Impact of Fan Engagement on Football Game Outcomes: A Data-Driven Analysis

ELIOTT COUTAZ, University of Uvic, Spain
CHRISTIAN SAIDI, University of Uvic, Spain
CLEMENT-LOUIS JOYAT, University of Uvic, Spain

This study investigates the influence of fan engagement, as expressed through social media platforms, on the outcomes of football games. Utilizing sentiment analysis and data mining techniques, we analyze fan sentiments on Twitter and correlate them with game results, particularly focusing on the 2022 World Cup matches involving the Morocco team.

Additional Key Words and Phrases: Fan Engagement, Football, Sentiment Analysis, Data Mining, Social Media, 2022 World Cup

ACM Reference Format:
Eliott Coutaz, Christian Saidi, and Clement-Louis Joyat. 2024. Impact of Fan Engagement on Football Game Outcomes: A Data-Driven Analysis. 1, 1 (January 2024), 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

1.1 The Enchantment of Football

Football, universally acclaimed as the most captivating sport worldwide, holds a unique place in the hearts of millions. This enchanting game is not only about the players' skillfulness, tactical strategies, and physical endurance, but also about the extraordinary atmosphere created by its fervent fans. This study delves into an often-overlooked aspect: how fan engagement, particularly through social media platforms, impacts the outcomes of football matches.

1.2 Exploring the Fan Influence

The core of our research is to explore and quantify the influence of fans' online engagement on their teams' performances. By analyzing social media interactions, primarily Twitter, we aim to uncover whether the sentiments and emotions expressed by fans correlate with the performance and outcomes of football games. Our focus is on examining this dynamic in the context of high-stakes tournaments, using the 2022 World Cup as a primary case study.

1.3 Methodological Approach

To achieve our objectives, we employ advanced data science techniques. These include sentiment analysis and data mining to process and analyze large volumes of tweets related to specific matches. This innovative approach allows us to draw meaningful insights from the digital footprints of fan engagement, offering a new perspective on the intersection of sports, social media, and collective emotions. We utilize tools such as the Twitter API, TextBlob for sentiment polarity

Authors' addresses: Eliott Coutaz, eliott.coutaz@email.com, University of Uvic, Vic, Spain; Christian Saidi, Christian.saidi.2003@gmail.com, University of Uvic, Vic, Spain; Clement-Louis Joyat, Clementlouisjoyat@gmail.com, University of Uvic, Vic, Spain.

Manuscript submitted to ACM

classification, and Pandas for data processing. Our methods also involve predictive modeling and comparative analysis to understand the relationship between fan sentiment and game outcomes.

2 BACKGROUND AND RATIONALE

2.1 Inspiration and Objectives

For this project, we created our dataset and also found a relevant dataset on Kaggle. Our study is inspired by previous research examining the role of social media in enhancing engagement between fans and sports teams. This background research underscores the importance of sentiment analysis and content categorization in the sports industry, employing an AI-driven framework for effective social media usage by sports organizations.

2.2 Case Study of Morocco in the 2022 World Cup

Our research focuses on Morocco's remarkable performance in the 2022 World Cup. We examine the intense support from their fans and its potential impact on the team's success. The analysis includes pivotal matches such as Morocco vs Portugal and Morocco vs Spain, exploring the hypothesis that fan support on social media platforms, particularly Twitter, influenced the team's performance.

3 METHODOLOGY

3.1 Data Collection

Our data collection utilizes two primary sources: the Twitter API and a sports data API. We collected tweets related to football matches, focusing on specific hashtags and keywords, and gathered football game results from the sports data API, including team scores and standings.

3.2 Data Processing

The collected tweets were processed using the TextBlob library for sentiment analysis and Pandas for data organization. The sports API data was handled using Python's Requests library, extracting relevant information like team names and final scores.The sentiment score is situated between 0 and 2, with "0" being negative, "1" neutral and "2" positive.

3.3 Sentiment Analysis and Comparative Analysis

The sentiment of each tweet was classified as positive, negative, or neutral. By comparing sentiment distributions before and after games, we aimed to identify correlations between public sentiment and game outcomes.

3.4 Correlating Fan Sentiment and Game Outcomes using predictive modeling

The final aspect of our methodology involves developing a machine learning model to correlate fan sentiment with game outcomes. This model aims to validate our hypothesis that fan engagement on social media platforms has a tangible impact on football game results.

4 SENTIMENT ANALYSIS OF THE 2022 MOROCCO VS SPAIN MATCH

4.0.1 This sentiment analysis is done on the Moroccans fans.

4.1 Sentiment Classification and Distribution

Our analysis begins with classifying the sentiments of Moroccan fans before the match. We conducted sentiment analysis to understand the emotional undercurrents among supporters.

4.1.1 Figure 1: Pie Chart of Sentiment Distribution. The first graphic, a pie chart, presents an overview of the sentiments with percentages for positive, negative, and neutral tweets. This graphic helps illustrate the general mood among Moroccan fans prior to the game.

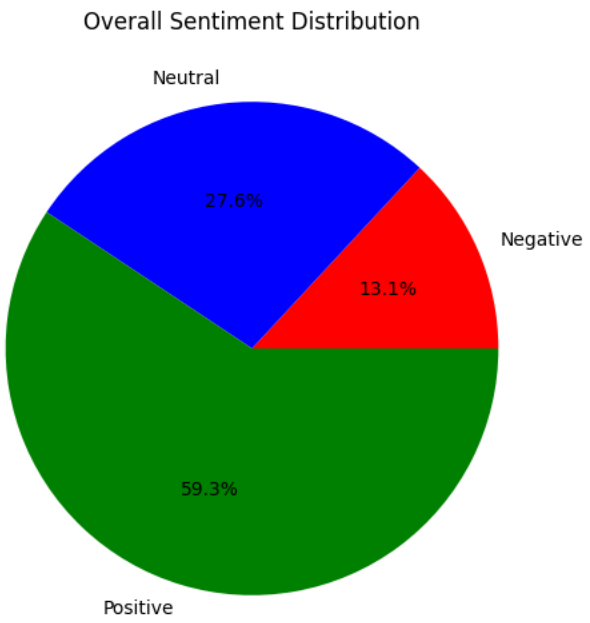

Fig. 1. Pie Chart of Sentiment Distribution - 2022 Game

4.2 Contextual Analysis of Fan Sentiments

Despite the favorable odds for Spain, characterized by their formidable team formation and experience in high-stakes games, our analysis revealed a contrasting sentiment among Moroccan fans. A significant majority expressed confidence in their team's victory, with only a small fraction anticipating the end of the Atlas Lions' journey.

4.2.1 Figure 2: Bar Graph of Sentiment Intensity. The second graphic, a bar graph, explores sentiment intensity distribution, complementing the pie chart by highlighting the frequency of each sentiment expression.

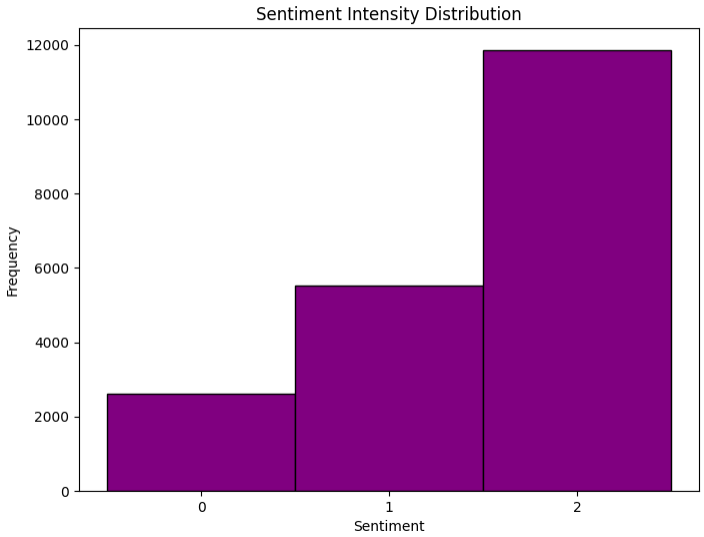

Fig. 2. Bar Graph of Sentiment Intensity Distribution - 2022 Game

Despite the favorable odds for Spain, characterized by their formidable team formation and experience in high-stakes games, our analysis revealed a contrasting sentiment among Moroccan fans. A significant majority expressed confidence in their team's victory, with only a small fraction anticipating the end of the Atlas Lions' journey.

4.3 Sentiment Dynamics on Match Day

4.3.1 Figure 3: Sentiment Over Time. Our third graph illustrates the sentiment trends over time on the match day. It calculates the average sentiment hourly, with the scale ranging from 0 (negative) to 2 (positive).

This graph indicates a consistent trend of positive sentiment, intensifying as the match approached. The Moroccan team's spirited performance and acknowledgment of their supporters' trust reinforce the presumed correlation between fan engagement and on-field performance.

4.4 Sentiment Comparison: Pre- and Post-Game

4.4.1 Figure 4: Comparative Sentiment Analysis. The fourth graph provides a comparative analysis of sentiments before and after the game. It explores how fan sentiments influence the game and, conversely, how game outcomes impact fan moods.

While it is intuitive that fans' moods shift in response to game results, our analysis extends beyond this to examine how these sentiments align with broader trends in fan engagement and game outcomes.

5 SENTIMENT ANALYSIS OF THE 2018 MOROCCO VS SPAIN MATCH

5.0.1 This sentiment analysis is done on the Moroccans fans. In a similar vein to our analysis of the 2022 match, we extended our sentiment analysis to the 2018 Morocco vs Spain game. Employing the same methodology, we analyzed

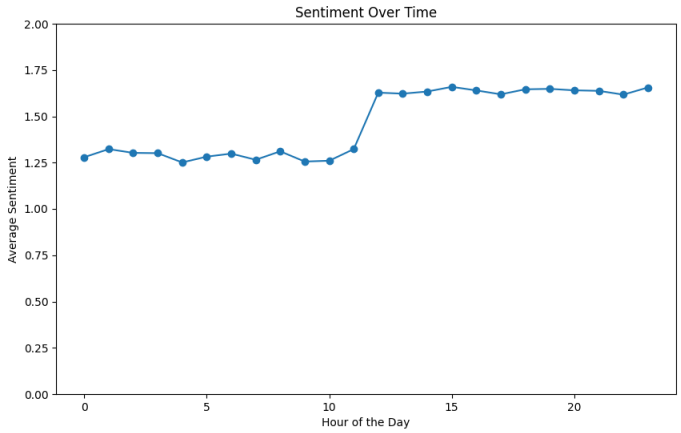

Fig. 3. Sentiment Over Time on Match Day - 2022 Game

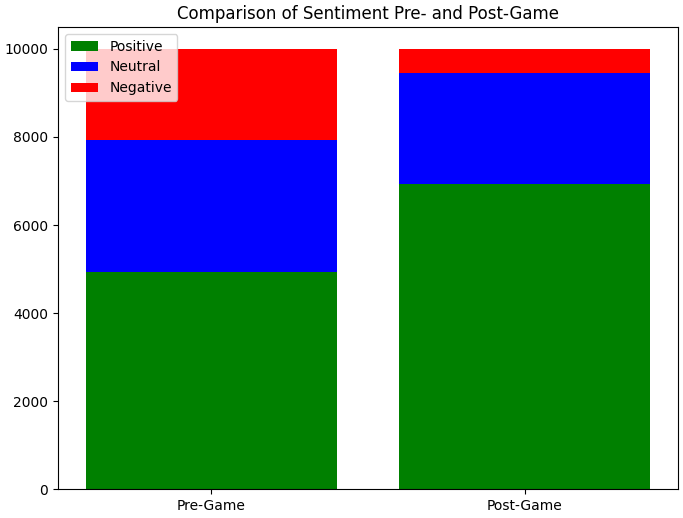

Fig. 4. Comparative Sentiment Analysis Pre- and Post-Game - 2022 Game

fan sentiments to draw comparisons and observe changes over time. Below are the four graphics representing various aspects of the sentiment analysis for the 2018 match, presented in pairs for a comparative view.

5.1 Graphics for the 2018 Game

This approach allows us to juxtapose the data from the two different years, offering insights into the evolution of fan sentiments and their potential impact on the team's performance in these high-profile matches.

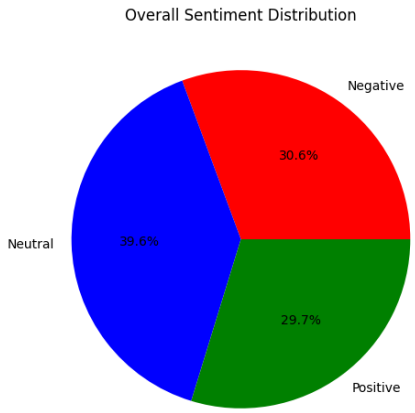

Fig. 5. First Graphic - 2018 Game

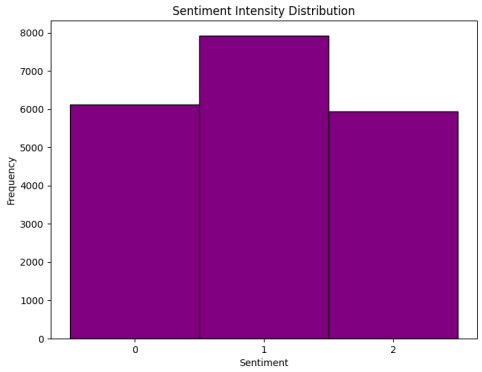

Fig. 6. Second Graphic - 2018 Game

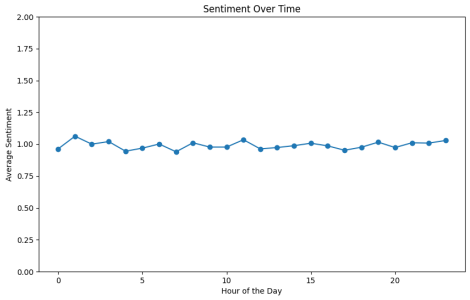

Fig. 7. Third Graphic - 2018 Game

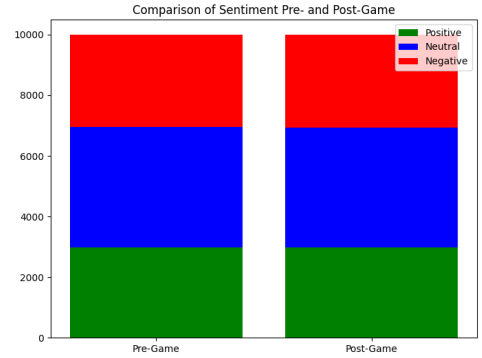

Fig. 8. Fourth Graphic - 2018 Game

6 PREDICTIVE MODELING FOR CORRELATING FAN SENTIMENT AND GAME OUTCOMES

6.1 Model Development

To predict football match outcomes based on Twitter fan sentiment, we developed a machine learning model using the RandomForest classifier. The model was trained on a dataset comprising sentiment texts and corresponding game outcomes, utilizing TF-IDF for feature extraction and a RandomForest classifier for prediction.

6.2 Model Evaluation and Results

Upon evaluating the model against 200 games, we observed the following:

- The model predicts losses with a 55% probability and wins with a 69% probability.
- The precision rates for predicting losses and wins are 55% and 69%, respectively.
- The recall rates are 80% for losses and 40% for wins.
- The F1-scores are 0.65 for losses and 0.51 for wins, indicating a better balance between precision and recall for losses.
- Overall model accuracy stands at 59%, suggesting it performs marginally better than random chance.

The classification report is as follows:

Table 1. Classification Report

Class	Precision	Recall	F1-Score	Support
Loss	0.55	0.80	0.65	95
Win	0.69	0.40	0.51	105
Accuracy			0.59	200
Macro Avg	0.62	0.60	0.58	200
Weighted Avg	0.62	0.59	0.57	200

6.3 Comparison of Predictions and Actual Results

In addition to the previously discussed metrics, Figure 9 provides a visual comparison of the model's predictions against the actual game outcomes. Each point on the graph represents a game, with the horizontal axis numbering the games and the vertical axis denoting the outcome (win or loss). The proximity of the predicted outcomes (red) to the actual outcomes (blue) illustrates the model's accuracy on a game-by-game basis. Notably, the graph demonstrates the model's propensity to correctly predict losses more frequently than wins, which aligns with our earlier findings on the model's higher precision and recall for losses.

6.4 ROC and Precision-Recall Curves

The model's performance was further visualized using ROC and Precision-Recall curves, with the ROC curve demonstrating a 60% chance of correctly distinguishing between classes (AUC = 0.60) as seen on 10, and the PR curve indicating moderate performance (area = 0.59).

The area under each curve suggests that while the model performs moderately well, there is room for improvement, particularly in enhancing the prediction of wins based on pre-game fan sentiment.

6.5 Evaluation

Our findings suggest a correlation between pre-game fan sentiment and match outcomes, with a 60% ROC AUC reinforcing this relationship. This model, despite its limitations, provides a foundation for further exploration into the predictive power of fan sentiment in sports outcomes.

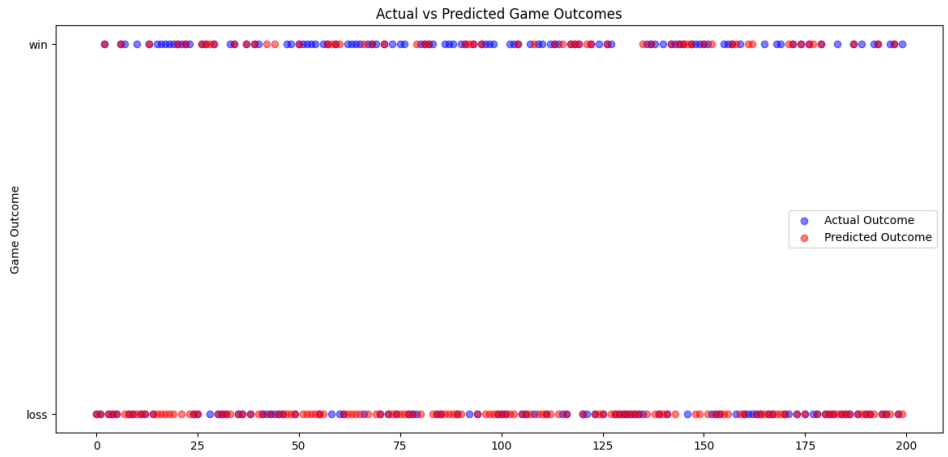

Fig. 9. Comparison of Actual vs. Predicted Game Outcomes

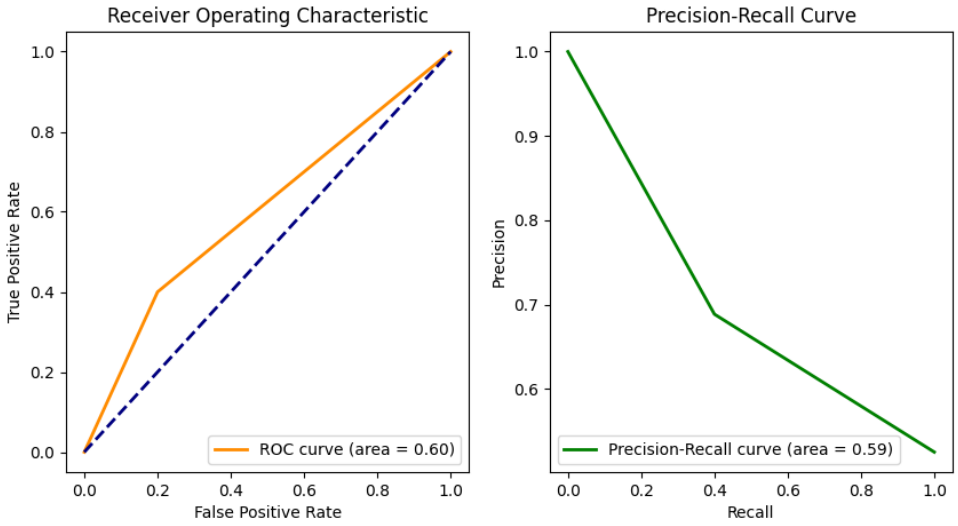

Fig. 10. Receiver Operating Characteristic Curve and Precision-Recall Curve

7 CONCLUSION

7.1 Synthesis of Findings

Our comparative analysis of fan sentiment before the two critical Morocco vs. Spain matches—namely, the 2018 World Cup group stage and the 2022 World Cup quarter-final—demonstrates a significant difference in fan engagement and its apparent influence on the team's performance.

7.1.1 2018 World Cup Group Stage Sentiment. In 2018, Morocco entered their final group stage match against Spain with minimal fan enthusiasm, as the team had already been eliminated from the competition after two losses. This lack of support is reflected in the team's performance, culminating in a 2-2 draw—a scoreline that experts suggest could have tilted further in Spain's favor.

7.1.2 2022 World Cup Quarter-Final Sentiment. Contrastingly, the 2022 quarter-final saw a starkly different scenario. Morocco's unexpected success in the group stages, coupled with their triumph over Portugal, ignited a wave of fan optimism. This enhanced sentiment is believed to have positively impacted the team's performance against Spain, leading to a commendable 0-0 result, defying the predicted dominance by Spain.

7.1.3 Correlation Between Fan Sentiment and Game Outcome. With the confirmation of the model we built using the RandomForest classifier, we could predict with a certainty of 60% the outcome of the game using the fan sentiment, proving the correlation between the fan sentiment and the final score of the game.

7.2 Impact of Fan Sentiment on Team Performance

The correlation between fan sentiment and game outcomes not only underscores the team's victories but also the heightened confidence and support from the fans. The bar chart below illustrates the average fan sentiment in 2018 and 2022, providing a visual representation of this correlation.

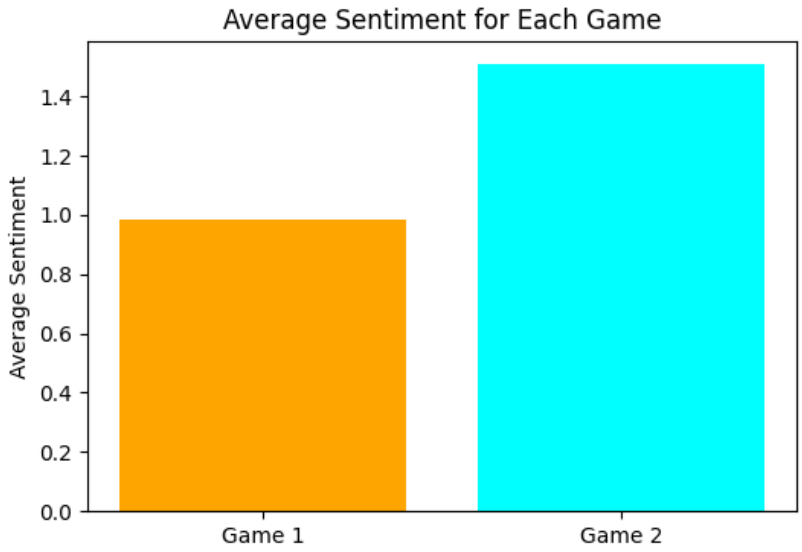

Fig. 11. Average Fan Sentiment in 2018 vs. 2022

Referencing Figures 5, 6, 7,8, 3, 2,4, and 1 from our analysis, we can conclude that the Moroccan team's performance is not only a reflection of their on-field strategies and capabilities but also a product of the emotional support and confidence instilled by their fans.

7.3 Concluding Thoughts

This study reveals the nuanced yet discernible impact of fan sentiment on football match outcomes. The Moroccan team's comparative performances in 2018 and 2022 serve as a testament to the potential of fan engagement as a contributing factor to sports success. It invites a broader consideration of how emotional investment and support can play pivotal roles in the dynamics of competitive sports.

ACKNOWLEDGMENTS

We would like to extend our deepest gratitude to Professor Jordi Villà Freixa, whose insightful questions and guidance were instrumental in the conception and development of this research. The direction and expertise provided by Professor Villà Freixa not only shaped our analytical approach but also enriched our understanding of the subject matter. This work benefitted immensely from their scholarly perspective and academic rigor.

REFERENCES

[1] Coutaz, E., *Final Project Data Science*. GitHub repository for this paper. Available at: https://github.com/GnomeGuerrier/Final-project_Data_Science.git.
[2] Singh, T., *Twitter Sentiment Extraction - Analysis, EDA and Model*, Kaggle (2020). Available at: https://www.kaggle.com/code/tanulsingh077/twitter-sentiment-extaction-analysis-eda-and-model.
[3] Entity Sentiment Analysis, *Kaggle Dataset*. Available at: https://www.kaggle.com/datasets/jp797498e/twitter-entity-sentiment-analysis.
[4] Biocomputing-Teaching, *Data Science with Python Course Materials*, GitHub repository. Available at: https://github.com/Biocomputing-Teaching/Data-Science-with-Python.
[5] Kaggle Discussion, *General Discussion on Twitter Sentiment Analysis*, Kaggle (2020). Available at: https://www.kaggle.com/discussions/general/327257.
[6] Pal, S., *Awesome Twitter Data*. GitHub repository. Available at: https://github.com/shaypal5/awesome-twitter-data.