TWITTER SENTIMENT ANALYZER

Introduction

This project involves developing a sentiment analyzer for Twitter data, employing techniques from data processing, machine learning, and natural language processing.

Data Loading and Preprocessing

Data Loading: Tweets are loaded from a CSV file, including columns like TweetID, Entity, Sentiment, and TweetContent.

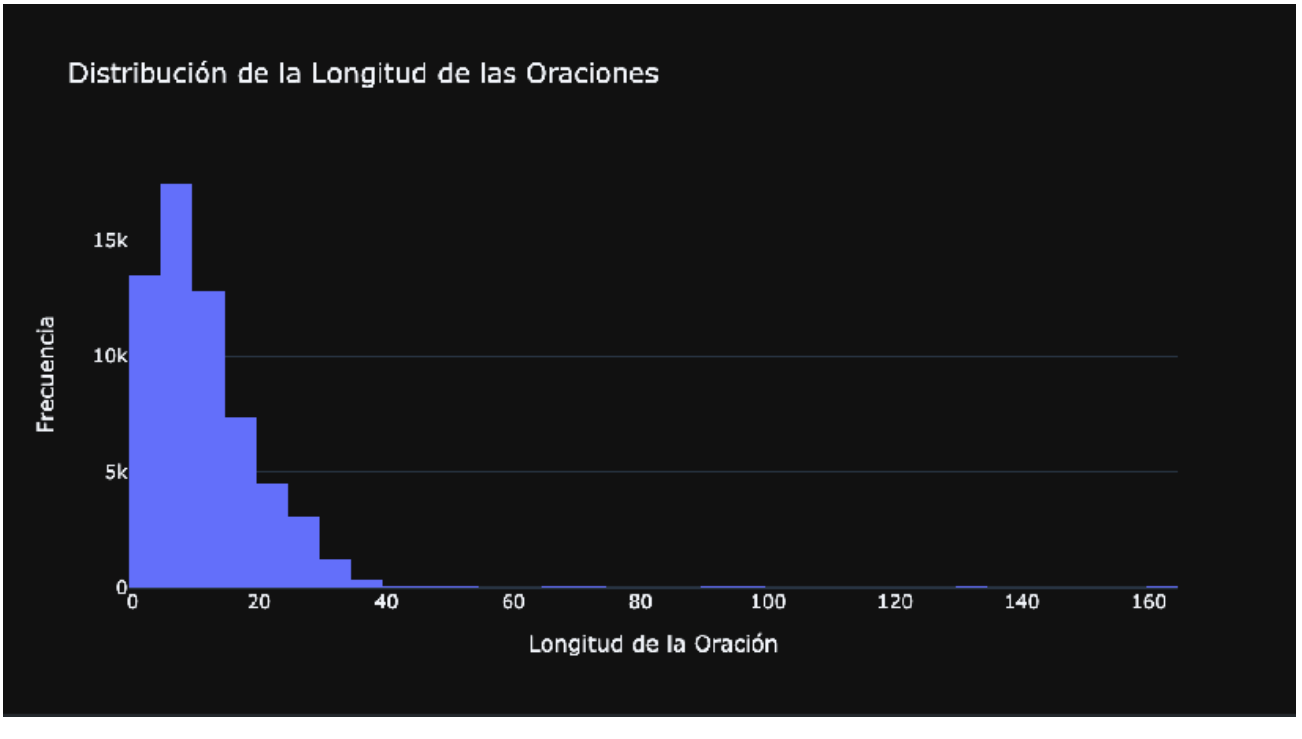
Data Preprocessing: Tweets are cleaned and normalized, removing URLs, special characters, and applying lemmatization.

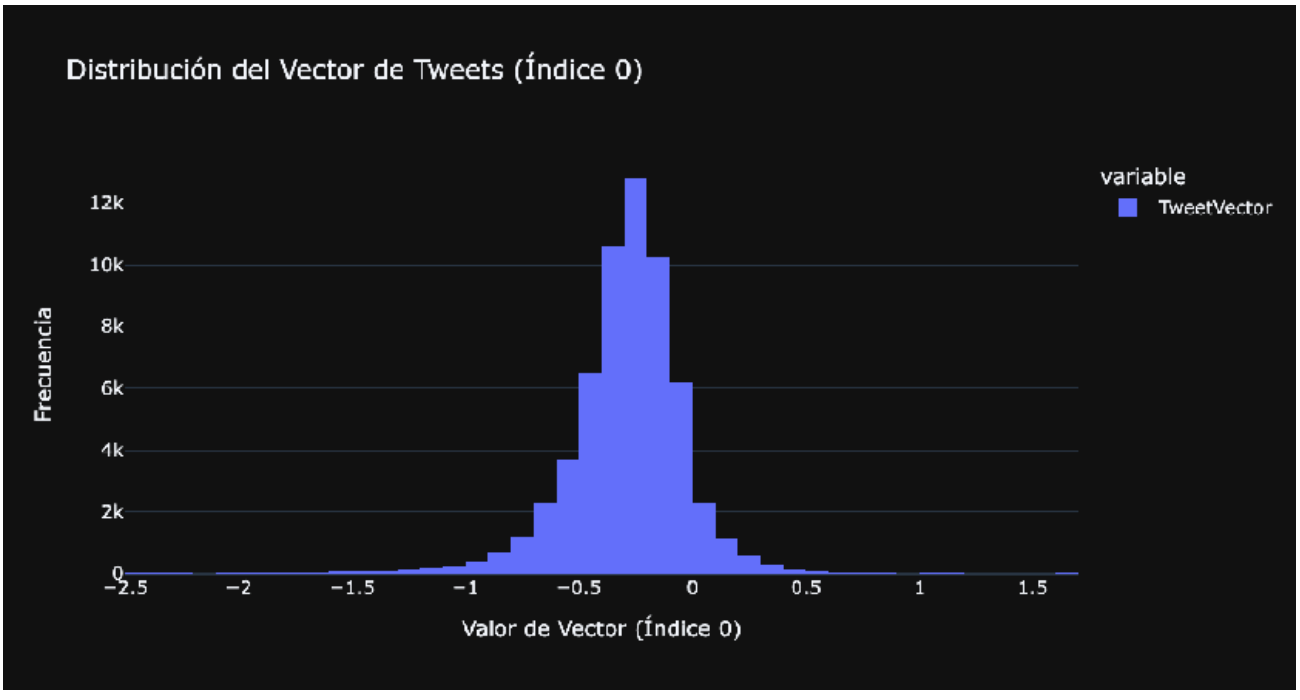
Feature Enhancement

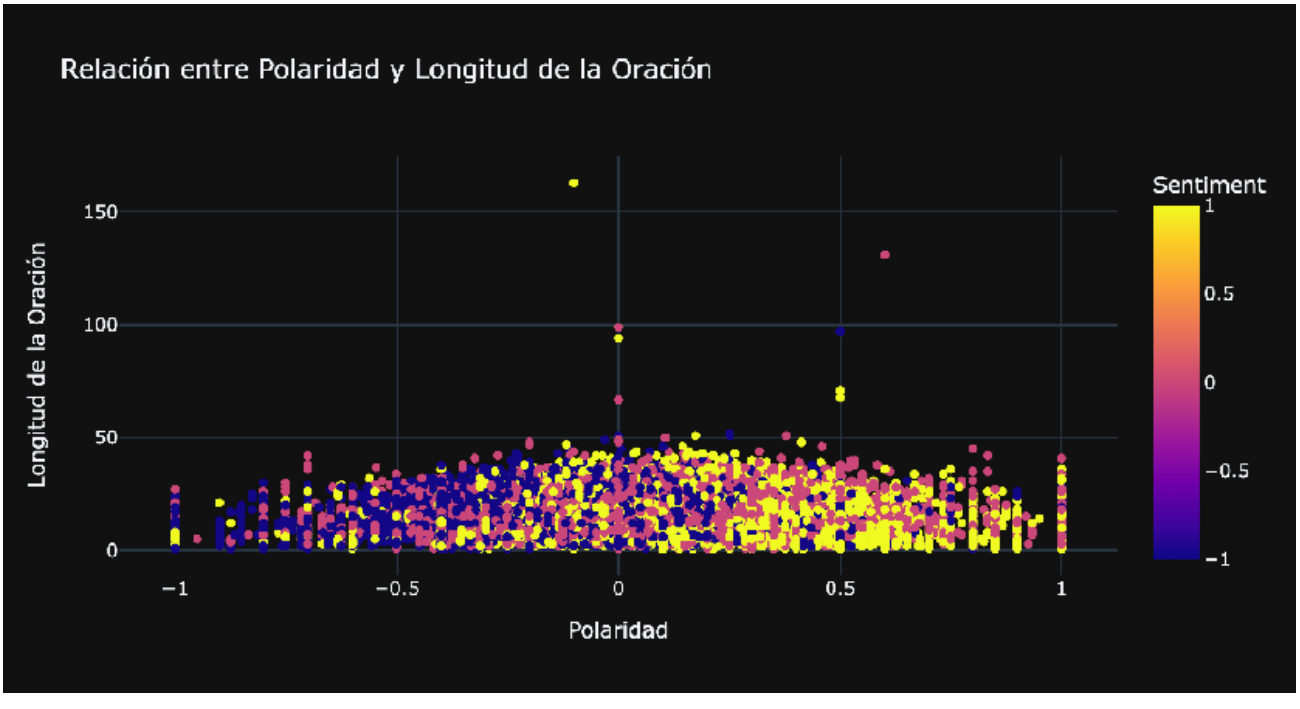
- **Word2Vec Training:** Tweets are used to train a Word2Vec model, generating vector representations.
- **Sentence Length and Polarity:** Each tweet's length and polarity are calculated.

Data Visualization

- Visualization includes histogram plots for sentence length and sentiment distribution, scatter plots for feature relationships, and word cloud generation.







Feature Selection and Data Splitting

- Selects features like Entity, TweetContent, and others for model training and testing.

Sentiment Analysis

- Utilizes cardiffnlp/twitter-roberta-base-sentiment model for analyzing sentiments of tweets.
- **Evaluation:** For example, a tweet "I hate you" is analyzed, resulting in a sentiment "Negative" with a confidence score of 96.54%. This demonstrates the model's ability to categorize sentiments with high confidence.

Conclusion

The sentiment analyzer successfully categorizes tweets, providing insights into public opinion trends on social media. This tool has applications in market analysis, political campaigns, and more.