

Name: Mayuri Birari

Semester: Fall 2023

**Quantitative Analysis of Twitter Conversations on the MeToo Movement Using
Sentiment Analysis and Topic Modeling**

(1544)

1. Introduction

The MeToo movement, initiated in 2006 by Tarana Burke, gained widespread attention in recent years. It trended as a hashtag used by survivors and victims of sexual assault and harassment. This movement encouraged victims to come out in the open and share their stories. This led to a surge of personal stories and discussions concerning sexual misconduct, particularly in the workplace. The global impact of the hashtag emphasizes the topic's importance. The impact of the movement on cultural expectations and legal systems and policies makes it an important subject of study. This study intends to use the insights gained through the research of Reyes-Menendez, Saura, and Filipe (2020), as well as Reyes-Menendez, Saura, and Thomas (2020). We do so by evaluating the origin and evolution of the MeToo movement discussion on Twitter. This study will look specifically at the general sentiments and most discussed topics in these Twitter conversations, to shed light on broader societal involvement with the MeToo movement.

2. Research Questions

- What is the general sentiment of Twitter users about the MeToo Movement?
- What are the most discussed MeToo Movement topics on Twitter?

3. Method

3.1 Data

This existing dataset, "Hatred on Twitter During the MeToo Movement," was chosen because it corresponds directly to the MeToo movement, a prominent social campaign against sexual harassment and assault. With 807,175 tweets, this dataset—hosted on Kaggle—offers a sizable amount of data and a solid foundation for quantitative research. These tweets not only align with the MeToo movement but also span a critical period of highest activity and public interest, from September 30, 2018, to February 18, 2019. This period is especially important because it marks the movement's peak and covers a wide variety of public opinion and conversation. The dataset's time frame was further divided into two unique periods within the limits in order to do a thorough sentiment analysis.

1. Early stages of the MeToo movement - September 30 to December 31, 2018
2. Peak of the MeToo movement - January 1 to February 18, 2019

Rather than creating a new dataset, the "Hatred on Twitter During the MeToo Movement" dataset was chosen because of its extensiveness. It captured a wide range of Twitter conversations. It provides insights into the nature of engagement, including favorites, retweets, and other user interactions. The study seeks to provide answers to the research questions posed during a period when it was very significant using topic modeling and sentiment analysis on this dataset.

3.2 Analysis

In this study, two different approaches to data analysis were employed: sentiment analysis using EMPATH and topic modeling using LDA (Latent Dirichlet Allocation). These processes are applied to data from both timeframes. The initial step in sentiment analysis was importing all necessary packages. The data was then timestamped from UTC. Sentiment analysis was performed using empathy, and scores for both positive and negative emotions were calculated for each comment.

The total positive and negative ratings for each period are then computed to get the overall emotions experienced during the period (refer to Figures 1 and 2). The graphs demonstrate that negative emotion had increased and positive emotion had decreased between the early stages of the MeToo movement and the peak of the MeToo movement. Graphs were plotted to show the trend of the positive and negative scores over different time periods, and the results show that positive emotion is consistently lower than negative emotion in both cases. Refer to Figures 3 and 8. However, as the scores indicated, negative emotion has increased and positive emotion has decreased as the time periods have changed.

The Natural Language Toolkit (NLTK) stop words approach was used to eliminate punctuation and excessively common words from the corpus during the data preprocessing stage for topic modeling. Next, word clouds, frequency tables, and bar graphs were used to visually represent the most frequently occurring terms in the corpus, both before and after stop words were eliminated (refer to Figures 9 and 10). Consequently, the topic word distribution is apparent (refer to Figures 13 and 14). To find the number of topics, the topic coherence score was computed after the model was trained using genism to produce a bag-of-words. Four topics were developed for every instance of LDA use. The words and their associated weights are displayed in Figures 15 and 16.

Common terms associated with the early MeToo movement were discovered through topic modeling, including "sexual," "amp," "women," "victim," "assault," "men," and "survivor." Topic 0 focused on gender dynamics and possibly social interactions or attitudes, with words like "amp," "women," "men," and "thank" appearing frequently. Topic 1 deals with sexual assault and related social issues, using words like "sexual," "assault," "know," "tell," "ford," and "woman," whereas Topic 2 deals with the movement, using words like "movement" as well as "make," "kavanaugh," "want," and "founder," which specifically refers to discussions about the MeToo movement. Topic 3 appears to deal with

beliefs, opinions, and possibly personal narratives, using words such as "get," "go," "women," "believe," "think," "say," "would," "one," "use," and "women."

Overall, in our analysis Topic 0 looks at how men and women interact, with a focus on trust and feelings. Topic 1 is about sexual violence, talking about awareness and how society deals with it. Topic 2 discusses a social or political movement, focusing on personal stories and feelings related to gender issues. Topic 3 is about personal views and stories on gender, covering a wide range of talks on experiences and changes in society about gender. We used LDA to identify common words among topics, which helps in understanding how these topics are related. For instance, words like "gender" and "movement" appear frequently across different topics, it suggests a connection between discussions on gender dynamics and the social movement.

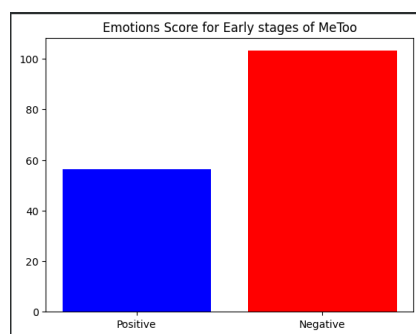


Figure 1: Emotion scores for early stages of MeToo

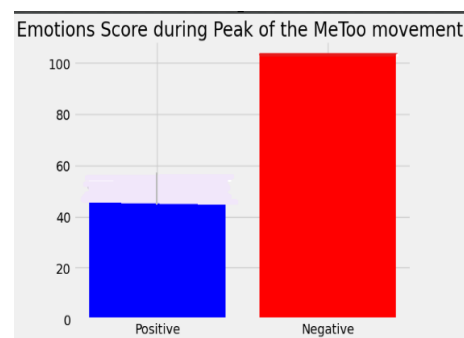


Figure 2: Emotion scores for the peak of the meToo movement

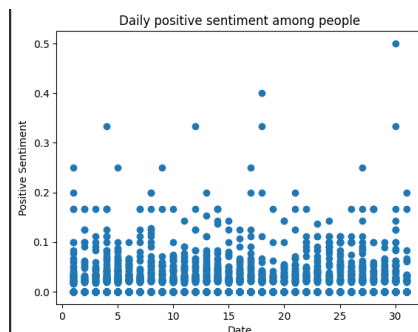


Figure 3 Positive Sentiment Vs Day graph early Metoo movement.

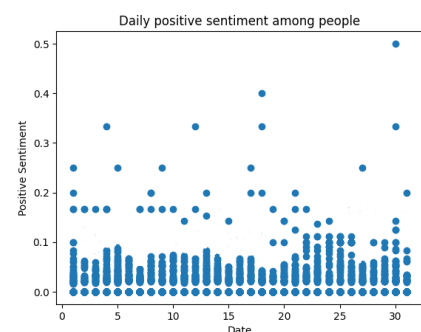


Figure 4 Positive Sentiment Vs Day graph peak Metoo movement

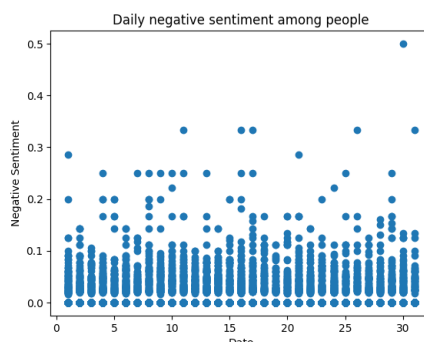


Figure 5 Negative Sentiment Vs Day graph early Metoo movement.

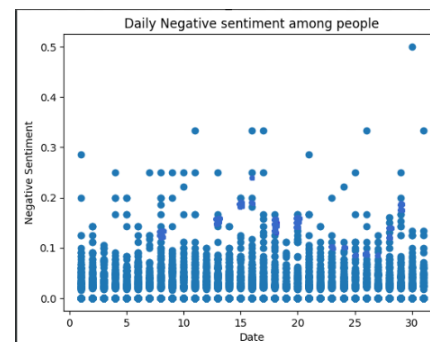


Figure 6 Negative Sentiment Vs Day graph peak Metoo movement.

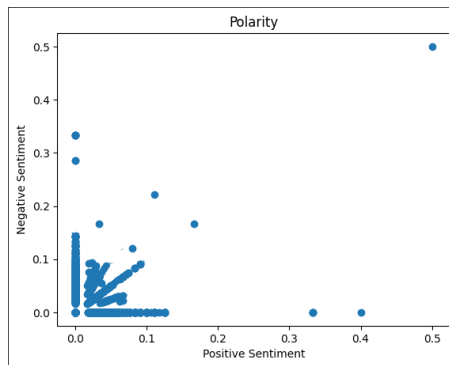


Figure 7 Positive Vs Negative Sentiment for Early stages of MeToo.

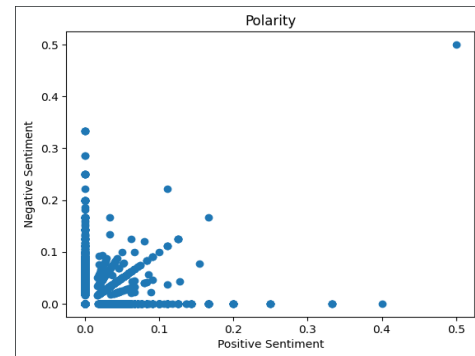


Figure 6 Positive Vs Negative Sentiment for peak of MeToo.

Topic Modelling Outcomes:

The outcomes indicate the most used words during both periods.

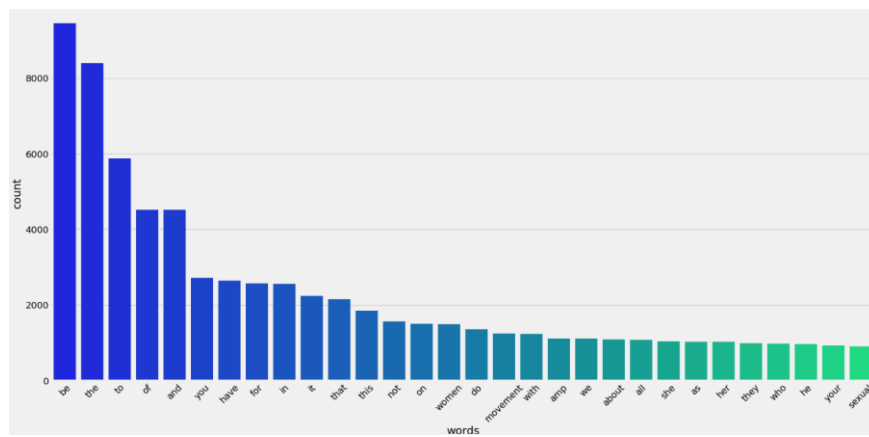


Figure 9 Word Vs Count for Early stages of Metoo movement

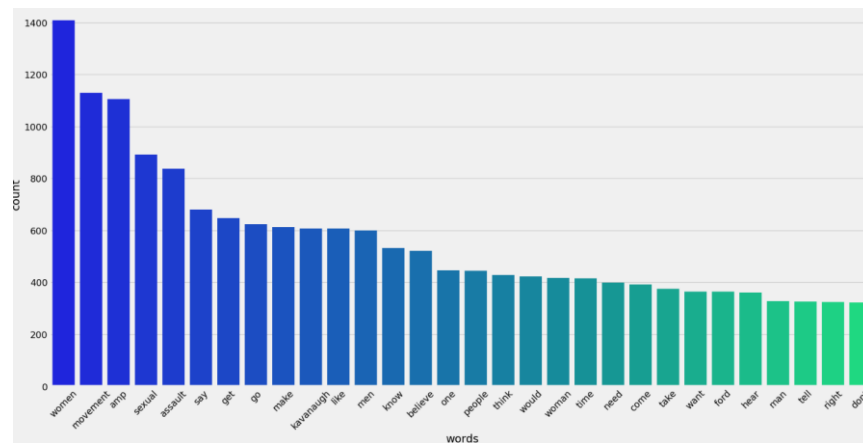


Figure 10 Word Vs Count for peak stage of Metoo movement

Word Clouds for both period:

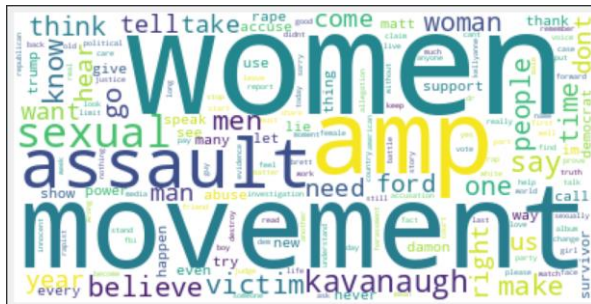


Figure 11 Word Cloud early Metoo movement.



Figure 12 Word Cloud peak Metoo movement.

The Topic-word distribution for each topic in both periods:

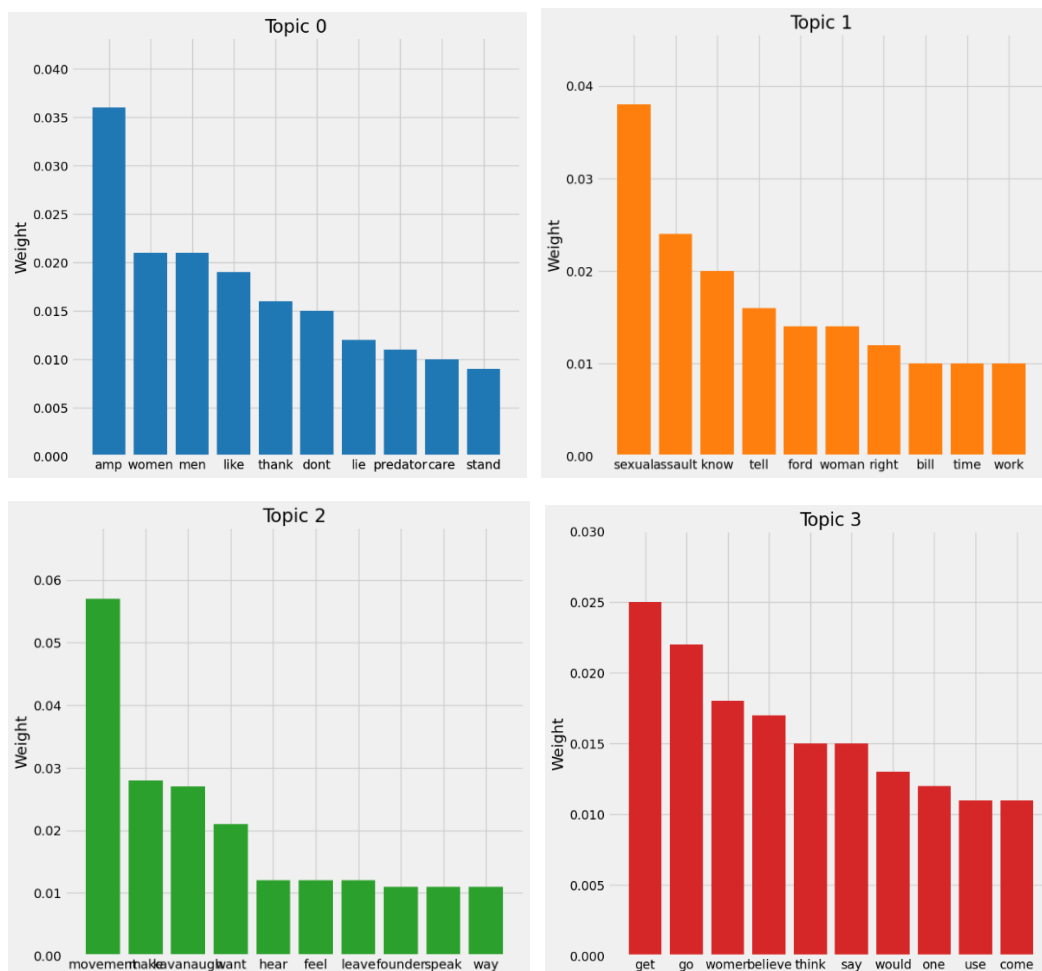


Figure 15 Word-Weight for each Topic early Metoo movement

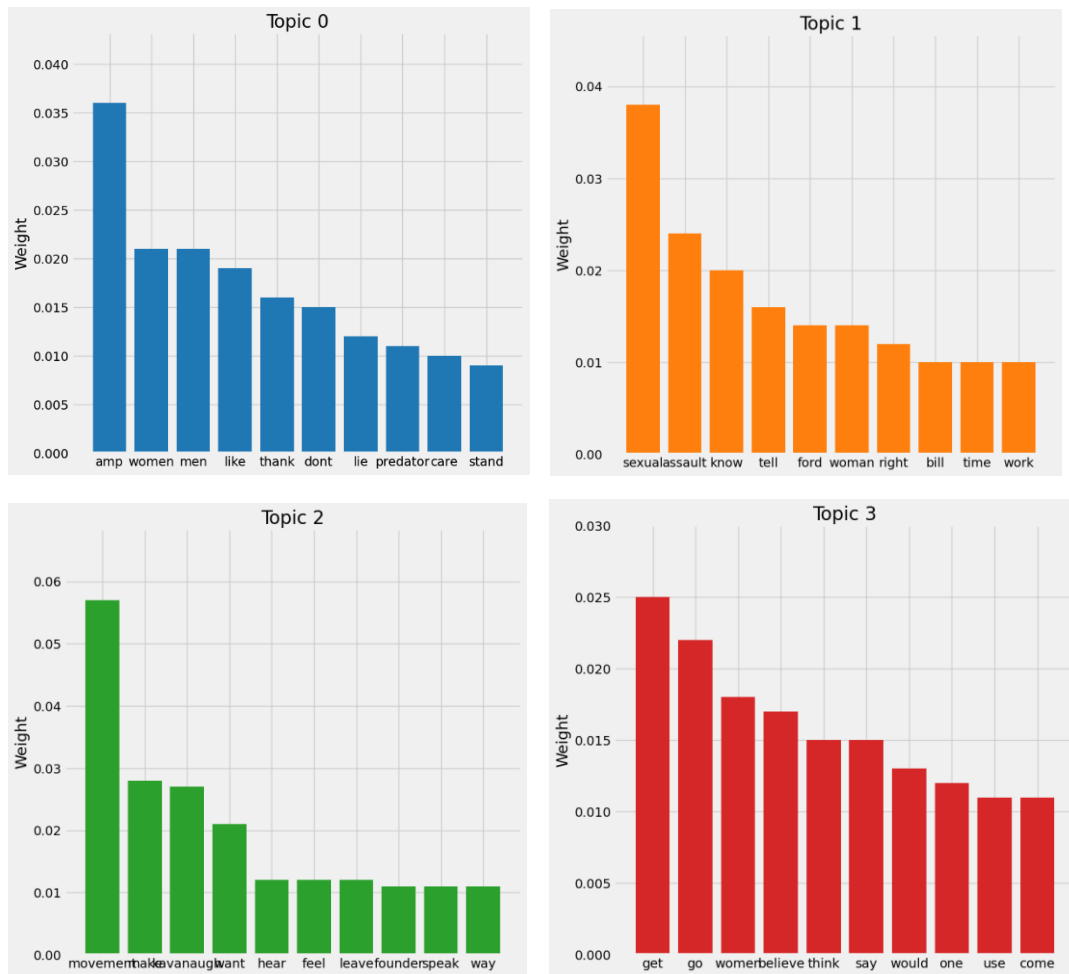


Figure 16 Word-Weight for each Topic peak Metoo movement

4. Conclusion and Limitations

Our study investigated an extensive collection of Twitter conversations from the MeToo movement and analyzed shifts in public sentiments about the same. The findings show a significant shift in the general attitude toward these sentiments. We noticed an increase in negative sentiments and a decrease in positive sentiments as the movement gained popularity. This shift in sentiment can be linked to increased awareness of sexual violence and harassment issues. The study found four major topics: gender dynamics, sexual violence awareness, the impact of a gender-focused social movement, and personal beliefs and narratives about gender issues. These findings indicate that the MeToo movement has been crucial in bringing critical discussions about gender inequality, sexual violence, and harassment to the forefront. In addition to drawing attention to these problems, the movement has helped foster a larger public discussion on the dynamics of gender. These results underlined how crucial systemic change is.

There are certain limitations to the study that need to be considered. Firstly, because the analysis was restricted to Twitter, the viewpoints gathered might not accurately reflect those of the general public.

The results may be skewed by the unique demographics and behaviors of Twitter users. Second, it's possible that some of the text's nuances, like sarcasm or tone inconsistencies, were not fully captured by the sentiment analysis techniques used, which could have led to misinterpretations of the actual sentiments expressed. Because the study relies primarily on Twitter data, it may not fully capture every relevant fact or the range of conversations around the MeToo movement.

5. References

Goel, R. (n.d.). Hatred on Twitter During MeToo Movement [Data set]. Kaggle. Retrieved from <https://www.kaggle.com/datasets/rahulgoel1106/hatred-on-twitter-during-metoo-movement>

Reyes-Menendez, A., Saura, J. R., & Filipe, F. (2020). Marketing challenges in the #MeToo era: Gaining business insights using an exploratory sentiment analysis. *Heliyon*, 6(4), e03747. doi:10.1016/j.heliyon.2020.e03747. Retrieved from

<https://pubmed.ncbi.nlm.nih.gov/32258475/#:~:text=,Rey%20Juan%20Carlos%20University%2C%20Spain>

Reyes-Menendez, A., Saura, J. R., & Thomas, S. B. (2020). Exploring key indicators of social identity in the #MeToo era: Using discourse analysis in UGC. *International Journal of Information Management*, 54, 102129. Retrieved from

<https://doi.org/10.1016/j.ijinfomgt.2020.102129>