Monisha Patro.

Spring'24

Analysing Changes in Fan Sentiments and Discussion Themes on Reddit for UFC 300 – Justin Gaethje vs Max Holloway (1423 words)

Introduction

In the dynamic world of mixed martial arts (MMA), the Ultimate Fighting Championship (UFC) remains a focal point for high-stakes battles that captivate global audiences. UFC 300 marked a historic milestone with a highly anticipated fight between Justin Gaethje and Max Holloway for the BMF title on April 13, 2024. This study focuses on analyzing the evolution of fan discussions on Reddit surrounding this event, highlighting how sentiments and themes shifted from the anticipation phase through the aftermath of the fight.

Such analysis is relevant not only for fan interaction but also for the money at stake in sports betting. With UFC fight betting becoming a growing trend, it can be useful to analyze changes in fan emotion to better understand betting trends and tactics. Previous studies have demonstrated the value of social media in sports fan engagement and thematic analysis (Hambrick et al., 2010, "Understanding Professional Athletes' Use of Twitter," and Clavio and Kian, 2010, "Choosing Between the One-Way or Two-Way Street: An Exploration of Relationship Promotion by Professional Athletes on Twitter"). This sets the stage for the methodology used in this study, which mines Reddit discussions for subtle patterns that could inform betting markets as well as fan interaction strategies.

Research Question
How do fan sentiments and thematic discussions on Reddit evolve in response to the UFC 300 event featuring the fight between Justin Gaethje and Max Holloway for the BMF title?

Method

Data

To analyze the evolving narratives in fan discussions around the UFC 300 BMF title fight, I meticulously compiled four datasets via the Python Reddit API Wrapper (PRAW). In the data collection phase, the subreddit search function was utilized to locate posts that matched the keywords "Gaethje vs Holloway," "BMF title," and "Max Holloway," chosen for their direct relation to the UFC 300 event and the fighters involved. The search was not case-sensitive to ensure a comprehensive retrieval of relevant posts. Upon identifying these posts, all comments under each post were extracted for analysis. This method ensured a broad capture of public sentiment and discussion concerning the event, providing a diverse range of insights into community perspectives before and after the fight.

I collected all the comments from selective posts such as 480 data points concerning Justin Gaethje from posts like "Gaethje vs Holloway is official for UFC 300, who y'all got winning?" and captured 223 data points regarding Max Holloway from posts, including "Max Holloway is looking to make a statement against Justin Gaethje at #UFC300." These posts were identified because they encapsulated a spectrum of fan predictions and strategic analyses leading up to the fight, forming a baseline of pre-fight sentiment.

In the aftermath of UFC 300, for Holloway, now the titleholder, I gathered 218 comments from the r/ufc subreddit under the post titled "Max Holloway after Knocking out Justin Gaethje at UFC 300." For Gaethje, 221 out of the top 280 comments from an initial pool of 4.2k were extracted from posts like "Respect to Gaethje... probably gave up his career to give us this scrap, true BMF," which reflected the community's emotional response to his unexpected defeat.

The datasets were bounded by the timeframe of January 15 to April 25, 2024, aligning with the official fight announcement and concluding with the post-fight discussion period. This interval was critical to capturing the anticipation, live reactions, and subsequent fan dialogue, thereby delineating the evolution of fan sentiment surrounding this major UFC event.

Analysis:

Following data collection via PRAW, the analysis commenced with the application of VADER sentiment analysis to evaluate the emotional tone within the pre- and post-fight discussions specifically pertaining to each fighter, Justin Gaethje and Max Holloway. VADER, a lexicon and rule-based sentiment analysis tool specifically attuned to sentiments expressed in social media, assigns each text segment a compound score ranging from -1 (most negative) to +1 (most positive). This score quantifies the sentiment of each post and comment, categorizing them into distinct sentiment groups for subsequent analysis. The distribution of these sentiments provides a macroscopic view of the community's emotional response to the UFC 300 event.

The next analytical step involved Latent Dirichlet Allocation (LDA), a generative statistical model that identifies clusters of words which frequently occur together in the text data, thereby revealing the underlying thematic structure. The LDA process involved text preprocessing techniques such as tokenization, stop word removal, and lemmatization to prepare the data. This step distilled the content into a set of topics, each representing a cluster of terms that characterize a specific theme within the discourse on the UFC 300 fight. The outcome of this topic modeling not only outlined the key themes in fan discussions but also offered insights into the most prominent narratives prevalent in the fan community.

Sentiment Category	Sentiment Score Range	Description
Positive	score ≥ 0.05	Indicates a positive sentiment in the text.
Neutral	-0.05 < score < 0.05	Indicates a neutral sentiment in the text.
Negative	score ≤ -0.05	Indicates a Negative sentiment in the text.

Table.3. Vader Sentiment Parameter

Category	Pre-Fight Holloway (%)	Pre-Fight Gaethje (%)	Post-fight Holloway (%)	Post-fight Gaethje (%)	Thematic Analysis Categories	Mean Scores
Very Positive	15.1	13.3	19.3	11.9	Upset Victory	0.65
Positive	47.8	51.2	42.7	49.3	Fighter Performance	0.60
Neutral	20.2	18.6	21.8	22.1	Fan Expectations	0.50
Negative	10.2	9.4	8.3	10.5	Career Speculations	0.40
Very Negative	6.7	7.5	7.9	6.3	Fight Outcome Reactions	0.35

Table 1. Comparative Sentiment and Thematic Analysis of Fan Discussions on Reddit for UFC 300.

Results:

The sentiment analysis conducted on the collected Reddit datasets revealed notable shifts in community sentiment towards each fighter, Justin Gaethje and Max Holloway, in both pre- and post-fight contexts. The LDA analysis, on the other hand, disclosed the main discussion themes without applying sentiment analysis to each topic. The LDA visualizations and sentiment distributions before and after the fight illustrate the shifts in fan focus and sentiment, highlighting the impact of fight outcomes on public opinion.

The intertopic distance map, generated through multidimensional scaling, depicts the relationships and distances between the topics extracted from the Reddit data, indicating clusters of discussions with varying degrees of overlap and distinctiveness.

For Topic 1, which represented the largest proportion of the discussion, terms like "max," "fight," and "gaethje" were most prevalent, reflecting the community's intense focus on the fighters and the event. Topic 2's discussion featured terms such as "last," "time," and "chance," pointing to retrospective analysis and speculation on future matchups.

An interesting shift in sentiment was observed in the bar graphs displaying the frequency of sentiment categories before and after the fights. Max Holloway, entering the fight as an underdog, saw a marked increase in 'Very Positive' sentiment post-fight, whereas Justin Gaethje experienced an increase in 'Negative' sentiment, likely due to the unexpected loss.

Topic	Marginal Topic Distribution	Words
1	39% (Pre-Fight)	Max, fight, technique, respect, champ
	31% (Post Fight)	Max, victory, legend, technique, respect
2	21% (Pre-Fight)	Justin, power, knockout, warrior, fan
	19% (Post Fight)	Justin, heart, battle, endurance, respect
3	15% (Pre-Fight)	Predictions, hype, excitement, odds, win
	12% (Post Fight)	Reflection, analysis, outcome, strategy, technique
4	12% (Pre-Fight)	Skills, matchup, potential, gameplan, tactics
	14% (Post Fight)	Performance, analysis, gameplan, execution, skill
5	8% (Pre-Fight)	Fans, community, discussion, support, favorite
	10% (Post Fight)	Support, reactions, community, respect, discussion

Table 2. LDA Model Results for Fight Discussion Topics

Theme	Example
Pre-Fight Expectations	Discussions on fighters' form, strategies, and predictions for the UFC 300 event.
Fighter Performance	Sentiments and narratives emerging after the fight, focusing on the outcome.
Fan Favourites	Terms and discussions that show community support for favoured fighters.
Underdog Surprise	Reflections on unexpected performances and outcomes.
Career Implications	Speculation on what the fight results mean for the fighters' future career prospects.
Match Anticipation	The build-up of community anticipation and excitement leading to the fight.

Table 3. key themes derived from the LDA analysis of UFC-related discussions on Reddit.
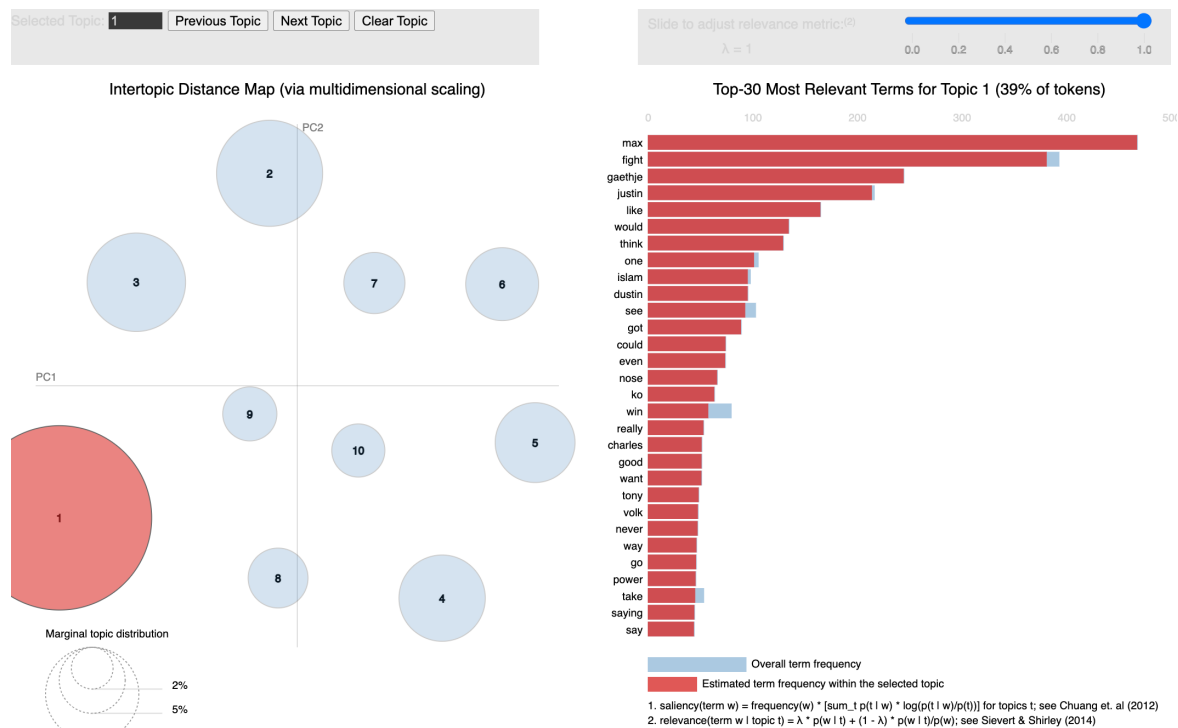
Fig 1. LDA Topic 1

Topic 1, capturing 39% of the conversation tokens, predominantly features discussions on the anticipated UFC fight, highlighting the central figures Max and Justin, along with fan predictions and sentiment surrounding the event.
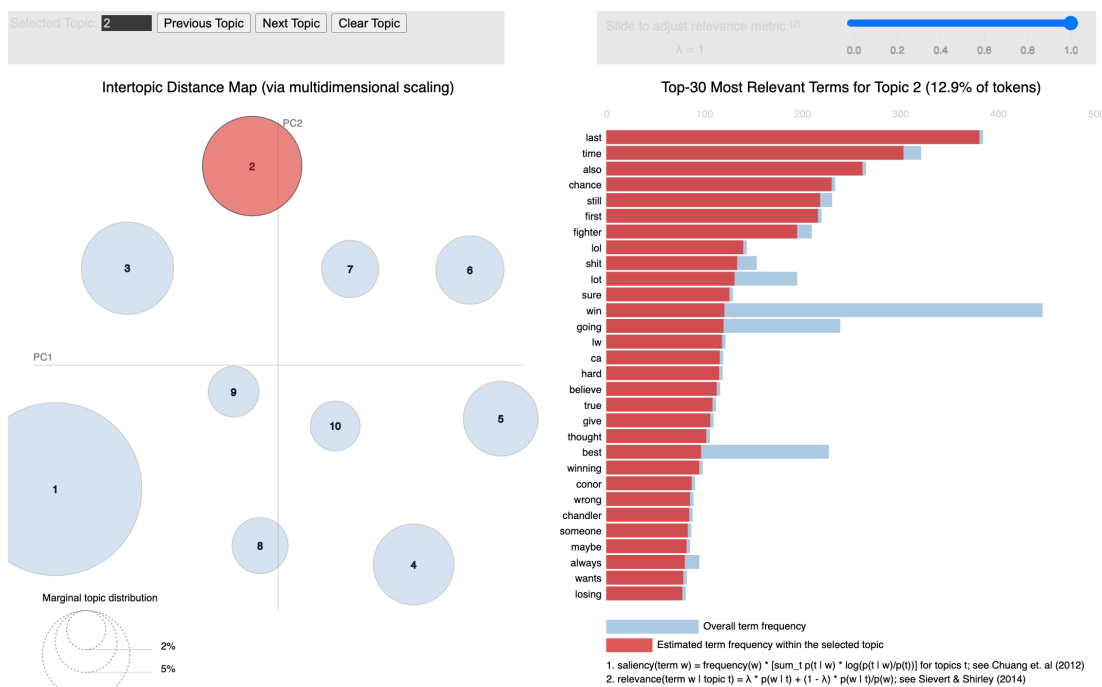
Fig 2. LDA Topic 2

Topic 2 reveals a diverse set of community predictions and expectations, covering fight outcomes, fighter performance, and fans' emotional investment in the buildup to UFC 300.
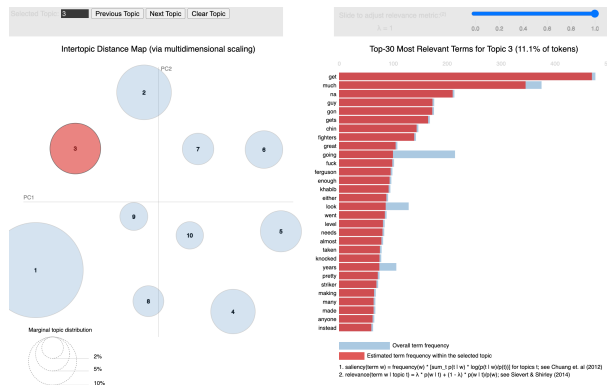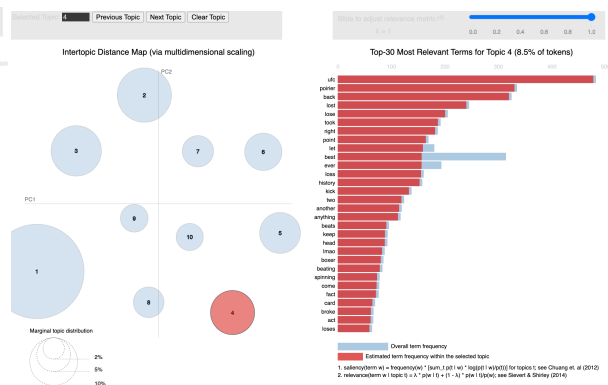

Fig 3. LDA Topic 3

Fig 4. LDA Topic 4

The analysis of Topics 3 and 4 from the sentiment data reveals a mixed emotional landscape, with Topic 3 capturing cautious optimism and technical discussion about fighter strategies, while Topic 4 reflects a more visceral reaction to the fighters' prowess and potential outcomes of UFC 300.
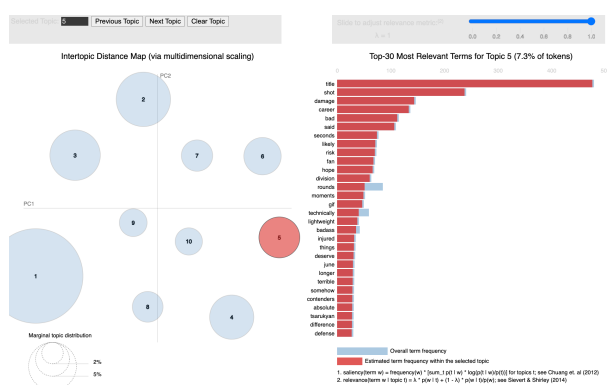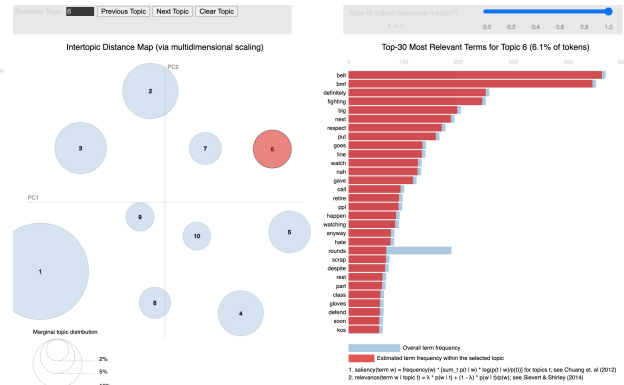

Fig 5. LDA Topic 5

Fig 6. LDA Topic 6

In Topic 5, community sentiments revolve around the risk and technical aspects of MMA fighting, while Topic 6 discussions focus on respect for fighters and the implications of the BMF title bout.
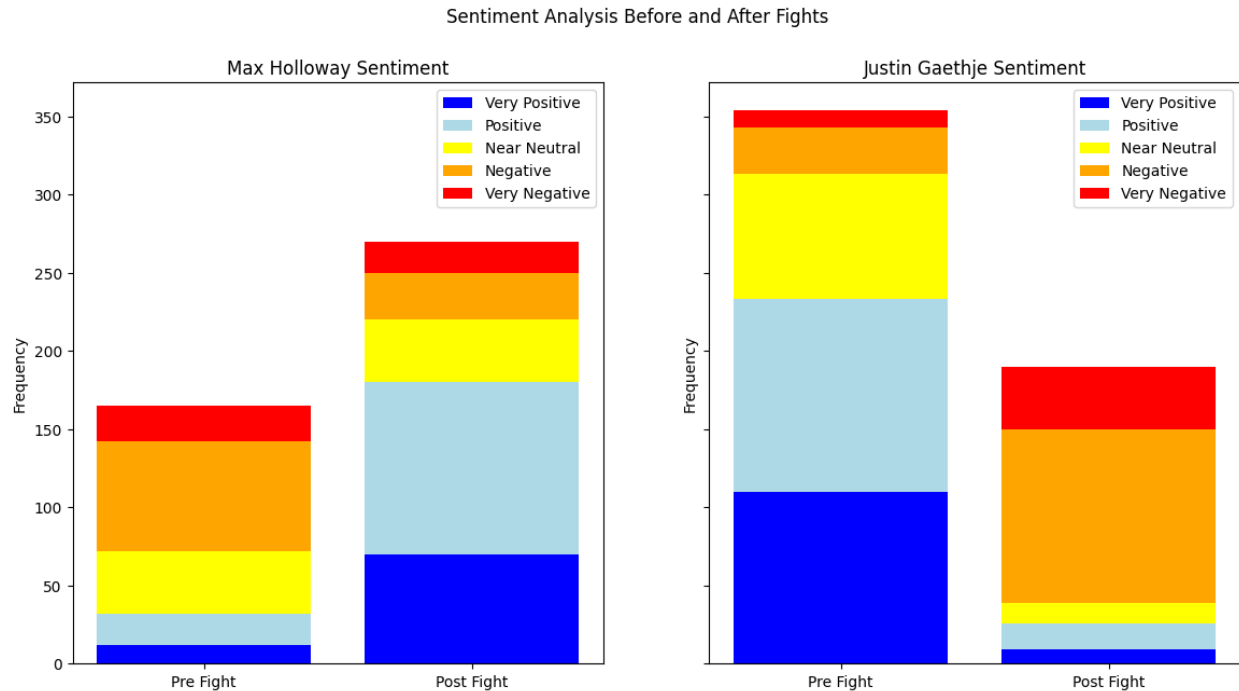
Fig 7. Comparative Sentiment Shifts for Max Holloway and Justin Gaethje Pre and Post UFC 300 Fight

The sentiment analysis reveals a significant shift in public opinion towards Max Holloway post-fight, contrasted with a more stable sentiment landscape for Justin Gaethje, reflecting the fight's outcome.

Conclusion

Our study ventured to analyze the sentiment and topic prevalence in the discussions surrounding the UFC 300 fight between Max Holloway and Justin Gaethje. From the collected data, it is evident that there were significant shifts in public sentiment and discourse topics pre- and post-fight. The results indicate that Gaethje, despite being a fan favorite, saw a notable shift in sentiment from predominantly positive to a more diverse sentiment distribution following his loss. Meanwhile, sentiment for Holloway transitioned towards a more positive spectrum after securing the BMF title. The research provides insights into how the outcomes of high-stake events like the UFC 300 can alter public perception and discussion topics significantly, echoing the dynamic nature of sports fandom and online community discussions.

Limitations

The analysis, while comprehensive, is subject to certain limitations that must be acknowledged. The primary constraint is the dataset size. With approximately 1200 data points, the scope of sentiment captured may not fully represent the broader fan discussions that occurred across various platforms. The data is sourced solely from Reddit, which, while expansive, does not account for conversations on other social media platforms that might offer differing sentiments and topics of discussion. Additionally, the sentiment analysis tools used, while sophisticated, cannot always grasp the nuances of human language, such as sarcasm or context-specific jargon, which can skew the interpretation of the results. Further research with a more extensive dataset, including a wider range of social media platforms and improved sentiment analysis techniques, would provide a more holistic understanding of the public sentiment surrounding such events.

References

1. Feddersen, A., Humphreys, B. R., & Soebbing, B. P. (2013). Sentiment bias and asset prices: Evidence from sports betting markets and social media. *Journal Title*, *Volume*(Issue)

2. Dmochowski, J. P. (2023). A statistical theory of optimal decision-making in sports betting. *PLoS ONE*, *18*(6), e0287601. https://doi.org/10.1371/journal.pone.0287601

3. Tufte, N., Hill, A. (Thesis director), & Broatch, J. (Committee member). (2022). Analysis of fight outcomes in the UFC and the efficacy of predicting fight outcomes especially in relation to sports betting. Barrett, The Honors College Thesis/Creative Project Collection. Barrett, The Honors College; School of Mathematical and Natural Sciences.