Ramitha V 22070126082

Sakshi Sah 22070126096

Samriddhi Kumari 22070126100

AIML - B1

#MeToo Movement Twitter Sentiment Analysis

1. Introduction

Social Media such as Twitter plays a vital role in the globalization of issues and it acts as a catalyst for uniting people for a cause. The MeToo movement is a prime example for this. Twitter played a crucial role in providing a platform for survivors of sexual assault and harrasment to speak up about their experiences as well as finding community and empowerment in others as well as themselves, inspiring many other survivors to share their stories. This has allowed voices that had been silenced for too long to be heard by not just the people they knew but the entire world.

Analyzing this movement will help us grasp the true essence of this movement and what it stands for. The sheer volume of tweets and the diversity of voices revealed the extent of the problem. It provided a data-rich environment for researchers and activists to study patterns, trends, and the public's reactions, shedding light on the nuances of the movement.

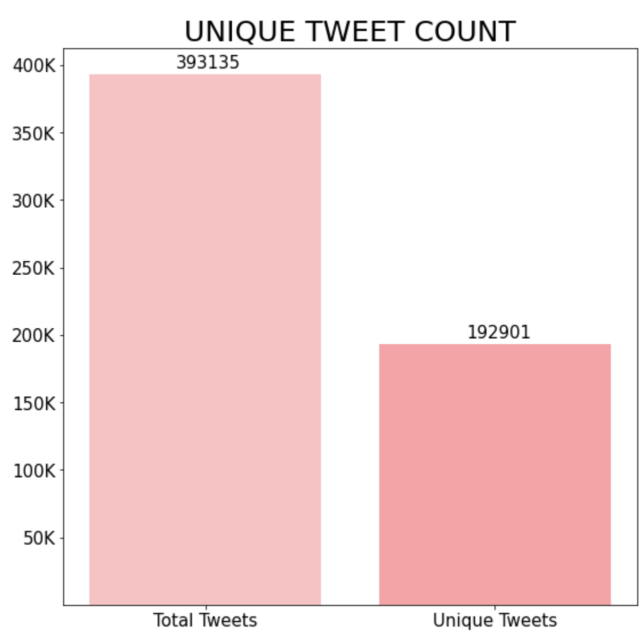
The #MeToo movement lacks a thorough and systematic analysis of its multifaceted dimensions. There is an imperative to conduct an in-depth evaluation that encompasses public reactions, assesses the movement's influence on incidents of sexual harassment and assault, and monitors the engagement of key figures within the movement. This analysis seeks to address the need for a comprehensive understanding of the movement's impact, challenges, and key contributors.

2. Problem statement

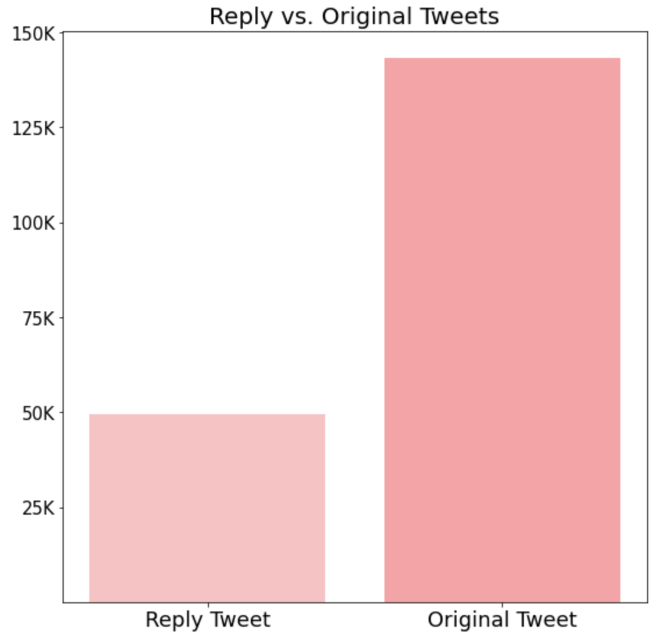
The problem we aim to address is to conduct sentiment analysis on #MeToo movement twitter data to evaluate the reaction of the public towards the movement as well as to gauge the media's portrayal of the movement, to identify the key issues in the movement that needs addressing , to analyze the variation of the movement across different regions of the world, and also to monitor the involvement of key individuals.

3. Dataset Description

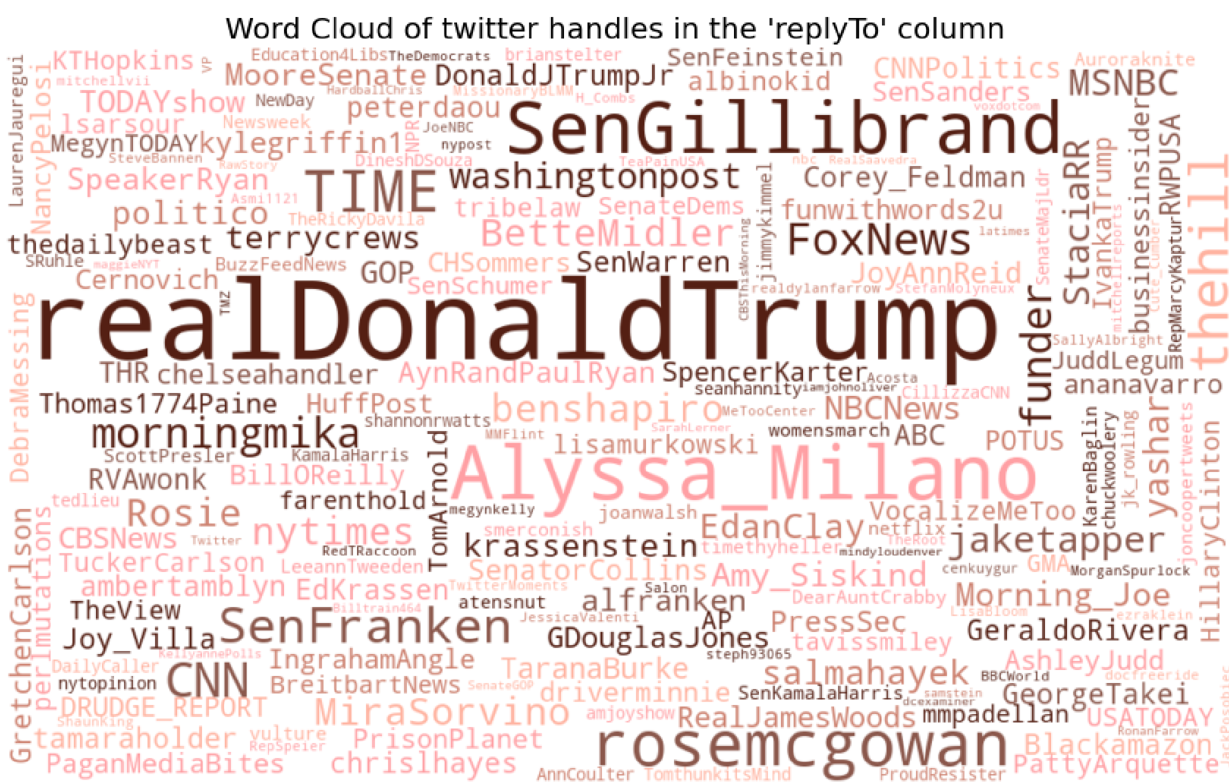
- The dataset includes text content of tweets posted between November 29th and December 25th, 2017.
- Each tweet is associated with metadata, including the timestamp, user information (username), and engagement metrics.
- The data may contain multimedia elements such as links and hashtags shared in tweets.
- Dataset is completely raw data scraped from Twitter by the author with no preprocessing done.



From this graph we can infer that the dataset we have is data-rich and has a multitude of opinions from different people



From this graph we can see that from the entire dataset, most tweets are original and we lose context for less amounts of tweets that are in reply to a different tweet.



This word cloud helps us in visualizing the comparison of the number of tweets that were posted in reply to public figures.

```
Most Replied-To People:
realDonaldTrump: 2062 replies
Alyssa_Milano: 746 replies
SenGillibrand: 662 replies
rosemcgowan: 511 replies
TIME: 435 replies
thehill: 374 replies
SenFranken: 313 replies
CNN: 308 replies
FoxNews: 271 replies
morningmika: 225 replies
```

Here we can see the key people who were highlighted in this movement as they were the most replied to.

4. Justification.

4.1 Business Prospects of out project:

- Brand Alignment and Advocacy: Understanding the sentiment around movements like #MeToo enables companies to align their brand values with prevailing public sentiments. They can identify opportunities to support social causes and engage in advocacy initiatives that resonate with their customers.
- Influencer Marketing and Partnership Opportunities: For companies, monitoring the involvement of key figures within movements like #MeToo can inform influencer marketing strategies. They can identify influential advocates and partner with them to amplify their brand message and align with social causes.
- Corporate Responsibility Reporting: Many businesses produce corporate social responsibility (CSR) reports. Sentiment analysis can provide data and insights to support these reports, showcasing a company's commitment to social issues like combating sexual harassment and assault.
- Risk Mitigation: By monitoring sentiment, businesses can identify potential risks associated with social issues. They can take proactive measures to mitigate these risks and prevent adverse impacts on their brand, financial performance, and stakeholders.

4.2 Literary Review:

[1] Sentiment analysis in comments, critiques, or other feedback offers helpful markers for a variety of applications. These feelings can be divided into two groups: positive and negative, or they can be ranked on an n-point scale with options like very good, good, satisfactory, bad, and very bad. Businesses can use sentiment analysis to assess how well their products are received and to identify ways to raise their level of quality. Politicians and policy makers can also use it to analyse public opinion on problems related to politics, public services, and policies.
The actual findings of a comparative research comparing the performance of several classifiers are presented in this work, which demonstrates how using many classifiers in a hybrid fashion can increase sentiment analysis's efficacy. The procedure is that if one classifier fails to classify a document, the classifier will pass the document onto the next classifier, until the document is classified or no other classifier exists.

[2]Topic models are a family of algorithms that allow the analysis of unlabelled large collections of documents in order to discover and identify hidden topic patterns in the form of a cluster of words. In the present paper we aim at topic modeling the CompWHoB Corpus (Esposito et al., 2015), a political corpus collecting the transcripts of the White House Press Briefings.

[3] Methods for document clustering and topic modeling in online social networks (OSNs) offer a means of categorizing, annotating and making sense of large volumes of user generated content. This paper benchmarks four different feature representations derived from term-frequency inverse-document-frequency (tf-idf) matrices and word embedding models combined with four clustering methods, and includes a Latent Dirichlet Allocation topic model for comparison. Several different evaluation measures are used in the literature, so this paper provides a discussion and recommendations for the most appropriate extrinsic measures for this task. It also demonstrates the performance of the methods over data sets with different document lengths.

[4] The aim of this study is to document, characterize, and quantify early public discourse and conversation of the #MeToo movement from Twitter data in the United States. We focus on posts with public first-person revelations of sexual assault/abuse and early life experiences of such events.
We purchased full tweets and associated metadata from the Twitter Premium application programming interface between October 14 and 21, 2017 (ie, the first week of the movement). We examined the content of novel English language tweets with the phrase “MeToo” from within the United States (N=11,935). We used machine learning methods, least absolute shrinkage and selection operator regression, and support vector machine

models to summarize and classify the content of individual tweets with revelations of sexual assault and abuse and early life experiences of sexual assault and abuse.

[5] The present project collected real-time tweets from U.S. soccer fans during five 2014 FIFA World Cup games (three games between the U.S. team and another opponent and two games between other teams) using Twitter search API. We used sentiment analysis to examine U.S. soccer fans' emotional responses in their tweets, particularly, the emotional changes after goals (either own or the opponent's). We found that during the matches that the U.S. team played, fear and anger were the most common negative emotions and in general, increased when the opponent team scored and decreased when the U.S. team scored. Anticipation and joy were also generally consistent with the goal results and the associated circumstances during the games. Furthermore, we found that during the matches between other teams, U.S. tweets showed more joy and anticipation than negative emotions (e.g., anger and fear) and that the patterns in response to goal or loss were unclear. This project revealed that sports fans use Twitter for emotional purposes and that the big data approach to analyze sports fans' sentiment showed results generally consistent with the predictions of the disposition theory when the fanship was clear and showed good predictive validity.

[6] Recent technological developments have created novel opportunities for analyzing and identifying patterns in large volumes of digital content. However, many content analysis tools require researchers to choose between the validity of human-based coding and the ability to analyze large volumes of content through computer-based techniques. This study argues for the use of supervised content analysis tools that capitalize on the strengths of human- and computer-based coding for assessing opinion expression. This study argues for the use of supervised content analysis tools that capitalize on the strengths of human- and computer-based coding for assessing opinion expression. We begin by outlining the key methodological issues surrounding content analysis as performed by human coders and existing computational algorithms. After reviewing the most popular analytic approaches, we introduce an alternative, hybrid method that is aimed at improving reliability, validity, and efficiency when analyzing social media content. To demonstrate the usefulness of this method, we track nuclear energy- and nanotechnology-related opinion expression on Twitter surrounding the Fukushima Daiichi accident to examine the extent to which the volume and tone of tweets shift in directions consistent with the expected external influence of the event.

[7] Sentiment Analysis is a Natural Language Processing and Information Extraction task that aims to obtain writer's feelings expressed in positive or negative reviews, questions and requests, by analyzing a large numbers of documents. In this paper, we extend our ideas pertaining to Sentiment Analysis to the regional language Kannada, spoken mainly

in Karnataka, a state in southern part of India. We have explored the usefulness of semantic approaches and machine learning approaches, used predominately on English language data set, from Kannada web documents. We found the average accuracy of machine learning approaches to be better than the average accuracy of semantic learning approaches for Kannada data set.

[8] While employers already have plenty of reason to eliminate sexual harassment in the workplace, they can add the desire to stay union-free. In this climate, female driven union campaigns will likely increase, creating unique issues for employers and an increased need for well trained female members of management who can persuasively assure female employees that a union is not necessary to stop harassment, achieve pay equity, or otherwise improve the workplace for women.

[9] During the #MeToo movement, social movement organizations (SMOs) played a crucial role in the online mobilization by utilizing various message frames and appealing hashtags during the social movement. Applying a co-creational approach and using framing as a theoretical framework, the study explored how SMOs use words and hashtags to participate in the #MeToo movement through Twitter. Based on both semantic network analysis and thematic analysis methods, findings of the study enhance literature of social movement organizations and activism as well as provide practical implications for effective social movement campaigns.

[10] Role of social media in organizing public protests and mass movements has been the subject of debate. Events such as the Arab revolution (a.k.a Arab Spring) and the Egyptian protests have been cited as prime examples of the socio-political potential of social media (Ghannam 2011; Tufekci & Wilson 2012). Researchers believe that social media plays an important role in aiding protest movements (Ali 2011) and significantly influences the chances of individuals participating or attending protest meets (Tufekci & Wilson 2012). Researchers, policy makers and governments are also starting to take notice. The Internet is being used as a medium to counteract traditional media's (often controlled by handful elites) influence on how information is diffused to the public at large (Della Porta & Mosca 2005). Additionally, ICTs are also bringing about changes in the repertoire of tactics available to ordinary citizens. They are providing new tools like online petitions, hacktivism, online activism etc. (Van Laer & Van Aelst 2010) aiding easy participation and mobilization. As a result, agencies across the world (governments, policy makers and not-for-profit organizations) have been commissioning various studies and reports to understand the role of social media in citizen protests.

References:

[1] Prabowo, R., & Thelwall, M. (2009). **Sentiment analysis: A combined approach**. Journal of Informetrics, 3(2), 143–157. doi:10.1016/j.joi.2009.01.003

[2] **Topic Modelling with Word Embeddings** Fabrizio Esposito, Anna Corazza and Francesco Cutugno

[3] **Stephan A. Curiskis, Barry Drake, Thomas R. Osborn, Paul J. Kennedy,**
An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit,
Information Processing & Management,
Volume 57, Issue 2,
2020,
102034,
ISSN 0306-4573,
https://doi.org/10.1016/j.ipm.2019.04.002.
(https://www.sciencedirect.com/science/article/pii/S0306457318307805)

[4] Modrek S, Chakalov B
The #MeToo Movement in the United States: Text Analysis of Early Twitter Conversations
J Med Internet Res 2019;21(9):e13837
URL: https://www.jmir.org/2019/9/e13837
DOI: 10.2196/13837

[5] Yu, Y., & Wang, X. (2015). **World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans' tweets**. Computers in Human Behavior, 48, 392–400. doi:10.1016/j.chb.2015.01.075

[6] Leona Yi-Fan Su, Michael A. Cacciatore, Xuan Liang, Dominique Brossard, Dietram A. Scheufele & Michael A. Xenos (2017) **Analyzing public sentiments online: combining human- and computer-based content analysis, Information, Communication & Society**, 20:3, 406-427, DOI: 10.1080/1369118X.2016.1182197

[7] **Analysis of Users' Sentiments from Kannada Web Documents** K. M. Anil Kumar, N. Rajasimha*, Manovikas Reddy, A. Rajanarayana and Kewal Nadgir

[8] **#MeToo Movement: Potential Union Organizing Impact?** (2018). Management Report for Nonunion Organizations, 41(4), 2–2. doi:10.1002/mare.30376

[9] Xiong, Y., Cho, M., & Boatwright, B. (2018). **Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of Twitter during the #MeToo movement.** Public Relations Review. doi:10.1016/j.pubrev.2018.10.014

[10] Ray, Deepa and Tarafdar, Monideepa, (2017). "**HOW DOES TWITTER INFLUENCE A SOCIAL MOVEMENT?**". In Proceedings of the 25th European Conference on Information Systems (ECIS), Guimarães, Portugal, June 5-10, 2017 (pp. 3123-3132). ISBN 978-0-9915567-0-0 Research-in-Progress Papers. http://aisel.aisnet.org/ecis2017_rip/60