

MONISHA PATRO

Spring 2024

Title: Evolving UFC Fan Sentiments and Themes on Reddit (2022-2024) (487 words)

Research Question:

How do fan sentiments and discussion themes related to UFC fighters evolve on Reddit in the period leading up to and following major fights between 2022 and 2024?

Data:

I will be using Reddit's API to collect data for this study, with a particular focus on subreddits like r/UFC, r/MMA, and r/knockout that are devoted to the UFC and related subjects. Posts will be filtered using key terms like "brutal" and "undefeated" to capture the most active debates around big fights and fighters, ensuring the relevancy and specificity of the data. The data covers a large number of fan discussions and runs from January 2022 to March 2024. Eight weight classes will be used to segment the data, and prominent competitors from each class will be chosen, such as Ilia Topuria, the Men's Featherweight class champion. About 120 data points will be gathered for each chosen fighter's fight card, split equally between pre- and post-fight talks: 60 data points will be gathered prior to the fight decision and another 60 following. This guarantees that we have a wide range of fan viewpoints, both before and after the fights, covering different angles. The data collection will prioritize the top 10 most upvoted posts and, where available, the most commented-on discussions within the fighter-specific subreddits, such as r/Topuria. This method ensures a dataset rich with fan insights and reactions across different weight classes and fight outcomes by enabling an in-depth study of fan attitudes and theme discussions related to UFC fighters.

Method:

To analyze UFC discussions on Reddit, I will employ a combination of computational methods, starting with the Gensim library to perform Latent Dirichlet Allocation (LDA) topic modeling. This technique will help identify 10 to 20 key themes, such as "Fight Predictions," "Fighter Performance," and "Fan Favourites." The initial phase involves cleaning the text and tokenizing it, which means breaking it down into meaningful elements and removing irrelevant words to improve the data's quality. After training the LDA model on UFC fan discussions, I'll fine-tune how it recognizes themes by adjusting parameters, such as the optimal number of topics and word contribution levels, to ensure the themes identified are precise and distinct. In addition to LDA, I'll use the EMPATH tool to analyze sentiment and category themes in the chats in order to capture the wider emotional landscape. This will offer a more profound comprehension of the background and emotion underlying the conversations, such as enthusiasm, disillusionment, or appreciation, which are just as important as the subjects at hand. Using these techniques, my goal is to map out the primary topics of conversation and the feelings they arouse in the UFC community between January 2022 and March 2024. Additionally, pyLDAvis visualizations will be created to provide a clear and understandable visual interpretation of the results, highlighting the main themes and the emotions they convey.

Referenecs:

Řehůřek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45-50.

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63-70.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.