MILESTONE

Eliott Coutaz, Clément-Louis Joyat, Christian Saidi

WHY

Football. Probably the most entertaining sport on the planet. In every country, we play, we follow football.

What makes this sport magic ? Is it only the technicity of the players, the beautiful geometry of every tactic, the terrible efforts everyone has to do during an hour and a half ? Maybe there is something else. Maybe football has another dimension.

It might not only be a match but the modern way to fight another city or another country and that's why in the pulsating world of football, fan engagement is the lifeblood that fuels the sporting spectacle. During a match, the supporters are often called the 12th man, they give life to the stadium and the match.

But let me ask a question, can this influence start before the match ? That's the first question we will try to answer, by seeking to tap into the vast reservoirs of fan conversations across social media platforms and forums to understand and quantify their engagement.

On the same network, people talk before the game but after too. We plan to look how the messages can influence the match but we can try to find how the result will influence the messages after the match too. These two questions will be our guideline. Our goal is to analyze sentiments expressed online to provide insights that could shape more responsive and engaging fan interaction strategies.

For this project, we are able to recover our proper dataset. We also found one already on Kaggle.

In the same will, but not completely following neither the goal or the tools, we found a study that delves into the crucial role of social media content in fostering engagement between fans and collegiate sports teams. It reviews existing literature on sentiment analysis and content categorization in the sports industry, presenting an analytic framework for sports organizations to leverage AI and social media effectively. The framework spans data collection, preparation, sentiment mining, and content categorization.

The study showcases the application of text mining and sentiment analysis for enhanced data understanding by athletic departments, leading to improved fan engagement without escalating costs. Analyzing social media data from a major U.S. university's basketball team, the study develops analytical models to classify fan comments' polarity and subject matter.

So now that's the WHY, explaining the origin of the idea, our goals and what was already done in this subject is treated. Let's deepen our method and its tools.

WHAT

Hypothesis and Objectives:

As previously highlighted, the intricate relationship between the support of fans, often referred to as the "12th man," and its impact on football outcomes is an intriguing aspect worth exploring. Moreover, the reciprocal nature of a team's performance and the passionate backing from its supporters hints at a nuanced connection that warrants in-depth investigation.

Turning our focus to the remarkable journey of Morocco during the 2022 World Cup, their rise from underdog status was accompanied by unwavering support from the vast Arab football fanbase. Placed in a daunting group comprising Canada, Croatia, and Belgium, their progression to the later stages of the tournament defied expectations, marking a triumph against formidable opponents. The unexpected victories against Cristiano Ronaldo's Portugal in the Last 16 and a subsequent triumphant clash against Spain showcased the team's resilience in the face of adversity.

Central to our investigation is the conviction that the fervent support of Morocco's fanbase served as a catalyst for the team's remarkable success. To substantiate this claim, our study will focus on dissecting selected matches from Morocco's World Cup campaign. Specifically, we aim to analyze the impact of fan sentiments expressed on Twitter during critical encounters such as Morocco vs Portugal and Morocco vs Spain.

Our methodology entails leveraging advanced machine learning techniques to meticulously scrutinize the sentiments conveyed through fan comments on Twitter. By employing sentiment analysis and data mining, we seek to extract meaningful insights, identifying trends, sentiment patterns, and correlations between fan engagement and the team's performance. Through this meticulous analysis, we aim to validate our hypothesis concerning the influence of fan support on the outcomes of these pivotal matches. Now that we've established the Why and What of our study, our focus shifts to the methodology that will guide us in unearthing insightful results.

HOW

Data Collections

Our project leverages two primary data sources: Twitter and a sports data API. We utilized the Twitter API to collect tweets related to football games, focusing on specific hashtags and keywords tied to team names and match events. The time frame for data collection was strategically set to periods before and after the football matches, allowing us to capture the changing sentiment of the audience. For the football game results, we accessed a sports data API, which provided up-to-date results including team scores and standings.

Data Processing

For processing the collected tweets, we employed the TextBlob library, a powerful tool for natural language processing, particularly effective in sentiment analysis. The TextBlob library allowed us to analyze each tweet and assign a polarity score, indicating whether the

sentiment was positive, neutral, or negative. Additionally, we used the Pandas library for its robust data handling capabilities, organizing the tweet data into a DataFrame format for more efficient processing and analysis.

The game results obtained from the sports API were handled using Python's Requests library to make HTTP requests, and the resulting data, typically in JSON format, was parsed to extract relevant information such as the names of the teams involved and the final scores.

Sentiment Analysis and Comparative Analysis

The core of our analysis involved determining the polarity of each tweet to categorize them into positive, negative, or neutral sentiments. By comparing the sentiment distribution before and after the games, we aimed to observe significant shifts in public opinion and mood. This sentiment data was then juxtaposed with the actual game results to explore any correlations between public expectations or reactions and the game outcomes.

Visualization

We chose Matplotlib, a versatile plotting library in Python, for visualizing our analysis results. This included creating bar graphs that effectively represented the distribution of sentiments across different categories and times, providing a clear visual representation of our findings.

Data Storage and accessibility

All project-related code and documentation are maintained in a GitHub repository, ensuring version control and widespread accessibility. Given the size and sensitivity of tweet data, the raw data itself was not stored in the repository. Instead, we structured the code to enable users to fetch and process the data independently, ensuring both scalability and privacy.

Here is some early data showing that our sentiment analysis works!

Sentiment Analysis of Morocco vs Spain Game (Adjusted Data)
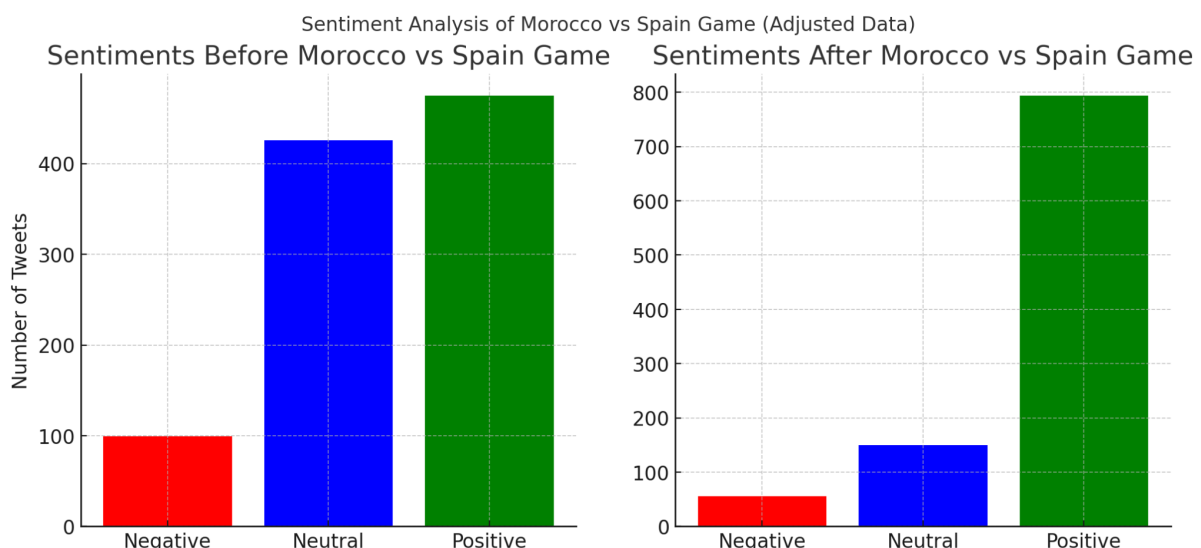
The hashtags used were #Morocco #MoroccoWin and different iterations based on Twitter searches.

The link to our github Repo is https://github.com/GnomeGuerrier/Final-project_Data_Science