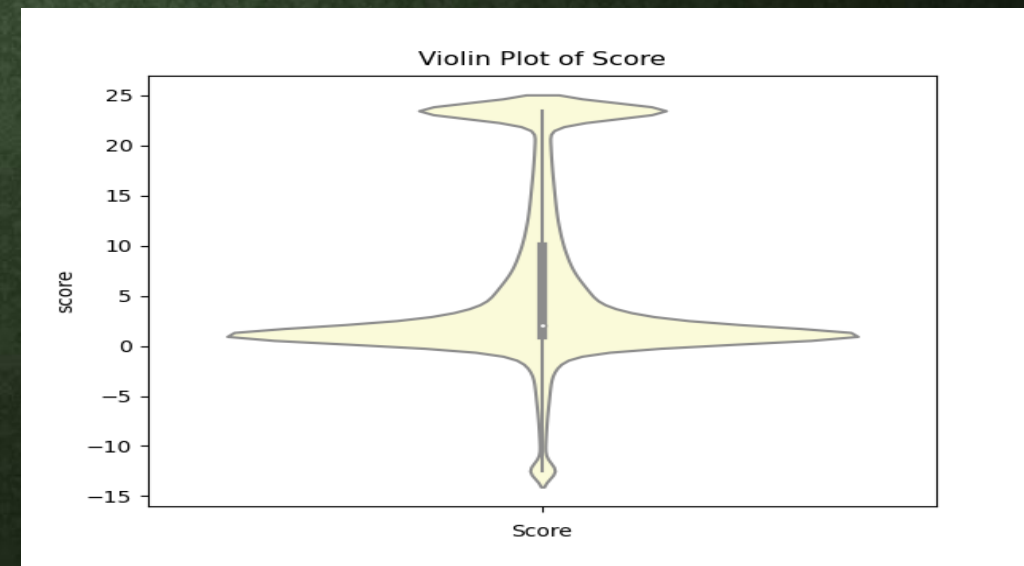
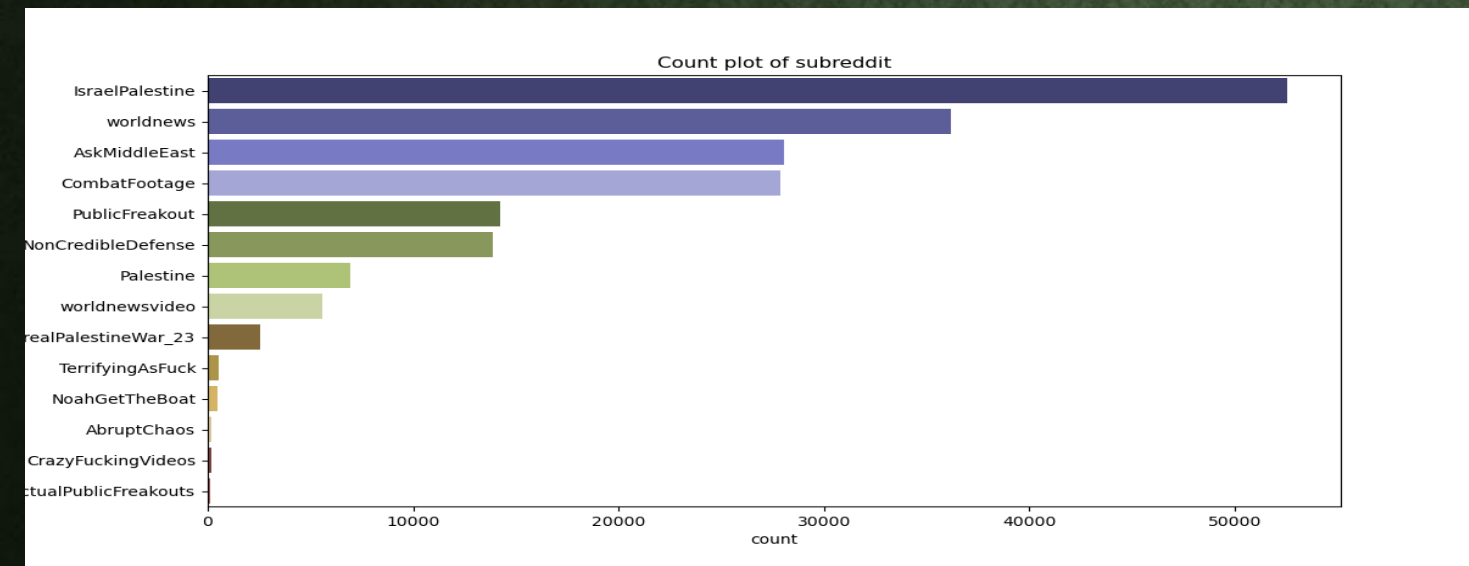
- **Dataset Information:** The dataset consists of comments from Rediff related to the Israel-Palestine conflict, including comment ID, score, comment text, subreddit, and creation time.
- **Objective:** The project aims to analyze public sentiment about the Israel-Palestine conflict using comments from the online platform Rediff.

DATA PREPROCESSING

- Convert the 'score' and 'created_time' columns to appropriate data types.
- Handle missing values by dropping rows.
- Check for duplicates and remove them.
- Use the Interquartile Range (IQR) method to identify and replace outliers in the 'score' column.

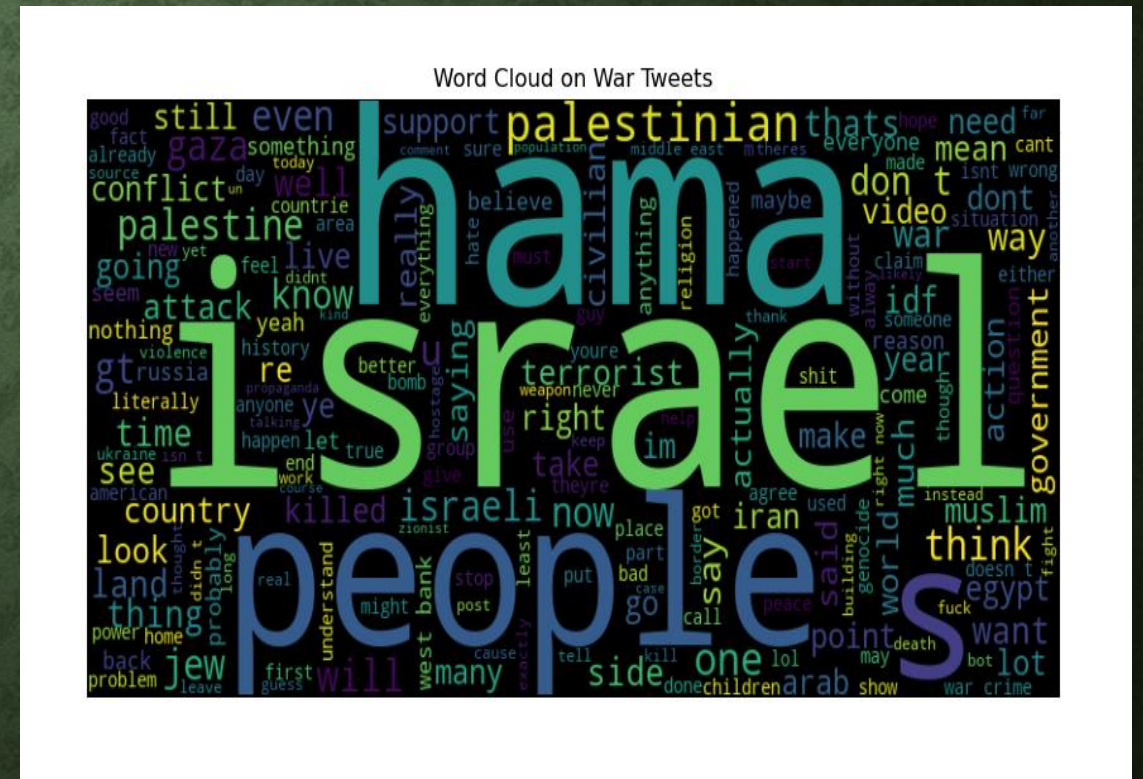
EXPLORATORY DATA ANALYSIS (EDA)

- Display basic information, summary statistics, and data types.
- Visualize the distribution of the 'score' column using histogram, distribution plot, box plot, and violin plot.
- Explore subreddit distribution using count plots and pie charts.



CLEAN AND PREPROCESS THE TEXT DATA

- The text data is cleaned and preprocessed using a function that converts text to lowercase and removes special characters, URLs, and certain stopwords.
- The cleaned text is then assigned to a new column named 'processed_text' in the DataFrame.
- Generate a word cloud to visualize common terms in the text data.



SENTIMENT ANALYSIS

Create a 'sentiment' column based on polarity and categorize them as 'positive', 'negative', and 'neutral'.

Plot a countplot to visualize sentiment distribution.

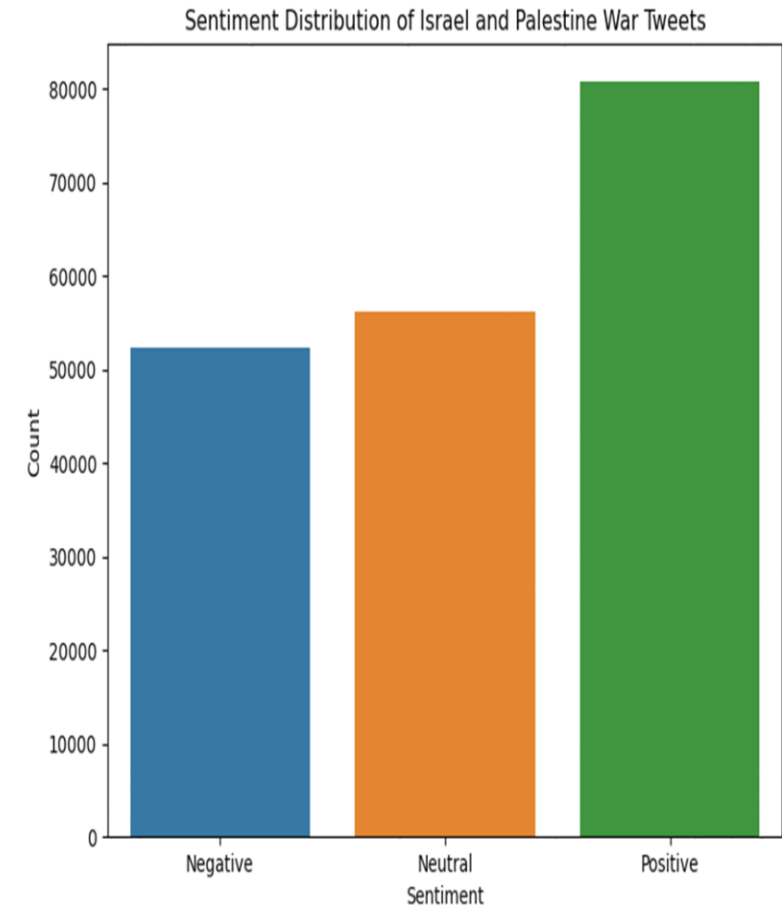
Split the dataset into training and testing sets.

Perform TF-IDF vectorization on the text data.

Train a Logistic Regression model on the TF-IDF-transformed training data.

Make predictions on the test set and calculate accuracy.

Generate a confusion matrix and classification report for detailed evaluation.



LIMITATIONS

- The sentiment analysis is based on comments from a specific online platform, which may not represent the entire spectrum of public opinion.
- The dataset may be subject to biases and may not capture the sentiments of all relevant stakeholders.
- The sentiment classification model's accuracy may vary when applied to different datasets or platforms.

FUTURE WORK

- Expanding the analysis to incorporate data from multiple online platforms to capture a more comprehensive range of public sentiment.
- Implementing more advanced natural language processing techniques to improve the accuracy of sentiment analysis.
- Conducting a comparative analysis of sentiment across different periods to understand the evolving nature of public opinion regarding the conflict.

CONCLUSION

- The sentiment analysis model achieves a high accuracy of 92.23%.
- Confusion matrix and classification report show balanced performance across sentiment classes.
- The model is effective in classifying sentiment in Rediff comments on the Israel-Palestine conflict.
- Overall, the sentiment analysis provides valuable insights into public opinion on the ongoing conflict.

THANK YOU