Understanding the Influence of Unprecedented Events on People's Perception

Authors: Komal Ahuja, Ritika Gupta

Abstract

This paper focuses on the influence of the recent pandemic COVID19 on people's behavior using the data from Twitter. The tweets from before the pandemic and during the pandemic are used to identify the change in the topics around which people have expressed hatred. The policy of twitter to remove such tweets makes this a challenge since getting the data of the actual tweet is very difficult. Our study uses NLP to refine the tweets and draw visualizations depicting the most common words used during both times. The dataset considered is a labelled dataset (1200 rows) with pre covid tweets and during covid tweets with an almost equal distribution of hate (60%) and non-hate (40%) speech. The data mining approach that we use here is topic modelling to understand the most common topics of such hatred during the pre and during covid times.

Introduction

Twitter is a public, global, distributed, and real-time social and information network in which users post short messages called "tweets". Users on Twitter follow other users to form a network such that a user receives all the tweets posted by the users he follows. Tweets are restricted to contain no more than 280 characters of text, including any links. This constraint fosters immense creativity leading to many diverse types of styles and information carried in the tweets. On the other hand, this constraint is also considered a reason the thoughts are left incomplete and thus misunderstood and create a sense of confusion and abusiveness.

Tweets are generated from all over the world on a variety of topics and contain massive information and data that is leveraged for further analysis. The researchers and industries use that data to understand the sentiments of people, the changes in customer trends, crime forecasting etc. Social media text mining is also used in disease forecasting by tracking the spread of the diseases and assessing public awareness [10]. With the freedom of expression privilege granted to social media users, it has become easy to spread abusive/hate propaganda against individuals, groups, and communities.

As per the Twitter policy [1], the tweets that contain violent threats; wishing, hoping, or calling for serious harm on a person or group of people; incitement against protected categories; repeated or non-consensual slurs or any content that degrades someone; hateful imagery; are removed and the accounts are suspended in some cases. Since people are aware about such policies, they put their hate emotions in form of images and memes. The challenge here is that even after having policies in place, the hate speech spreads quickly and a lot of times even percolates into real violence against minorities.

In mid-February 2020, the COVID-19 virus started to spread in various parts of the world and became a pandemic very quickly affecting human life in various ways. People took to social media to express their feelings and understand the situation in the world [9]. The pandemic became the most trending topic on Twitter with various hashtags such as #COVID, #Chinesevirus, #COVID 19, #Coronavirus and open hatred towards communities.

In this paper, we study the effects of unprecedented events such as this pandemic on people's perceptions. To understand these effects, we will analyze the tweets before and during the pandemic, derive the topics using topic modelling, and group them based on their similarities using the clustering technique, and compare them. This would help us analyze if the topics associated with hate within the society have changed and if the concentration of hate towards specific topics has increased.

Related Work/Literature Review

In an experiment by [2], a labeled dataset of 16K tweets was used for hate speech detection through deep neural network architecture. Their approach was to generate embeddings for a tweet and use it for feature representation with a classifier. They concluded that embedding learned from deep neural networks when combined with gradient boosted decision trees (GBDT classifier) lead to best and accurate values with Precision and Recall around 0.930.

In a study [5], they addressed the anti-Asian hate speech during the COVID19 pandemic on Twitter. They used hand labelled dataset to train text-based machine learning classifier and identified three features for the classification. Amongst Linguistic Features (that is stylistic, metadata, and psycholinguistic categories representing the linguistic properties of text), Hashtag features and Bert Tweet Embedding the latter performed the best. The Bert classification model identifies the word level and sentence level semantics by embedding each tweet using BERT base uncased text embedding model [6]. The Hashtag feature performed the worst since the presence of hashtags is inaccurate to label the tweet as hate or non-hate. They also conclude that counter speech is effective in reducing the spread of hatred amongst people.

A paper by [7], investigates the efficacy of the different learning models in detecting abusive language. The accuracy of frequently studied machine learning classifiers (such as Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), Random Forests (RF), Gradient Boosted trees (GBT)) as well as recent neural network models (such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN)) have been compared. Among traditional machine learning models, the most accurate in classifying abusive language found was the LR model followed by ensemble models such as GBT and RF. For neural network models the highest F1 score for "spam" was from the RNN-LTC model (0.551), and the highest for "hateful" was from CNN with context tweets (0.309). Finally, it was concluded that each variant model excels in different metrics, therefore, it is expected to see additional improvements with the use of ensemble models of these variants in future works.

In 2010, Anand Karandikar presented a topic modeling feature of MALLET - a machine learning tool kit, to generate topic models from unlabeled data. The output from a topic modeler contained an inference file and a file consisting of top 'k' words associated with each topic. They used this inference file to infer topics from the disaster dataset. Then they used multidimensional scaling (MDS) to visualize N-dimensional data for exploring similarities and dissimilarities. Then the K-means function was applied to these reduced dimensions which resulted in a vector containing cluster association for each document in the dataset. Finally, they addressed the problem of determining which topic model is optimal for clustering tweets based on its clustering performances.

In research [4], sentiment analysis and topic modelling study are done to analyze public perception of the COVID-19 pandemic on Twitter. They first visualized the most common words (unigrams) through world cloud, then did sentiment analysis using National Research Council (NRC) sentiment lexicon which enabled examining the expressions of 10 terms related to basic emotions. Topic Modelling was then carried out using the latent Dirichlet allocation (LDA) algorithm which treats each document as a mixture of topics and each topic as a mixture of words. The results of the study suggested that people express negative emotions and share both information and misinformation and thus can help government understand the need of the people in such times. They concluded that Twitter is a good platform to understand public awareness and concern about the pandemic and sentiment analysis and topic modelling produce useful information about in trends in discussion of the same.

In [11], the research was done using the experiments Bag of words and term frequency-inverse document frequency values to five different classifiers. Their approach was to first find the probability that a tweet is hate or not using Bag of Words where a corpus of unique words was generated and then labelled. Then they used the TFIDF approach to make it more contextual that is giving weight to the abusive words and thus understand the context. The skewed nature of their dataset that is 93:7 ration of non-hate to hate tweets resulted in inaccurate results when classifying tweets with different contexts. They concluded that on comparing the Logistic Regression, Naïve Bayes, Decision Tree, Random Forest and Gradient boosting classifier, Logistics Regression gave the most accurate results with TFIDF.

Methodology

In this research, we used the Topic Modelling technique to distinguish the various topics on which people posted their hatred on the social media platform Twitter. This study is aimed to identify and compare the topics discussed before the pandemic and during the pandemic. We used tweets from twitter as a dataset for our analysis, since people tend to express their emotions, be they happy, sad, or angry on Twitter, and show support and understanding by retweeting the tweets which is re-posting the tweet. We conducted a series of steps that started from collecting the data for both the times that is pre covid and during covid. We manually labelled tweets as ‘Hate’ or ‘Non-Hate’. Labelled tweets were preprocessed to keep only the meaningful data for analysis. After pre-processing, we applied basic visualizations to understand the frequent words and their frequencies. Subsequently, we applied the data mining technique of Topic Modelling to extract the most relevant words in each topic from the tweets. Finally, we analyze the different topics that were extracted using the LDA algorithm for Topic Modelling and evaluated through Coherence and Perplexity score and the keywords present in each topic. The workflow of our methodology is depicted in Figure 1.

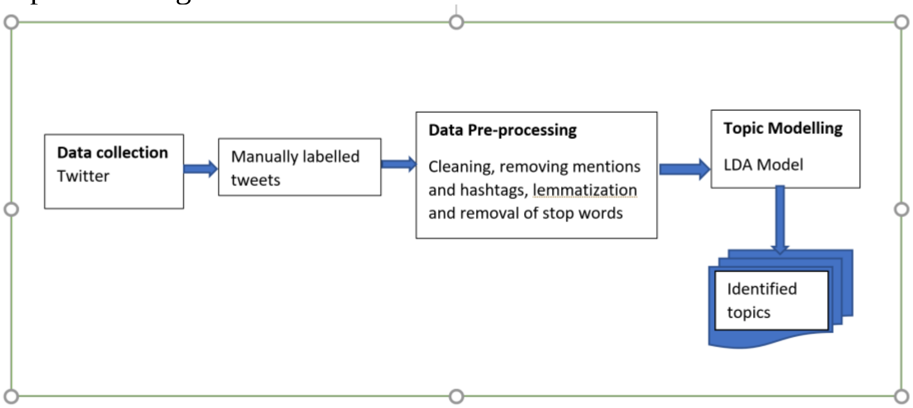


Figure 1: Methodology workflow.

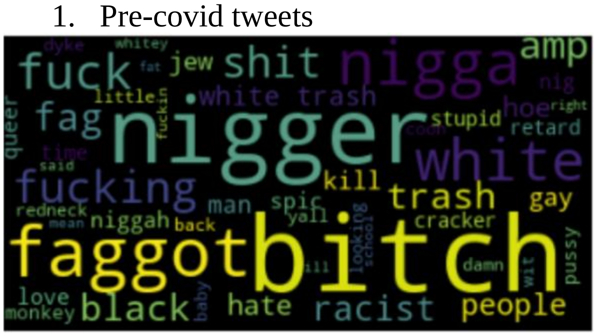
Data Collection

The data collected for the research is the Tweets data in English. The data is collected from both the time ranges that are pre covid and during covid. The pre covid tweets dataset [12] is from Kaggle. The during covid dataset [13] contains tweets created between January 15, 2020, and March 26, 2021. The Pre-covid dataset contained around 24,780 tweets classified as hate speech, offensive language, and neither. This dataset contains text that can be considered racist, sexist, homophobic, or generally offensive. A sample of around 1220 tweets were selected for our analysis purpose which comprised of 60% of tweets as Hate and 40% as non-Hate. The during covid dataset contains the Tweet ID and the corresponding tweet extracted from the Twitter containing COVID 19 hashtag. Twitter has the policy of automatically removing tweets that contain any hatred or abusive context, but the dataset was extracted before the automatic removal of these tweets from the platform. The data is comprised of 1273 tweets which are then labeled manually as Hatred and Non-Hatred denoted by 1 and 0, respectively. The research will use these attributes that are the tweet text and the corresponding indicator of hate to identify the common topics and themes in both the context of hatred and non-hatred.

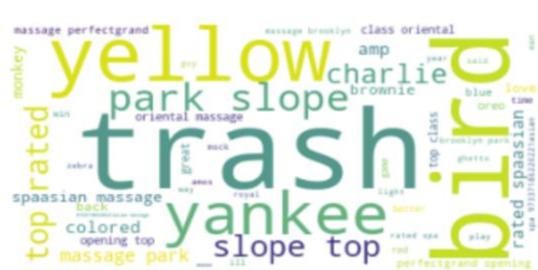
Data Pre-Processing

The raw data contains words such as stop words, irrelevant hashtags, tags which are higher in frequency, but lower in the value with respect to the usefulness in the analysis. These words are removed and only the relevant words are used for further analysis. The data Preprocessing is done in Python and starts with removing the URLs from all the tweets, followed by removing punctuations. Some tweets are multiline and thus they are combined into a single line and the extra blank spaces are removed. Twitter platform is diverse with people who are multilingual and tend to blend words from their native language with English to express their thoughts. Therefore, we removed the words that are not part of English language. Then, we removed the stop words and few extra words such as COVID19, coronavirus, covid 2019 etc. These extra words are the hashtags which were primarily used to extract the Covid related data from Twitter. The data is then divided into the tweets which have hatred and the tweets with no hatred and then n gram analysis is done. We also created visualizations such as world cloud (figure 2) for unigram and bigrams, word frequency bar graph to understand the most common words in the tweets for both the hatred and non-hatred. The figure below shows the word cloud of hate and non-hate words for pre and during covid times.

1. Pre-covid tweets

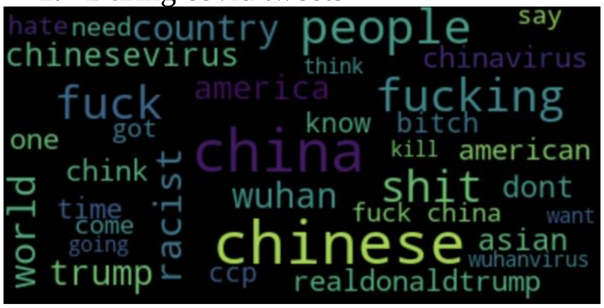


a) Hatred tweets

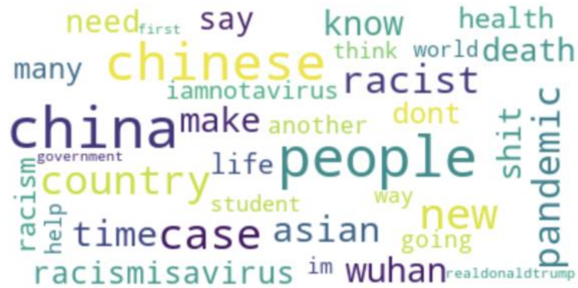


b) Non-Hatred tweets

2. During covid tweets



c) Hatred tweets



d) Non-Hatred tweets

Figure 2: Word clouds from the during Covid Tweets.

Data Mining

The processed data now contains only the relevant words in the tweets. Our study is geared towards the topics around which the hatred was depicted before and during the pandemic and thus we did text mining using the generative statistical model LDA (Latent Dirichlet Algorithm). It is one of the most popular topic modeling methods. LDA works by finding the words that belong to a topic or the probability of words belonging into a topic. The model goes through each document and randomly assign each word in the document to one of k topics (k is chosen beforehand according to the evaluation metrics). For each document d, model go through each word w and compute the following:

1. p(topic t | document d): the proportion of words in document d that are assigned to topic t. Tries to capture how many words belong to the topic t for a given document d. Excluding the current word. If a lot of words from d belongs to t, it is more probable that word w belongs to t.
   (#words in d with t +alpha/ #words in d with any topic+ k*alpha)
2. p(word w| topic t): the proportion of assignments to topic t over all documents that come from this word w. Tries to capture how many documents (or tweets) are in topic t because of word w. LDA represents documents as a mixture of topics. Similarly, a topic is a mixture of words. If a word has high probability of being in a topic, all the documents having w will be more strongly associated with t as well. Similarly, if w is not very probable to be in t, the documents which contain the w will be having a very low probability of being in t, because the rest of the words in d will belong to some other topic and hence d will have a higher probability for those topics. So even if w gets added to t, it will not be bringing many such documents to t.
3. Update the probability for the word w belonging to topic t, as:
   p(word w with topic t) = p(topic t | document d) * p(word w | topic t)

Applying LDA to our dataset, the model considers each tweet as a document and runs through it to find out the keywords and derive topics out of them. We divided the dataset into two, considering one data frame with all the hatred tweets and the other with non-hatred tweets. The model is then run on both the data frames to capture the relevant topics. An important aspect is to figure out the optimal number of topics and for that the evaluation metric Coherence score is used. The model with better coherence score is known to be a better fit. It indicates how closely the words are related in the topics. We also visualize the topics through an interactive chart with intertopic distance map on left and a bar chart (with 30 most salient words and frequencies on right). On hovering over the topic circles, we can visualize the words on the bar graph in those topics. We can adjust the words displayed for a topic by adjusting the lambda (λ) slider. Adjusting it closer to 0 highlights more rare and exclusive terms of the topic.

Experimental Results/Findings

We now provide a detailed analysis of the experimental results from our approach. As part of our evaluation process, we consider the coherence score with various topics numbers. The higher the coherence, better is the model. We generated different number of topics ranging from 2 to 9 and studied the pattern of relevant keywords derived for each topic. We also collected a sample of tweets to derive the topic theme for each topic.

Pre-Covid

On observing the coherence and perplexity score, we saw that the highest coherence score of 0.52 is achieved when the number of topics is 5. However, since most of the tweets had similar hatred keywords, in the visualization of data (see Figure 5a), we can see that the topics are also overlapping, thereby making it challenging to derive topic themes. We also tried adjusting lambda to identify the rare words however, we still did not get distinct topics. Therefore, we minimized the number of topics to 2 to obtain clear topic themes (see Figure 5b). With the number of topics as 2, we got a coherence score of 0.48 and a perplexity score of -7.36. The topics with the most relevant keywords are shown below in Figure 2. Next, we plotted the word cloud (Figure 6, 7) for the keywords generated for each topic for better visualization. Each tweet can be composed of multiple topics, but only one of the topics is dominant. We extracted the dominant topic in each tweet. Figure 3 depicts the dominant topic for each tweet and shows the weight of the topic and the keywords. Similarly, we applied the LDA model to non-Hate tweets and generated a different number of topics and observed coherence scores. For non-hate topics, we were getting the highest coherence of 0.66 with the number of topics as 6 and 4 (see Figure 4). But with the number of topics as 6, it was difficult to derive topic themes since topics were overlapping. Therefore, we selected the number of topics as 4 to identify topics, with a coherence score of 0.66 and perplexity score of –8.74.

```
[(0,
  '0.016*"white" + 0.015*"bitch" + 0.013*"trash" + 0.011*"fuck" + '
  '0.009*"faggot" + 0.007*"kill" + 0.006*"fag" + 0.006*"shit" + 0.005*"hoe" + '
  '0.005*"black"'),
 (1,
  '0.032*"faggot" + 0.026*"nigger" + 0.019*"nigga" + 0.015*"bitch" + '
  '0.014*"fag" + 0.009*"fucking" + 0.009*"white" + 0.008*"fuck" + 0.007*"hate" '
  '+ 0.006*"queer"')]
```

	Tweets	Dominant_Topic	Topic_Perc_Contrib	Keywords	Text
0	0	0	0.8856	white, trash, bitch, faggot, fuck, coon, fag, ...	[slang, racist, porch, monkies]
1	1	0	0.8678	white, trash, bitch, faggot, fuck, coon, fag, ...	[fucking, loser, wetback, sorrynotsorry]
2	2	0	0.9471	white, trash, bitch, faggot, fuck, coon, fag, ...	[blame, black, man, since, blame, whitey, equa...
3	3	1	0.9283	faggot, nigger, nigga, fag, bitch, fuck, fucki...	[forget, nappy, headed, butt, ugly, bitch, mar...
4	4	0	0.8146	white, trash, bitch, faggot, fuck, coon, fag, ...	[hate, fat, bitch]
5	5	1	0.7050	faggot, nigger, nigga, fag, bitch, fuck, fucki...	[faggot]
6	6	1	0.9452	faggot, nigger, nigga, fag, bitch, fuck, fucki...	[muslim, military, honour, whatsoever, sub, hu...
7	7	1	0.9428	faggot, nigger, nigga, fag, bitch, fuck, fucki...	[stupid, fucking, nigger, lebron, flopping, st...

Figure 3: Topics with most relevant keywords

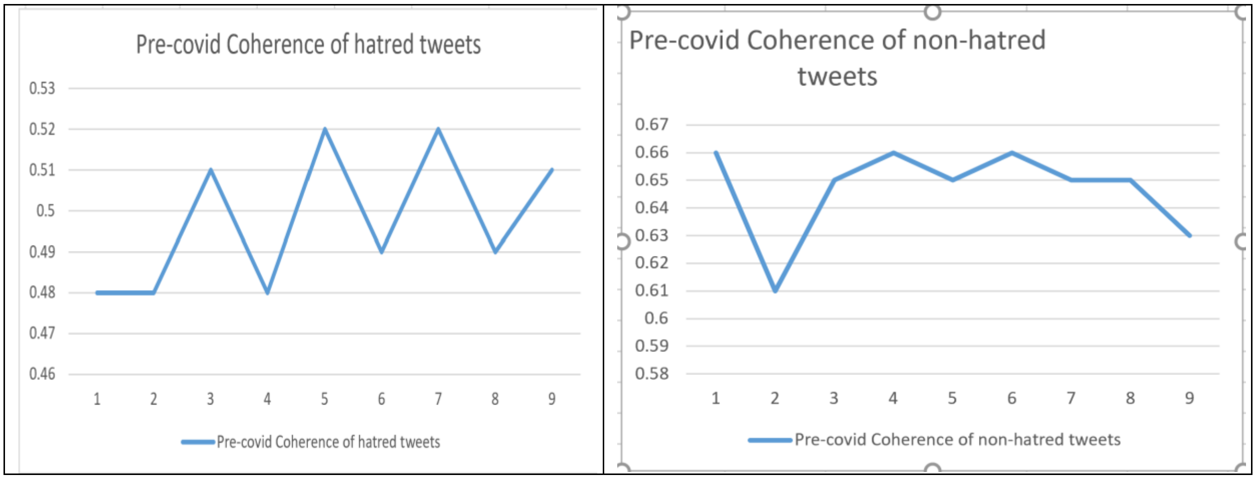


Figure 4: Coherence scores for number of topics from 2 to 9 for all the pre Covid datasets.

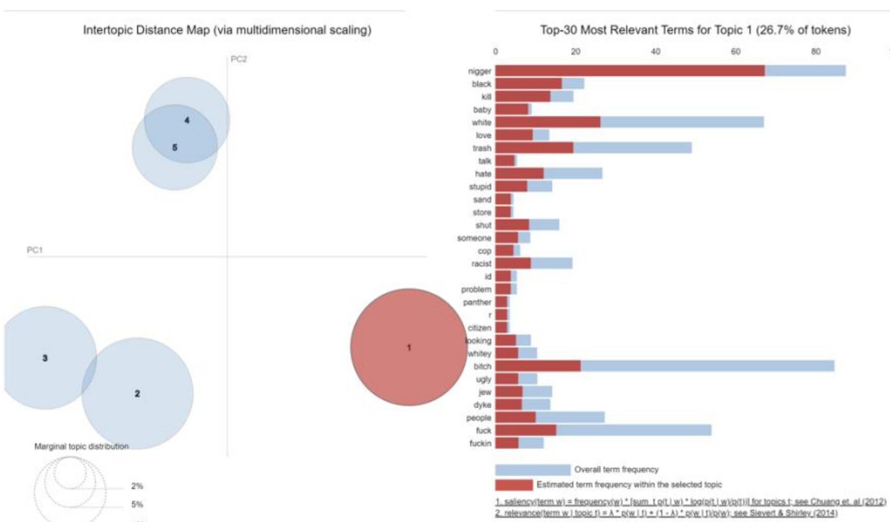


Figure 5a: Interactive chart

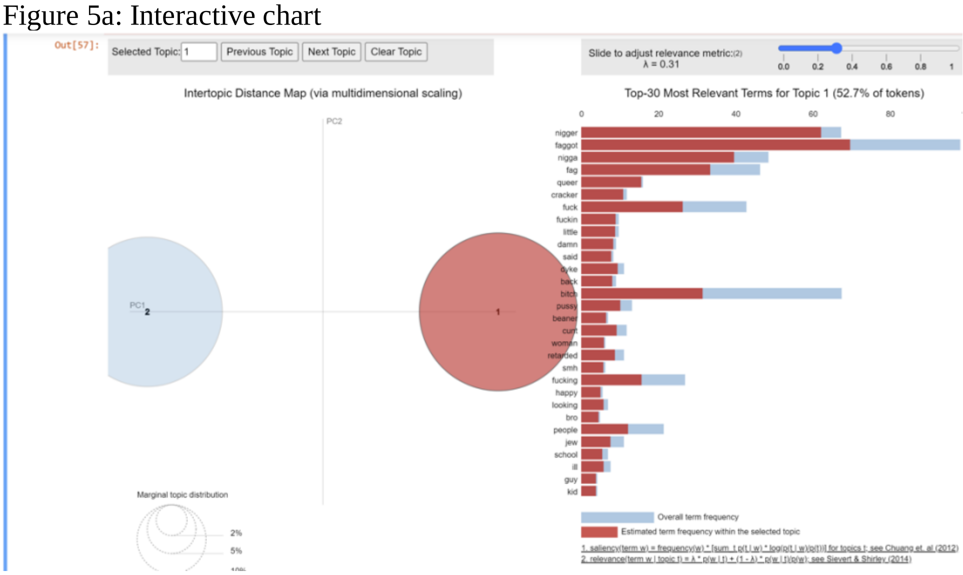
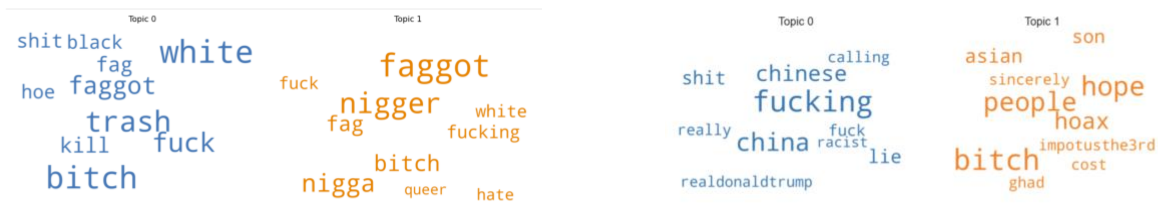


Figure 5b: Interactive chart

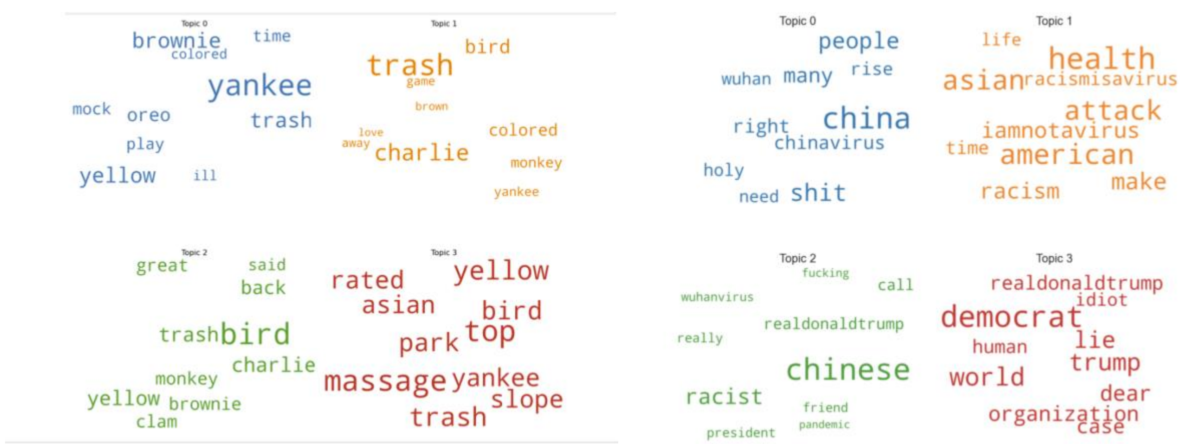
Pre-covid Hate Topics				
Topic Number	Top 5 words	Next Top 5 words	Sample Tweets	Topic theme
Topic 0	white, trash, fag, bitch, coon	fuck, black, fucking, nigga, amp	Tweet 1: @Hovaa_ ya I know all the slang I'm racist I h8 porch monkies	Racial slurs
			Tweet 2: @L1LTR4P fucking losers wetbacks #SorryNotSorry	
			Tweet 3: @infidelpamelaLC I'm going to blame the black man, since they always blame "whitey" I'm an equal opportunity hater.	
			Tweet 4: RT @JihadistJoe: We Muslims have no military honour whatsoever, we are sub human savages that slaughter unarmed men, women & children http:…	
			Tweet 5: @clinchmtn316 @sixonesixband AMERICA today, the rule of thumb is: when in doubt, blame "whitey"	
Topic 1	faggot, nigger, nigga, fag, bitch	fuck, fucking, queer, shit, people	Tweet 1: @kcSnowWhite7 @SamSaunders42 don't forget nappy headed, butt ugly bitch, who's married to a Muslim	Public humiliation
			Tweet 2: I hate fat bitches	
			Tweet 3: RT @Isa__Lopez: @D_Lo520 but you're still a faggot	
			Tweet 4:From now on, I will call all radical MUSLIMS niggers! It is very fitting, and it is racist! I AM RACIST AGAINST ANYONE WHO HATES AMERICA!	
			Tweet 5 : GEEZ..... I think #NorthKorea may be right. #BarackObama is a monkey! Surely acts like one. http://t.co/y09Waoe4EM”	

Table 1: Topic themes with actual tweets for pre-covid dataset



Pre Covid During Covid

Figure 6: Word cloud for hatred topics



Pre-Covid During Covid

Figure 7: Word cloud for non-hatred topics

During Covid

The results for the during covid data show that the coherence considering the hatred and non-hatred tweets together is highest when number of topics is 6 (0.2868). However, we did not see distinct topics themes from all the topics created. They contained very similar words. We also considered adjusting the lambda to see if the rare words helped better in the analysis. Lambda closer to zero gives the rare and exclusive terms that is the words that can differentiate the topics more from one another and define topic exclusivity. We then focused on the model considering the hatred and non-hatred tweets separately. The coherence for hatred tweets is highest when the number of topics chosen is 9 (0.2645) as can be seen in Figure 8, but the relevant words in all those topics were similar. However, on adjusting the lambda to lower values, we could see the rare words, but those rare words did not define the topic exclusivity. It was not significant to consider the optimal number of topics as 9. Therefore, on running the model with lesser number of topics, we saw that the optimal number of topics was 2 (with coherence 0.2331) to make understandable themes out of the topics. The themes mentioned in Table 2 could be derived from those topics by relating the keywords to the actual tweets with the help of dominant topics in every document that is tweet. Then we ran it for the non-hatred tweets and the highest coherence was with number of topics as 9 (0.52). In this case, we did see some relevant themes that could be derived from the topics, yet 9 was not an optimal choice in terms of the keywords and meaning in the topics. Therefore, we considered the next highest coherence with num topics 5 (0.4969). We also plotted word cloud (see Figure 6, 7) for the keywords generated for each topic for better visualization.

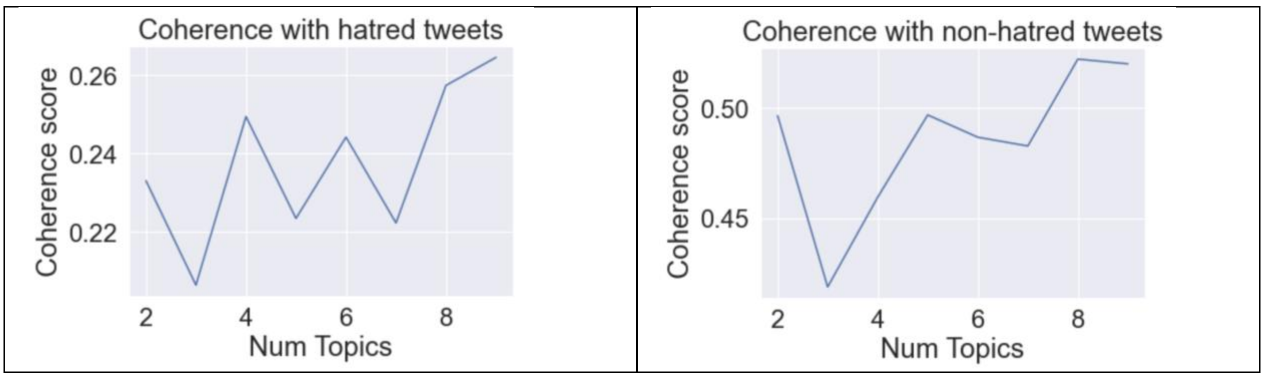
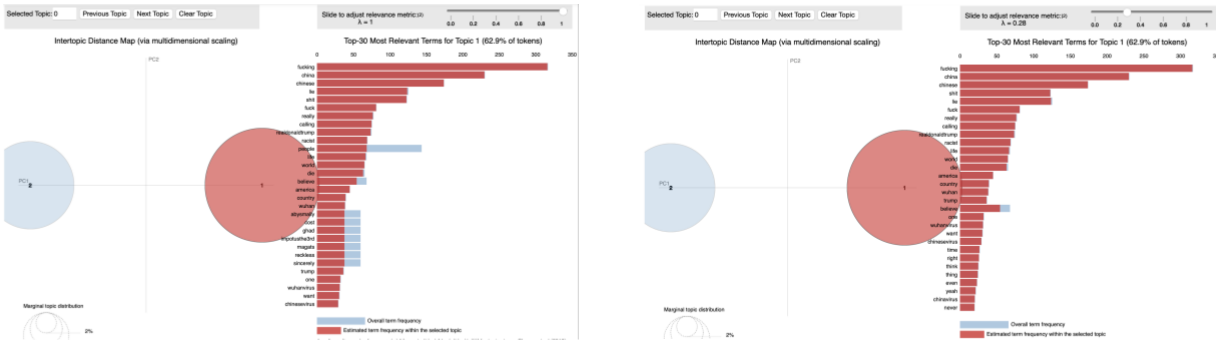


Figure 8: Coherence scores for number of topics from 2 to 9 for all the datasets during Covid.

During Covid Hate Topics				
Topic Number	**Top 5 words**	**Next Top 5 words**	**Sample Tweets**	**Topic Theme**
Topic 0	fucking, china, chinese, lie, shit	fuck, really, calling, realdonaldtrump, racist	Tweet 1: i mean we all kinda knew cause the whole cast had it but fuck u corona	**Government calling**
			Tweet 2: So Republicans want to blame Joe Biden for gas prices, but not hold Donald Trump accountable for 400,000 preventable Covid deaths on his watch or the loss of 9.37 million jobs in 2020? Got it.	
			Tweet 3: If you were serious you would start by calling the thing you're serious about by its actual name, COVID-19. #racist #unfit	
			Tweet 4: Would you support selfishly sitting on your arse for another 3 months being paid by the government to do nothing while everything goes to shit? 83% say yes.	
			Tweet 5: @realDonaldTrump so much for it being the Chinese virus..yo ass sponsored the fucking country,Äôs quarantine https://t.co/oxTRUbhJ6m	
Topic 1	bitch, hope, people, hoax, asian	son, abysmally, cost, ghad, impotusthe3rd	Tweet 1:China is saving the world! ?üò°?üò° Bitch Chinese Communist party and Chinese People spreaded #COVID2019 to more than 200 countries & killing thousands of people Fuck China and Fuck every chinese who eat bats We will never Never Forgive You #ChineseVirus19 #ChinaLiedPeopleDied https://t.co/RMZfpLYIQP	**Racist attacks**
			Tweet 2: Watch this and not cry...wtf America. Seriously W T F #CoronavirusPandemic #COVID19 #AsianAmericans #RacismIsAVirus https://t.co/Ui3CBfcNBv	
			Tweet 3: How abysmally reckless can #IMPOTUSthe3RD be?! Calling #COVID19 A FUCKING HOAX?? This is how people fucking die you son of a bitch! My fucking ghad I sincerely hope your #MAGATS don't REALLY believe your fucking lies. Your LIES cost LIVES @realDonaldTrump this ISNT China, ass.	
			Tweet 4: While out walking my dogs, a man yelled at me that "my kind" caused the pandemic and that I should "get out and go back to China" (I'm a US citizen). I reported it at the link below; please report anti-Asian #COVID19 incidents you witness. https://t.co/ZQQrqwqYQz #IAmNotAVirus	
			Tweet 5: You,Äôre a 5,Äô5 asian male whose no longer a doctor, you,Äôre the epitome of irrelevant, in fact you,Äôre a detriment(Chinese virus / intellectual theft / stealing us tech) Go Fuck yourself shrimp if u don,Äôt like TRUMP	

Table 2: Topic themes with actual tweets for during-covid dataset



Relevant terms for Topic 1 when lambda = 1 | Relevant terms for Topic 1 when lambda = 0.3

Figure 9: Interactive charts with keywords on adjusting the lambda

Limitations and Future Work:

One limitation of this study is that the Covid dataset is collected in a time period when certain events and topics dominated social media. This may have affected the analysis and prevented it from capturing new forms of hate, violence, and racism that could be exacerbated by the COVID-19 pandemic. Additional limitations are the limited amount of data used for the analysis purpose, short duration of the study and the chances of misclassification of tweets.

Future works could explore different Topic Modelling algorithms to identify and compare the topics themes that can be extracted through each algorithm in both time frames that is pre and during Covid. Additionally, follow-up analysis could be conducted to compare and classify the prevalence of hate with other psychological aspects, such as anxiety, fear, and depression, on Twitter data during the COVID-19 pandemic with more amount of data.

Conclusion

In this research, we have proposed the analysis of hatred on the social media platform Twitter before and during Covid. We have used the dataset with Tweets during both times and labeled them as hatred and non-hatred. Then we derived topics out of those tweets using the LDA model considering the keywords. From the pre-covid Dataset, Racial slurs and public humiliation seem to be the primary areas of display of hatred whereas during Covid the hatred clearly targeted towards Asians after the president called out it as Chinese Virus. The coherence value during both the times pre-covid and during covid for the hatred and non-hatred tweets was not of much relevance here due to the limited amount of data. Adjusting the lambda to lower values (around 0.3) gave us rare words which were not of much relevance in understanding the topics. Therefore, we decided to keep the minimal number of topics for the hate tweets. As for the change in people's perception the hate did not seem to increase or decrease but shifted towards the Asian communities.

References

1. Twitter. (n.d.). Twitter's policy on hateful conduct | twitter help. Twitter. Retrieved April 11, 2022, from https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy
2. Badjatiya, P., Gupta, S., Gupta, M., & Varma, V. (2017). Deep Learning for Hate Speech Detection in Tweets. *Proceedings of the 26th International Conference on World Wide Web Companion*.
3. Pereira-Kohatsu, J.C., Sánchez, L.Q., Liberatore, F., & Camacho-Collados, M. (2019). Detecting and Monitoring Hate Speech in Twitter. Sensors (Basel, Switzerland), 19.
4. Boon-Itt, S., & Skunkan, Y. (2020). Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study. JMIR public health and surveillance, 6(4), e21978. https://doi.org/10.2196/21978
5. Ziems, C., He, B., Soni, S., & Kumar, S. (2021). Racism is a virus: anti-asian hate and counterspeech in social media during the COVID-19 crisis. *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*.
6. Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL*.
7. Younghun Lee, Seunghyun Yoon, Kyomin Jung. Comparative Studies of Detecting Abusive Language on Twitter from https://www.researchgate.net/publication/327335971_Comparative_Studies_of_Detecting_Abusive_Language_on_Twitter
8. Anand Karandikar (2010). Clustering short status messages: A topic model-based approach from https://ebiquity.umbc.edu/_file_directory_/papers/518.pdf
9. Bartlett, C., & Wurtz, R. (2015). Twitter and public health. Journal of public health management and practice : JPHMP, 21(4), 375–383. https://doi.org/10.1097/PHH.0000000000000041
10. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EHY, Olsen JM, et al. Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review. PLoS One 2015 Oct 5;10(10):e0139701 [FREE Full text] [doi: 10.1371/journal.pone.0139701] [Medline: 26437454]
11. Kshirsagar, V. (2021, May 29). Detecting hate tweets - twitter sentiment analysis. Retrieved April 11, 2022, from https://towardsdatascience.com/detecting-hate-tweets-twitter-sentiment-analysis-780d8a82d4f6
12. Samoshyn, A. (2020, June 17). Hate speech and offensive language dataset. Kaggle. Retrieved May 6, 2022, from https://www.kaggle.com/datasets/mrmorj/hate-speech-and-offensive-language-dataset
13. Bing He, Caleb Ziems, Ziems, C., Sandeep Soni, , N. R. V., Ramakrishnan, Diyi Yang, Yang, D., Srijan Kumar. Racism is a virus: Proceedings of the 2021 IEEE/ACM International Conference on advances in social networks analysis and Mining. ACM Conferences. Retrieved April 27, 2022, from https://dl.acm.org/doi/10.1145/3487351.3488324
14. Kulshrestha, R. (2020, September 28). Latent dirichlet allocation(lda). Medium. Retrieved May 6, 2022, from https://towardsdatascience.com/latent-dirichlet-allocation-lda-9d1cd064ffa2
15. Prabhakaran, S. (2022, March 8). Topic modeling visualization - how to present results of LDA Model: ML+. Machine Learning Plus. Retrieved May 6, 2022, from https://www.machinelearningplus.com/nlp/topic-modeling-visualization-how-to-present-results-lda-models/
16. Klaifer Garcia, Lilian Berton. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. Retreived May 8, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7832522/#sec7

17. Raghad Alshalan, Hend Al-Khalifa, Duaa Alsaeed, Heyam Al-Baity, Shahad Alshalan. (2020, August 12). Detection of Hate Speech in COVID-19–Related Tweets in the Arab Region: Deep Learning and Topic Modeling Approach from https://www.jmir.org/2020/12/e22609/