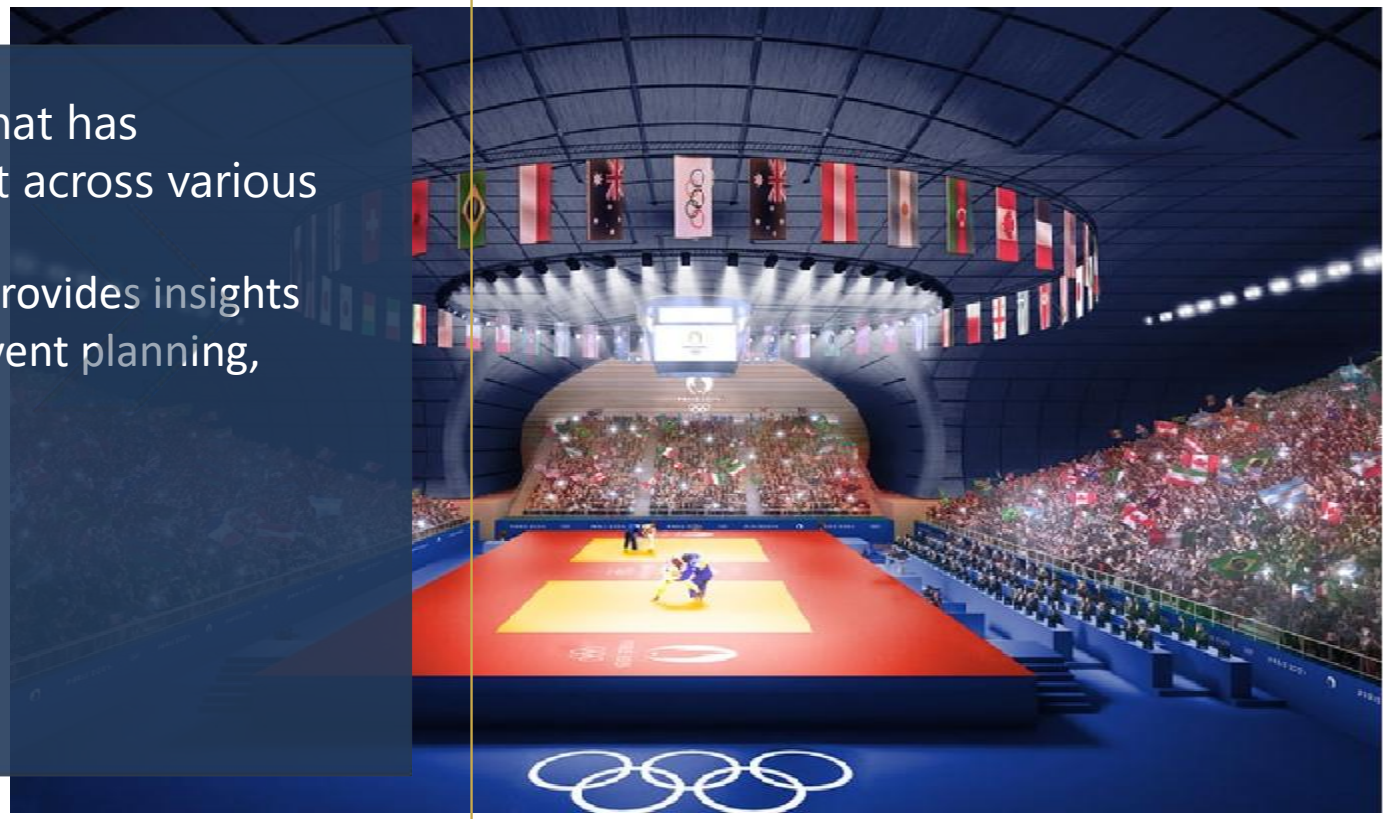
SENTIMENT ANALYSIS

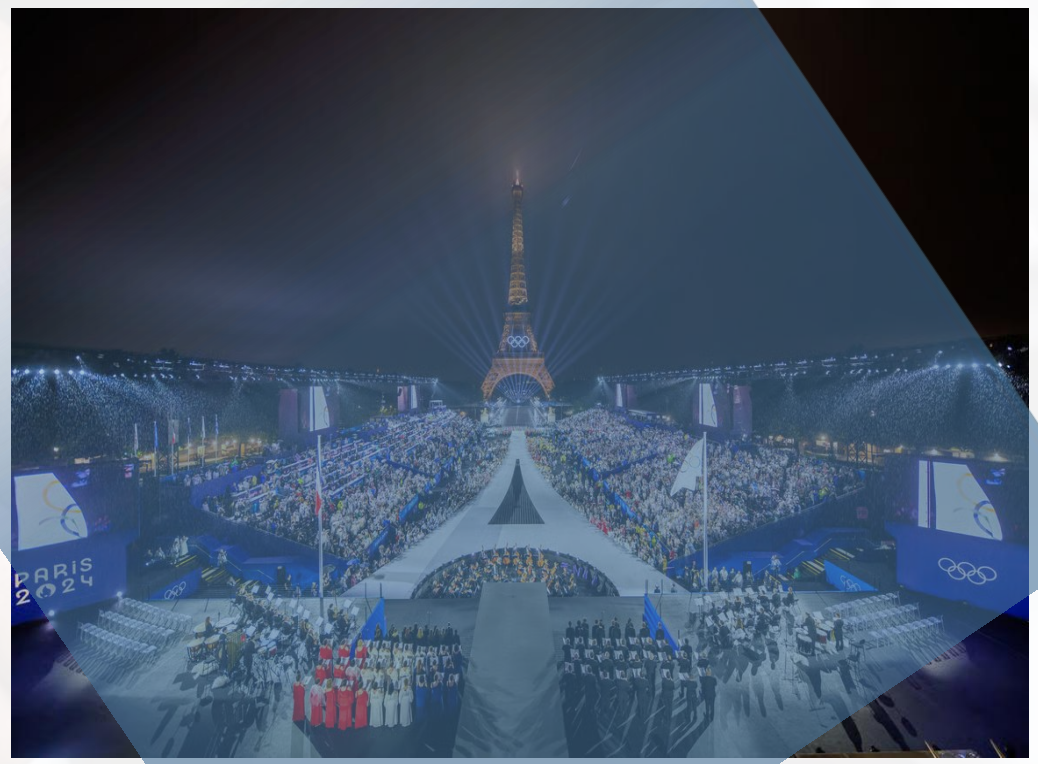
BUSINESS OVERVIEW

- The Paris Olympics is a global sporting event that has garnered significant attention and engagement across various social media platforms.
- Analysing public sentiment on the Paris Olympics provides insights for improving content, marketing strategies, and event planning, including health and sanitation.



PROBLEM STATEMENT

The Paris Olympics is a high-profile event that generates a substantial volume of unstructured social media data that reflects public sentiment. The challenge lies in effectively analyzing this vast and diverse stream of data while also tackling challenges such as language differences, sentiment variations and contextual meanings in order to provide accurate and actionable insights.



OBJECTIVES

Main Objective

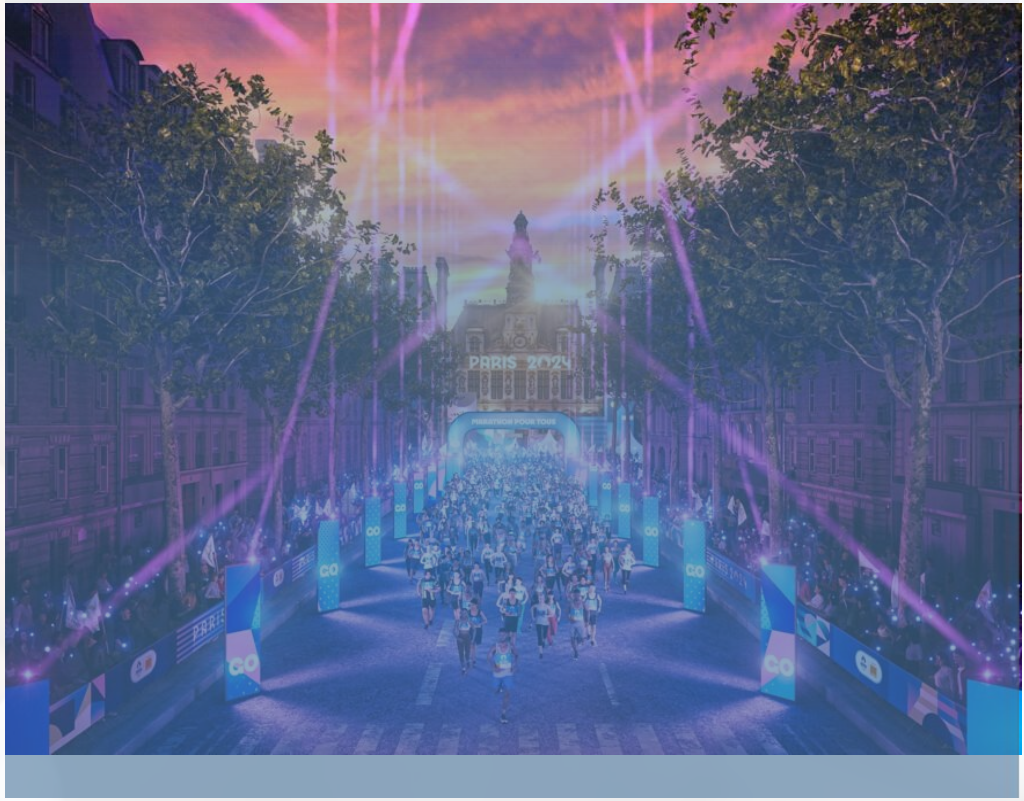
Develop a comprehensive social media sentiment analysis model that accurately captures and interprets public sentiment about the Paris Olympics from social media data.

Specific Objectives

To extract, preprocess and clean social media data from multiple platforms addressing quality issues and handling multilingual content related to the Paris Olympics.

To develop and train advanced natural language processing models to accurately classify sentiments incorporating techniques to handle sarcasm and contextual nuances.

To create interactive visualizations to display sentiment trends and key events providing actionable insights to stakeholders based on comprehensive analysis of public opinions.



DATA UNDERSTANDING

Data Sources

- Extracted data from X using Octorparse Webscraping Tool.

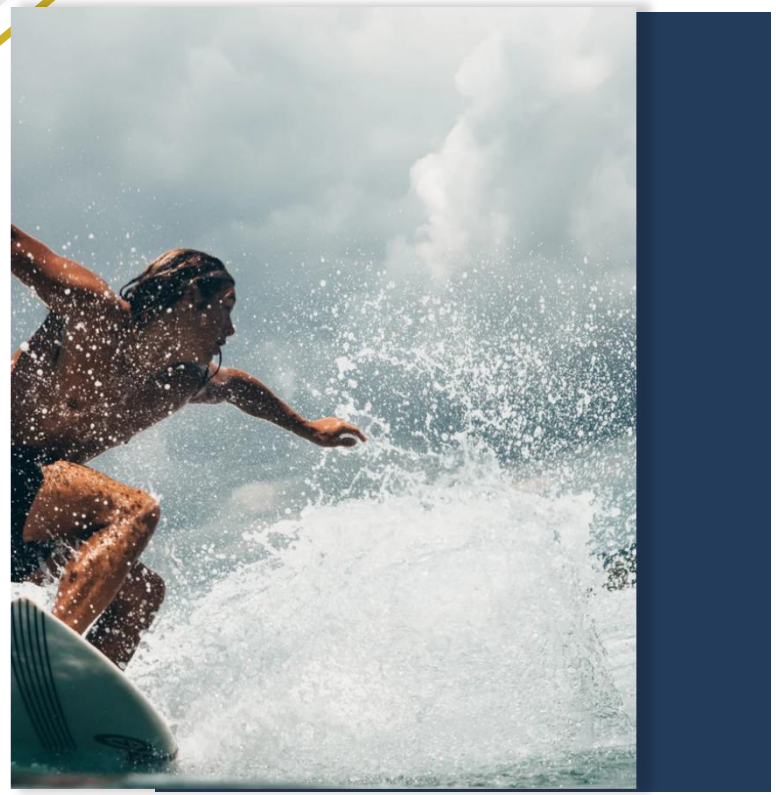
Datasets

- Tweets in the form of hashtags, comments and retweets discussing the various aspects of the Paris Olympics.
- The dataset has 15 columns

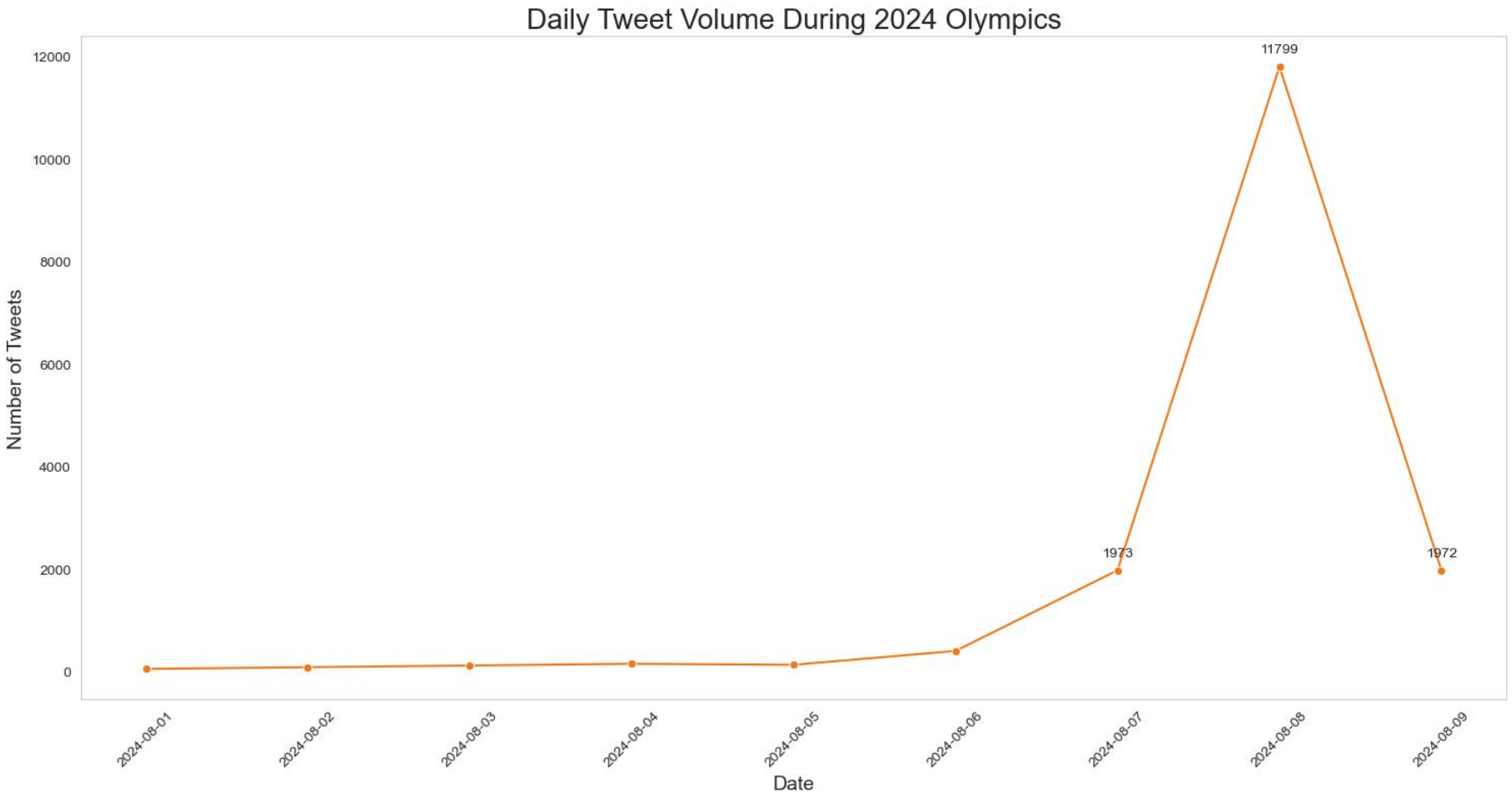
Relevance of the data

- Relevant hashtags allow for more targeted analysis potentially revealing topic-specific sentiments.

VISUALIZATIONS



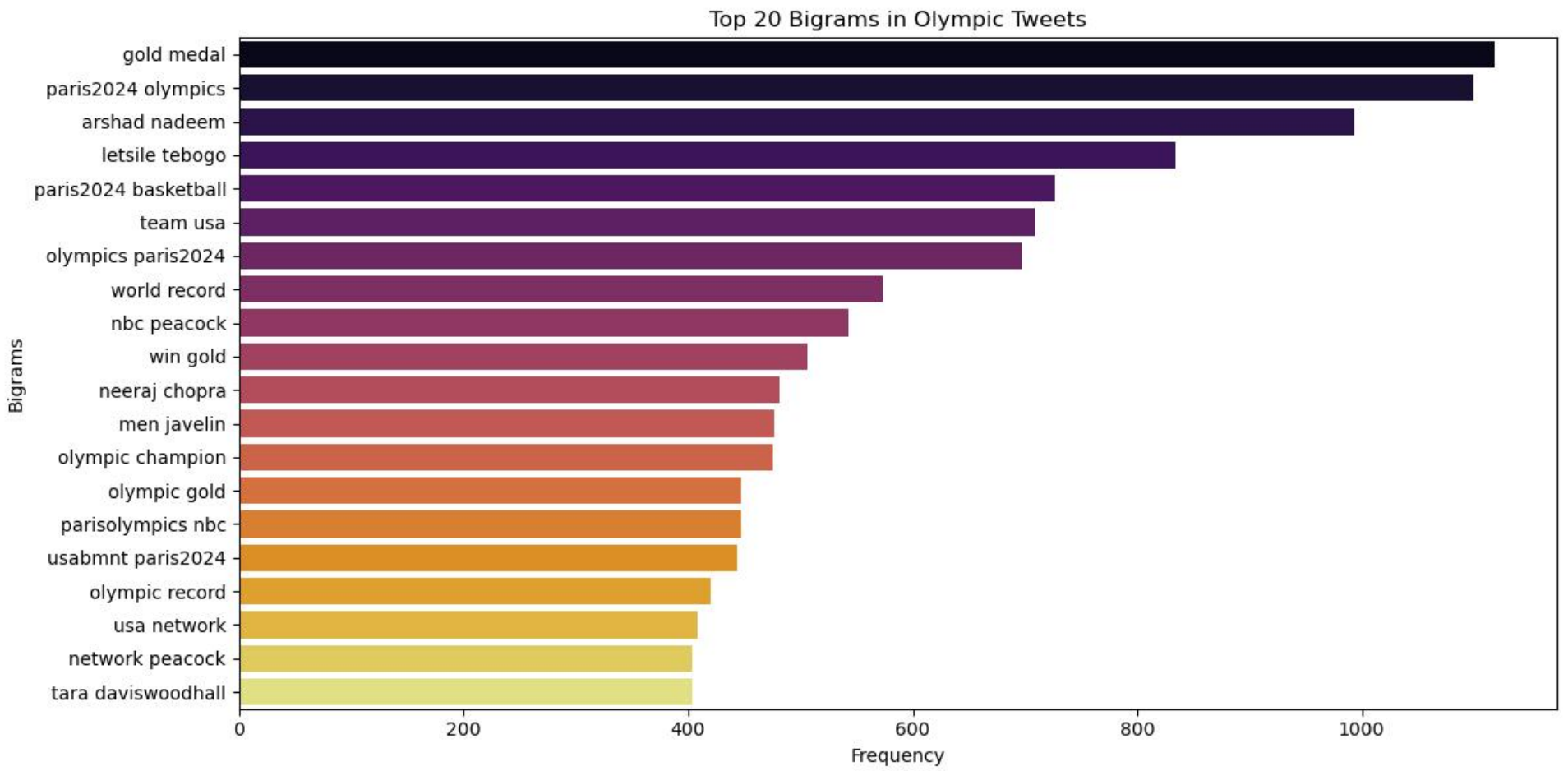
Tweet Volume During the Month Of August



Peak Activity: There's a dramatic spike in tweet volume on August 8, 2024, reaching 11,799 tweets. This suggests a major event or significant happening on this day.

Build-up:There is also a gradual increase in tweet volume leading up to the peak, with notable jumps on August 6 and 7.

Tweet Volume During the Month Of August



Observation

The plots show the most prominent bigrams include phrases like "gold medal," "paris2024 olympics" and "arshad nadeem," indicating that discussions around medal achievements, the Paris 2024 Olympics and specific athletes are among the most common topics in these tweets.

Most Prominent Hashtags During the Olympics

Observation

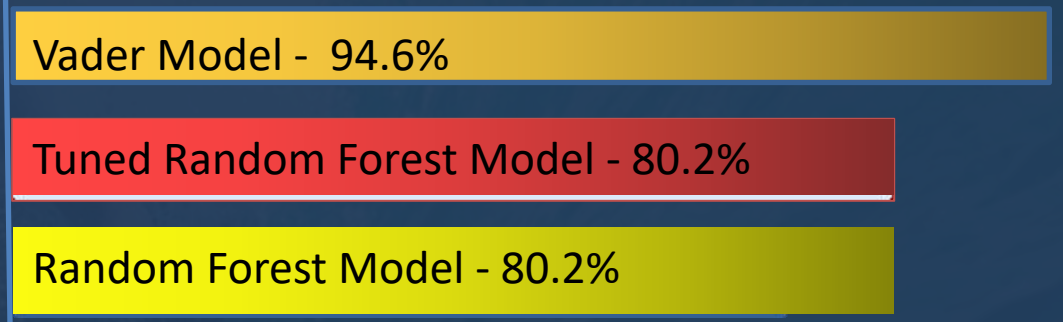
The word cloud highlights "#Paris2024," "#ParisOlympics and "#Olympics2024" as the most frequently used hashtags, emphasizing the focus on the Paris 2024 Olympic Games. Hashtags like "#NeerajChopra" and "#Basketball" also stand out, indicating key topics and athletes being discussed

MODELING RESULTS

Model Evaluation

The best Model is the Vader Model with an Accuracy of 0.9460, precision of 0.9474,Recall of 0.9460 and a F1Score 0.9465

ACCURACY



RECOMMENDATIONS

1. Implement a real-time sentiment tracking dashboard for organizers and media partners, allowing them to respond quickly to shifts in public opinion.

2. Develop a multi-lingual sentiment analysis capability to cater to the international nature of the Olympics, using language-specific versions of VADER where available.

3. Integrate sentiment analysis results with other data sources (e.g., ticket sales, TV ratings) to provide a comprehensive view of public engagement.

4. Use sentiment trends to guide content creation and social media strategies, focusing on themes and athletes that generate positive engagement.

NEXT STEPS

Step 1

Enhance Feature Engineering: Incorporate Olympics-specific features such as mentions of specific sports, athletes, or events to improve classification accuracy.

Step 2

.

Develop Custom Olympic VADER Lexicon: Create a specialized lexicon for VADER that includes Olympic-specific terms and their sentiment associations

Step 3

Extend the sentiment analysis to multiple social media platforms and news sources for a more comprehensive view.

Step 4

Develop user-friendly, interactive dashboards for stakeholders to explore sentiment data in real-time.

Step 5

Preprocess the multilingual olympics data and fine-tune a model to accurately detect sentiment across the different languages.

Step 6

Set up infrastructure for continued analysis post-Olympics to track the event's lasting impact on public sentiment towards Paris and the Olympic movement

THANK YOU