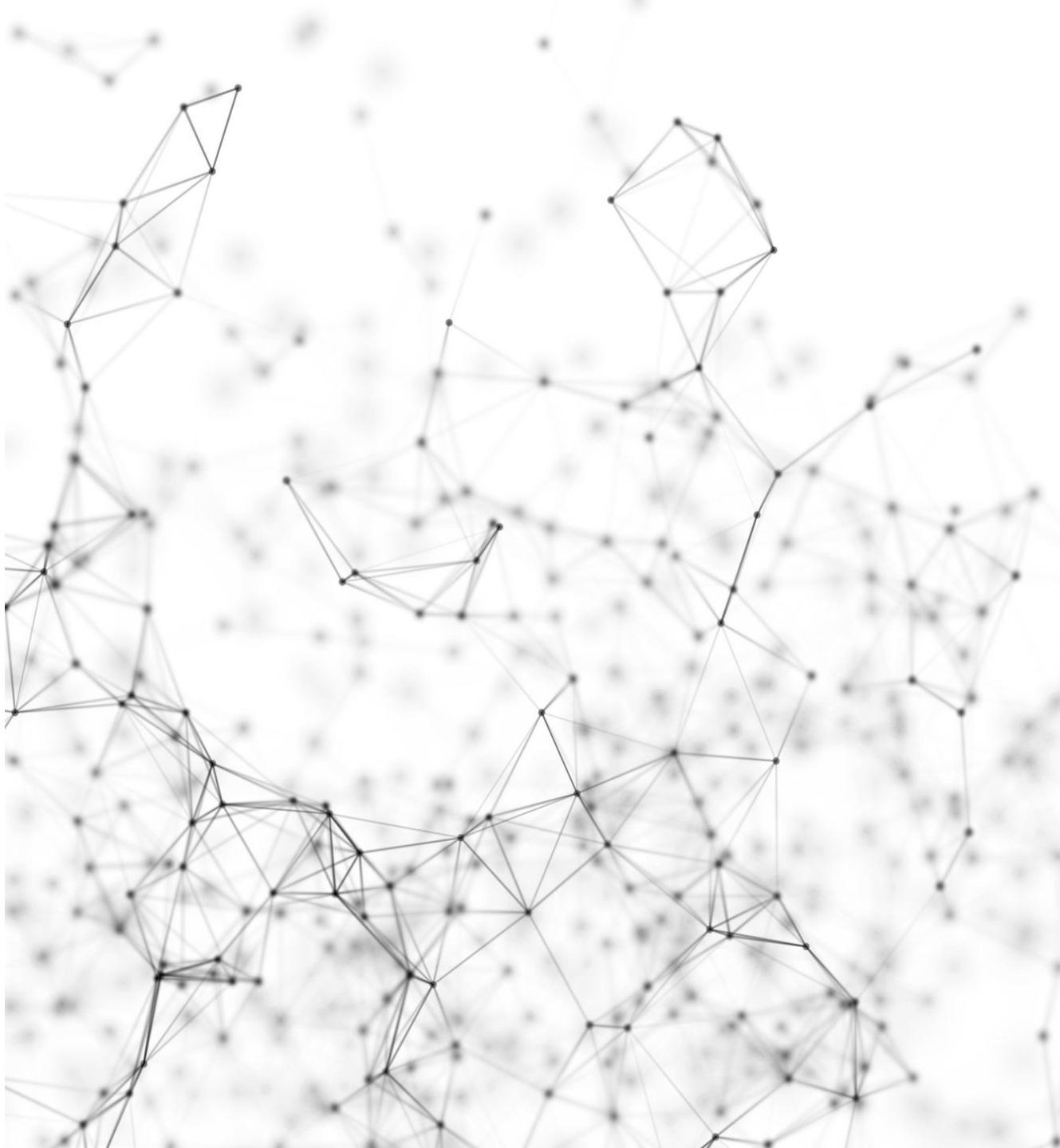
BEYOND THE SCORE: SENTIMENT ANALYSIS AS A TOOL FOR DIFFERENTIATING SOCCER AND FOOTBALL DISCUSSIONS ON REDDIT

By: Giovanni Lunetta



Introduction

Regardless of their intensity of fandom, most people globally engage with sports to some extent, often taking to online platforms like X (formerly Twitter) and Reddit for discussions, particularly when their teams show extreme performances, spurring increased internet interaction.

Previous work

Quantifying fan engagement in sports using text analytics

Original Article | Published: 11 June 2021

The Game-Changing Power of Text Annotation in Sports: Enhancing Performance and Analysis

Annotation Box
242 followers

+ Follow

November 10, 2023

Text Analytics for Sports Fan Engagement in Social Media

Amir Zadeh, *Wright State University*

Follow

8-10-2020

Text Analytics for Sports Fan Engagement in Social Media

Amir Zadeh
Wright State University, amir.zadeh@wright.edu

Research Objective

- Question: Which sports fans are more likely to make positive/negative comments about their teams and in what circumstances?
- Purpose: Use sentiment analysis and a random forest model to predict the sport (soccer or football) based on Reddit comments.

Dataset Overview

- Source: Reddit posts from football and soccer discussions.
- Time Frame: SuperBowl 47 (2013) to SuperBowl 57 (2023), 1 regular season NFL game, UEFA Champions League Final (2013 to 2023), 3 knockout games.
- Variables: Comment text, upvotes, sport, phase, category.

Analytical Approach

- Text Preprocessing: Tokenization, lemmatization, emojis to text, handling URLs.
- Sentiment Analysis: Using NLTK's SentimentIntensityAnalyzer for sentiment scoring.
- Modeling: Predictive modeling with random forests to classify comments by sport.

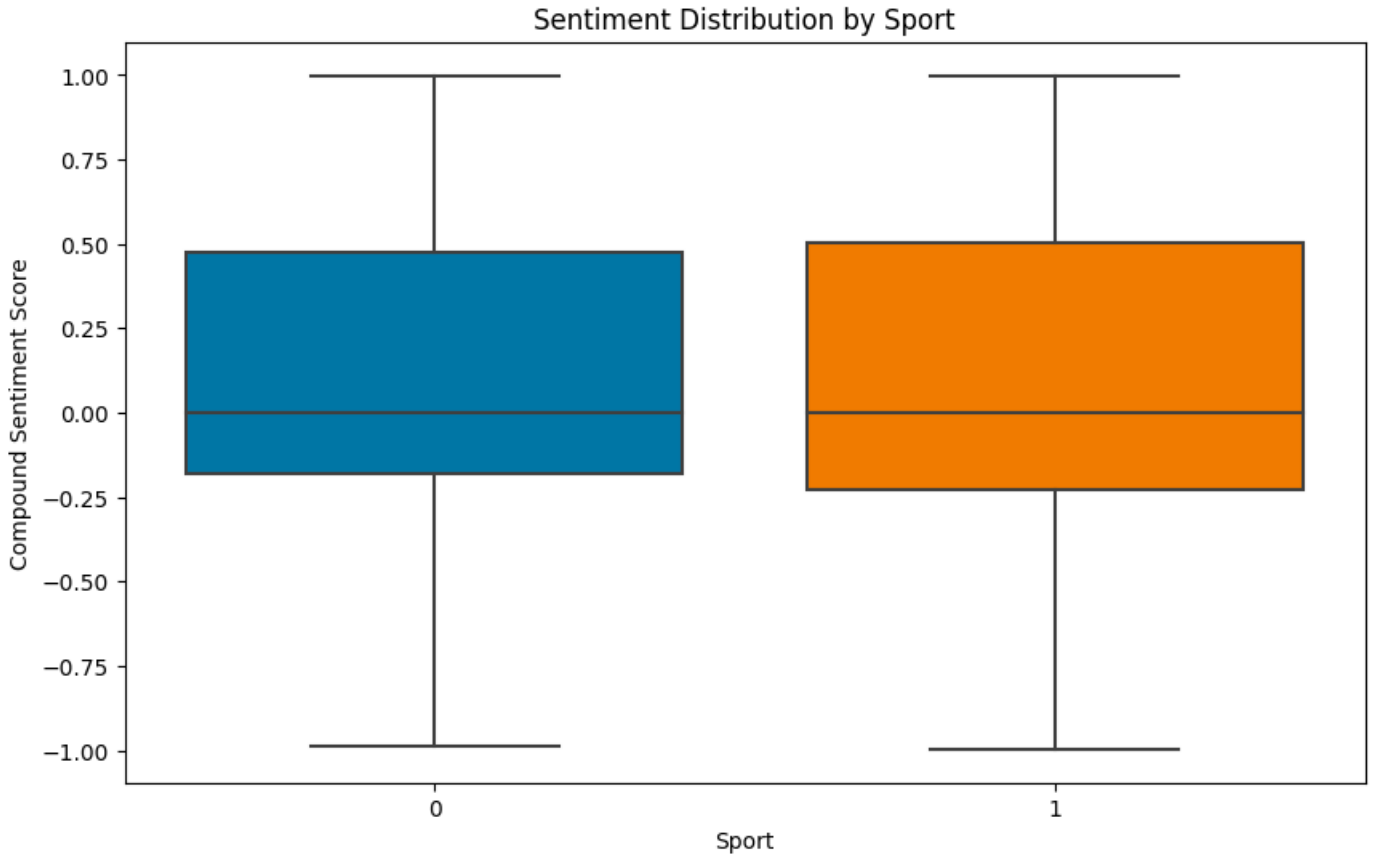
```
df["tokens"] = df.text.apply(
    process_text,
    lower_case = False,
    remove_punct = True,
    keep_specific_punct = True,
    remove_stopwords = False,
    lemma = True,
    emojis_to_words = True,
    convert_urls = True,
    string_or_list = "str")
```

Sentiment Analysis Insights

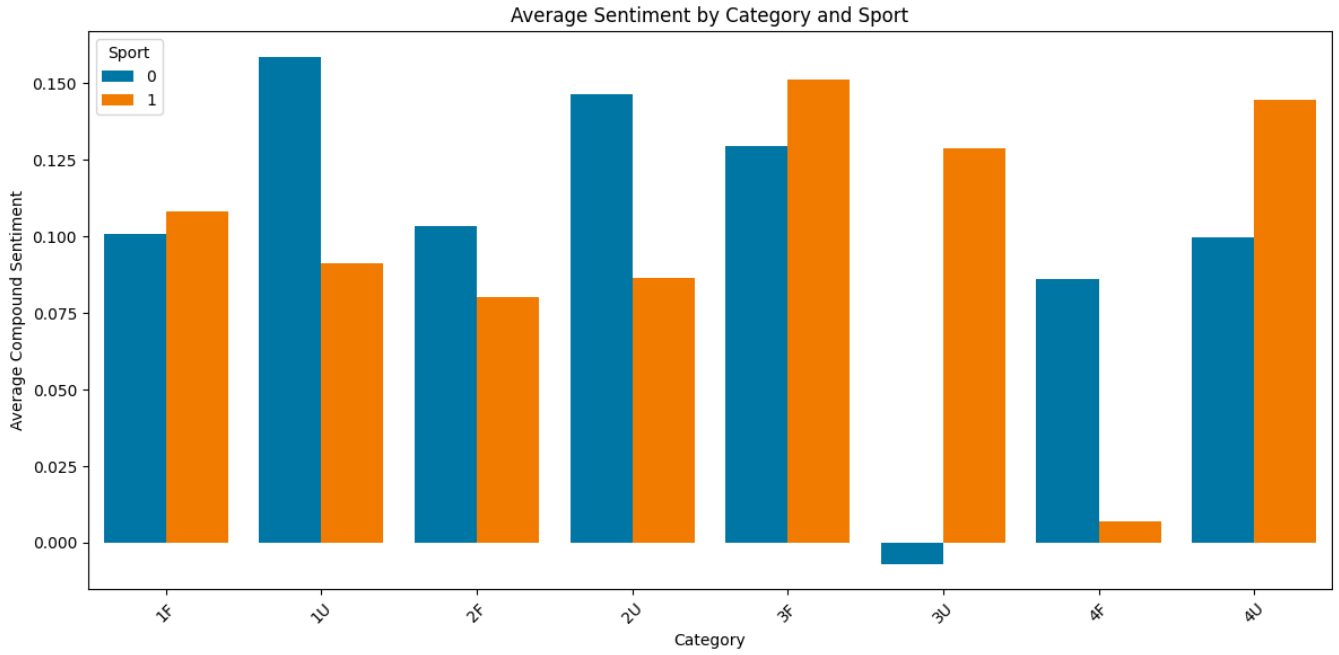
- Differences in sentiment between sports and conditions.
- Impact of game phase and team performance on sentiment.
- Visualization: Sentiment distribution across sports and game phases.

post_id	text	upvote	sport	phase	category	tokens	compound_sentiment
718	That's [redacted] terrifying.	1.0	1	1	2U	That [redacted] terrifying	-0.6115
4209	Alright Niners...it's a two possession game. W...	5.0	1	2	2F	Alright Niners it a two possession game We can...	0.4359
18518	Peyton is really [redacted] funny	5.0	1	1	3F	Peyton is really [redacted] funny	0.5379
75925	I am extremely biased but greatest run I've ev...	12.0	0	3	2U	I am extremely biased but greatest run I ever ...	0.7579
20002	The lowest scoring Super Bowl in history is al...	8.0	1	3	3F	The lowest scoring Super Bowl in history is al...	0.7269
33198	GET IN HERE BOYS	5.0	1	3	3F	GET IN HERE BOYS	0.0000
1746	YEAH BABY!	1.0	1	3	2U	YEAH BABY !	0.3595
84121	Please Sidibé, Raggi and Subasic have a good g...	2.0	0	2	4U	Please Sidibé Raggi and Subasic have a good ga...	0.5267
28667	Well I guess everyone is just getting offended...	5.0	1	1	1F	Well I guess everyone is just getting offended...	-0.8247
32905	If someone would have asked me who my 1st choi...	7.0	1	3	3U	If someone would have asked me who my choice w...	0.9382

Visualization 1



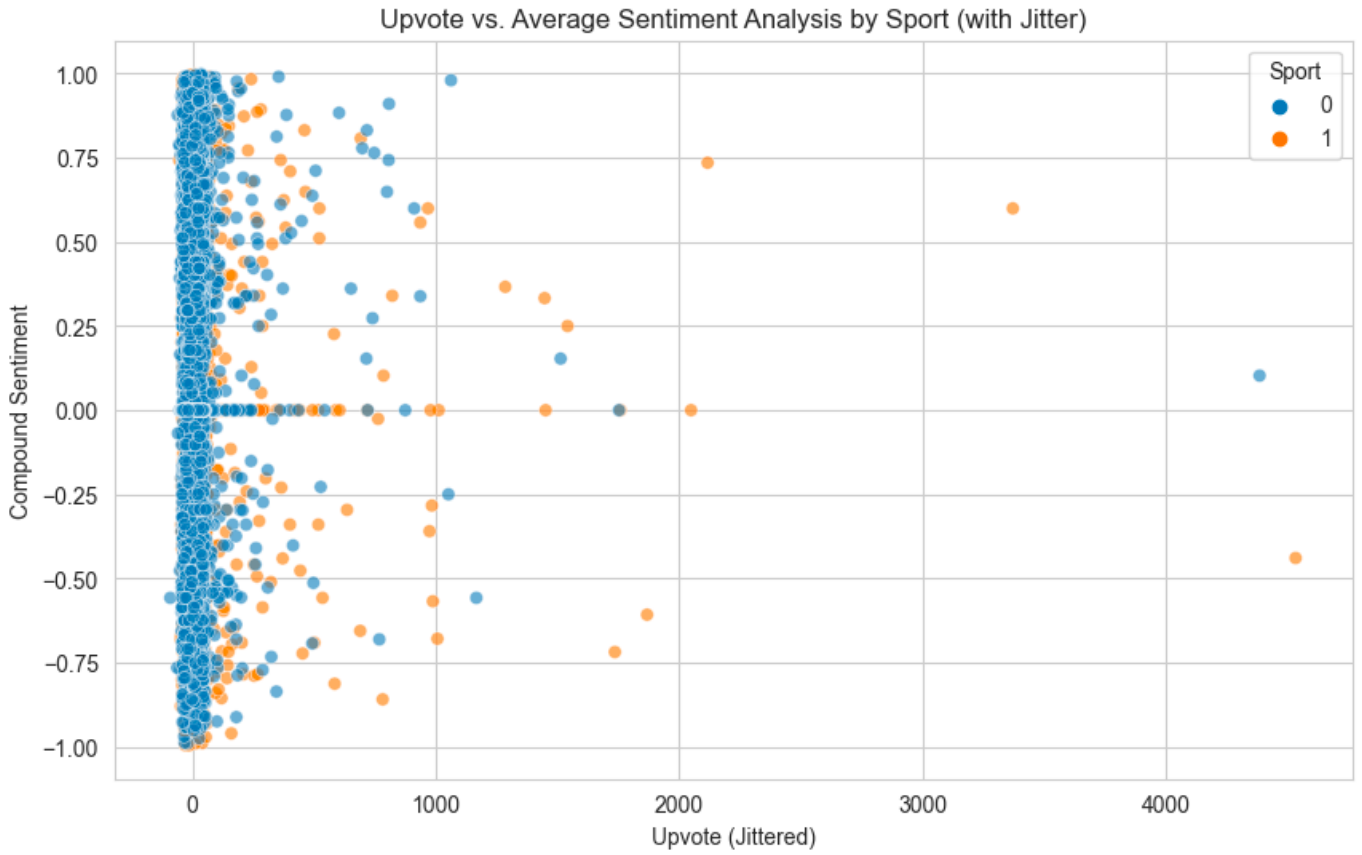
Visualization 2



Scenario	Sentiment Comparison Across Sports
1F: Favored team winning a close match (Favored team subreddit)	Relatively the same across sports
1U: Favored team winning a close match (Underdog team subreddit)	Underdog soccer fans' comments more positive than football
2F: Favored team winning a one-sided match (Favored team subreddit)	Relatively the same across sports
2U: Favored team winning a one-sided match (Underdog team subreddit)	Underdog soccer fans' comments more positive than football
3F: Underdog team winning a close match (Favored team subreddit)	Relatively the same across sports
3U: Underdog team winning a close match (Underdog team subreddit)	Underdog soccer fans' comments more negative than football
4F: Underdog team winning a one-sided match (Favored team subreddit)	Favored soccer fans' comments more positive than football
4U: Underdog team winning a one-sided match (Underdog team subreddit)	Underdog football fans' comments more positive than soccer

Visualization 3

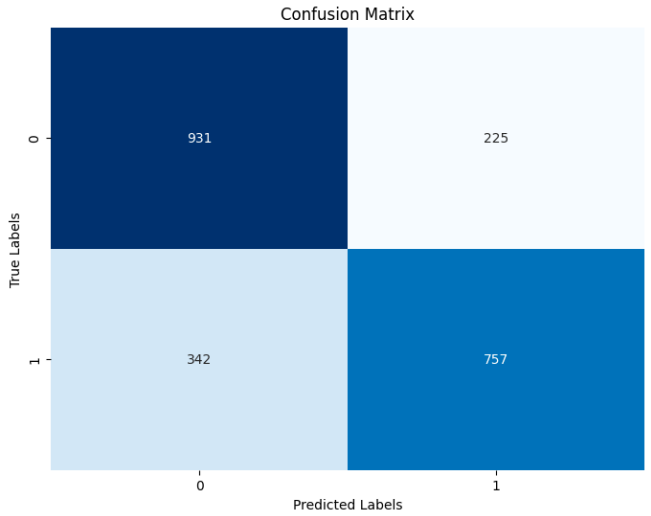
Visualization 4



Predictive Modeling Performance

- Model Accuracy: Approx. 75%.
- Precision, Recall, and F1 Score discussion.
- Feature Importance: Phase, upvote, category, and sentiment analysis score.

```
Accuracy: 0.7485587583148559
Precision: 0.7708757637474541
Recall: 0.6888080072793449
F1: 0.7275348390197021
```



Feature Importance

	importance
upvote	0.231281
phase	0.332900
compound_sentiment	0.084354

Key Takeaways & Future Directions

- Sentiment analysis reveals nuanced fan reactions across sports and scenarios.
- Predictive modeling shows potential for identifying sport based on sentiment.
- Implications for fan engagement and content strategy.
- Limitations and areas for future research: Language diversity, platform expansion.

Questions & Discussion

Citations

- https://www.linkedin.com/pulse/game-changing-power-text-annotation-sports-enhancing-performance-unsef/
- https://link.springer.com/article/10.1007/s42488-021-00052-4
- https://aisel.aisnet.org/treos_amcis2020/11/
- https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1010&context=treos_amcis2020