

Leveraging Natural Language Processing to Understand Public Outlook Towards the Influenza Vaccination

Ankita Agarwal
Department of Computer Science
Wright State University
Dayton, Ohio 45435 USA
agarwal.15@wright.edu

William L. Romine
Department of Biological Sciences
Wright State University
Dayton, Ohio 45435 USA
william.romine@wright.edu

Tanvi Banerjee
Department of Computer Science
Wright State University
Dayton, Ohio 45435 USA
tanvi.banerjee@wright.edu

Abstract—Understanding public outlook in healthcare management is important in the study of the various diseases. With respect to vaccinations, which play a major role in combating vaccine-preventable diseases, the study on their acceptance or rejection by the public becomes useful. In particular to the influenza vaccine, studies on the public opinion and views is ongoing. Social media platforms like Twitter help us to leverage thoughts and attitudes related to the flu vaccine. The data set used for our analysis contained tweets related to vaccines which were collected using vaccine-related keywords over a period of twelve months from February, 2018 to January, 2019. Out of these tweets, we filtered out the tweets specific to the flu vaccine and generated our corpus for further study. By using Latent Dirichlet Allocation (LDA), we identified eighteen topics comprising six major themes which best represented our corpus. In this paper, we discuss these six themes and subsequently analyze the trend observed in these themes over a period of twelve months. The themes identified covered various aspects related to the flu vaccine. Among the six major themes, four showed a distinctive temporal trend with respect to the annual flu season.

Index Terms—influenza, flu shot, machine learning, topic modeling, vaccinations

I. INTRODUCTION

Vaccinations play a critical role in the healthcare system to prevent diseases and to reduce the mortality and morbidity rate of many communicable diseases [1]. People have various views related to the effectiveness and the safety of vaccines which influence their acceptance or rejection [2]. Various campaigns from the healthcare system, government and non-profit organisations emphasise the importance of vaccines and promote their widespread use [3]. Critical study on pathogens along with their mutant variants are studied over time to help the production of vaccines [4]. Of particular interest is the flu vaccine due to its seasonal cycle [5]. Public opinion becomes an important factor to study concerns and views over the flu vaccine [6]. In this regard, data from social media platforms have become useful for analyzing the ongoing public discussions over health issues like vaccination [7]. One of these social media platforms, Twitter, has become a prominent platform to monitor and analyse public opinion in the context

of disasters like hurricanes [8], earthquakes [9], and cyclones [10], and outbreaks of various communicable diseases like flu [11], measles [12], Ebola [13], Zika [14] and COVID-19 [15]. Even discussions related to the methods to combat such diseases like vaccination [16], medications [17], and preventative measures [18] have been analysed using these social media platforms. Machine learning [19] and natural language processing [20] methodologies have been implemented to study the public discussions on various diseases and the ways to combat these diseases. Topic modelling [21] is one such methodology to identify the hidden themes in the text. In connection to flu vaccines, [22] evaluated topic modelling to identify the posts on Twitter that most likely contain vaccine safety signals. [23] analysed the geolocated tweets from U.S. counties using MALLET LDA (Latent Dirichlet allocation) topic modeling techniques to correlate the individual survey data about vaccine attitudes, actual vaccination, and real-life discussions about vaccines with family and friends during the 2018–2019 influenza season. [24] used LDA to describe their corpus with respect to the public outlook towards vaccines. Their corpus also included tweets extracted from Twitter using vaccine-related keywords, including those related to the influenza vaccine. With respect to the temporal trend observed in supporting (pro vaxxers) and rejecting (anti vaxxers) vaccines, [25] used topic modelling to examine how pro-vaccine and anti-vaccine topics and public interests have changed from 2007 to 2017. As far as we are aware, no previous work has implemented topic modelling to study the discussions related to the flu vaccine on Twitter over a period of a year. Given the seasonal nature of influenza, this could be useful to observe how the public interest and discussions on the flu vaccine changes with time.

In this study, we addressed the following research questions:

- 1) RQ1. What are the underlying themes discussed on Twitter related to flu vaccination?
- 2) RQ2. How do these underlying themes vary every month across a period of one year?

II. DATA

A. Data Collection

To analyze the above questions, we used the data set built by [24] which consisted of tweets extracted from Twitter from the beginning of February, 2018 through January, 2019. The tweets were collected using 78 vaccine-related keywords, some of which included ‘vaccine’, ‘vaccination’, ‘flu’, ‘autism’ and ‘MMR’ (Measles, Mumps, Rubella). Approximately 8 million tweets were collected over this time span. From this data set, we separated the tweets related to the flu vaccine by using the keywords “anti flu shot”, “antiflushot”, “flu shot”, “flushot”, “flu season”, “fluseason”, “flu vaccine”, “fluvaccine”, “flu vaccination”, “fluvaccination”, “influenza vaccine”, “influenzavaccine”, “influenza vaccination”, “influenzavaccination”, “flu”, “influenza”, and prepared a separate corpus. Thereafter the duplicate tweets from this corpus were removed and the remaining 83,774 were used for further analysis.

B. Data Preprocessing

Certain pre-processing steps were performed on our data set, consisting of 83,774 tweets, before performing topic modelling and trend analysis. First of all, normalization was done to convert the text to lowercase. Then, the punctuation marks and excessive white spaces were removed. Next, the symbol “#” from the hashtags, “@” from mentions were removed along with the Uniform Resource Locators and special characters. Stop words like ‘the’, ‘we’, and ‘has’ were also removed as they do not tend to convey important information. Keywords which were used to extract the tweets were also removed.

III. METHODOLOGY

In order to answer the research questions, we implemented topic modeling via Latent Dirichlet Allocation (LDA) on our data set consisting of 83,774 tweets. LDA [26] is an unsupervised probabilistic model to automatically identify themes (topics) in text documents. It classifies the documents (i.e. tweets) into topics representing the data set. After pre-processing the 83,774 tweets as discussed in Section II-B, we first converted the data to bag-of-words corpus and then used the Gaussian Mixture Model’s [27] Akaike’s Information Criterion (AIC) and Bayesian Information Criterion (BIC) [28] to determine the optimal number of topics, where lower AIC and BIC scores indicate a better model for balancing fit and parsimony. In order to do this, we used the Gaussian Mixture Model from Python Sklearn package¹. The corresponding AIC and BIC graphs were plotted to infer the optimal number of topics. Based on this, we ran a LDA model with this optimal number of topics using the Gensim package². The topics thus obtained consist of the top representative words along with their probabilities. Henceforth, the topic for each tweet was identified using the LDA model. The theme for each topic was identified through manual interpretation by inspecting the

words in each topic and some representative tweets belonging to each topic. The tweets were then divided by month, and then we calculated the percentage of tweets belonging to each topic per month. The topics were further grouped into broader themes manually and the overall trend of these broader themes was studied over the twelve-month time period.

IV. RESULTS

The GMM AIC and BIC score for the varying numbers of topics for our data set were similar and indicated the same number of optimal topics. The BIC graph for our data set is shown in Fig. 1. According to the graph, the GMM BIC score decreased until 18 topics and then gradually started to level off, hence we chose 18 to be the optimal number of topics for our data set.

After determining the topic for each tweet, we identified the theme related to each topic manually by looking at the keywords in each topic and some representative tweets for each topic. According to our interpretation, these 18 topics could be grouped into six broader themes. The percentage distribution of the tweets belonging to each broader theme is shown in Table I. The results of the 18-topic LDA model grouped into 6 broader themes are discussed below in Table II, along with the top 10 most relevant terms for each topic and representative tweets for each topic. The broader themes identified for our data set are: Importance of the annual flu vaccine, Research and testing on flu vaccines, Promotion of the flu vaccine, Nature of the flu virus, Anti vaxxers, and Vulnerable populations.

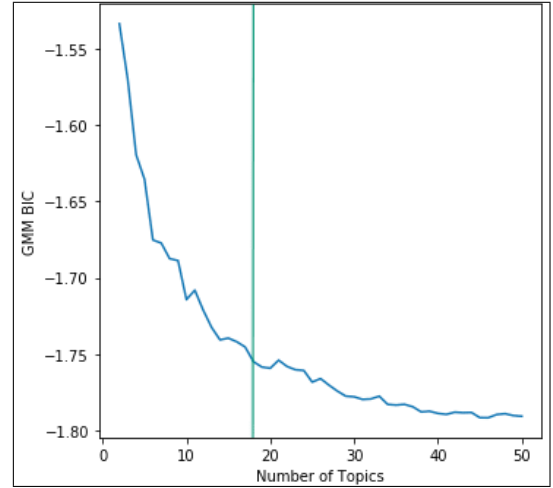


Figure 1. Bayesian Information Criterion (BIC) graph for topic modeling with varying number of topics. Green line shows that the GMM BIC score leveled off at 18 topics.

We discuss the themes for each topic as curated based on the top 10 relevant terms below.

- **Importance of the flu vaccine:** Within the “Importance of the flu vaccine” themes, Topic 1 emphasised that the flu vaccines reduce complications during the annual flu

¹<https://scikit-learn.org/stable/modules/mixture.html>

²<https://radimrehurek.com/gensim/>

Table I
BROADER THEMES DISTRIBUTION IN THE DATA SET

Topic Number	Theme	Tweets(%)
1	Importance of the flu vaccine	21.9
2	Research and testing on the flu vaccine	21.8
3	Promotion of the flu vaccine	16.7
4	Nature of the flu virus	17.2
5	Anti vaxxers	12.2
6	Vulnerable populations	10.2

season. Topic 2 contained tweets related to the people who support the influenza vaccine for fear that the flu may become a pandemic. Topic 3 reflected the opinions of pro vaxxers towards annual flu vaccination and Topic 4 talked about the information on how the flu vaccine protects against illness.

- **Research and testing on the flu vaccine:** Under this theme, Topic 1 discussed the study on the effectiveness of the flu vaccine and its role in the reduction of flu-related cases. Topic 2 contained tweets related to the ongoing research on the flu vaccine.
- **Promotion of the flu vaccine:** Under this theme, Topic 1 talked about the availability of flu shots at local pharmacies. Topic 2 covered tweets which asked people to protect their families from the flu through vaccination. Topics 3 and 4 depicted the tweets related to campaigns encouraging people to get vaccinated.
- **Nature of the flu virus:** In this theme, Topic 1 discussed that as the flu virus has different strains, getting a universal flu vaccine is a challenge. Topic 2 discussed the virulence of the flu virus, and Topic 3 stated that the flu virus strain changes yearly.
- **Anti vaxxers:** The theme ‘Anti vaxxers’ contained two topics. One topic mentioned that the flu vaccine is a doctor’s business and that the vaccines are not safe. The other topic depicted general skepticism and resistance towards the flu vaccine.
- **Vulnerable Populations:** This theme talked about the types of people who are at high risk of contracting the flu during the flu season. In this theme, Topic 1 talked about the special facilities for children for getting the flu vaccine. Topic 2 mentioned that the flu vaccine is recommended for pregnant women and Topic 3 said that the flu vaccine is also helpful for working age adults and elderly people.

After calculating the percentage of tweets belonging to each broader theme for every month, we plotted a figure of the temporal trend as shown in Fig. 2. In this figure, we plotted the percentage of tweets belonging to each broader theme per month, against that particular month. This was done in order to visualize the trend in each broader theme being discussed on Twitter related to the flu vaccine over a period of 12 months. This figure shows that the theme, “Importance of the flu vaccine” was discussed more during the months of October, 2018

and November, 2018 while the opposite trend was observed in the theme, “Promotion of the flu vaccine”. During the months between August, 2018 and October, 2018, there was a lot of discussion on “Research and testing of the flu vaccine” but very few tweets fell within the “Vulnerable populations” theme at that time. The trend in the “Nature of the flu virus” and “Anti vaxxers” did not show distinguishable differences over the time period.

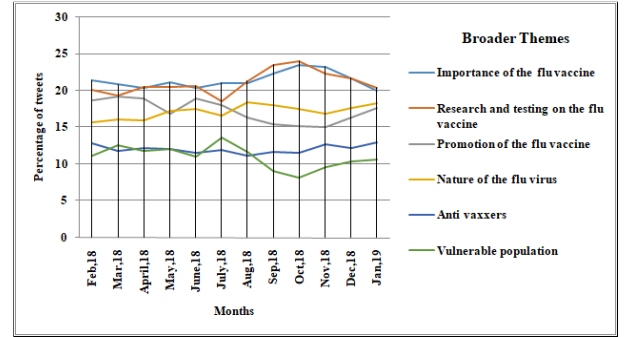


Figure 2. The prevalence of the broader themes over time

V. DISCUSSION

A. Topic Modeling

These six themes covered a broad spectrum of the conversations related to the flu vaccine on Twitter and are discussed in more detail below:

- **Importance of the flu vaccine:** The importance of the flu vaccine for public health has been explained by researchers in [29]. The “Importance of the flu vaccine” theme identified in our corpus also depicted this and consisted of four topics. The tweets in its first topic like “WINTER IS COMING! Annual vaccination is the most effective way to protect against influenza infection and is advised for people with asthma. People with asthma are at higher risk of experiencing severe complications following the flu virus” mentioned that vaccination could reduce severe complications in people who have other chronic diseases like asthma. Tweets in the second topic like “Currently we support flu vaccines. The pharma gets \$25/shot. And if people get sick because of the shot, it’s just the flu, so who cares, we don’t have to pay for it” depicted that people supported vaccination notwithstanding the cost involved or any vaccine related risks, if any.

Tweets tweeted by “Pro vaxxers” included statements like: “In case you haven’t been watching the news, influenza is widespread in 19 states and rising. Already this year 3 children have died. The flu vaccine this year is extremely effective. The time to vaccinate is NOW. #Vaccinate #Flu #Influenza #H1N1 #FluVaccine #WednesdayWisdom”. These statements advocated the importance of vaccination and associated risks of not getting vaccinated. Simultaneously, this topic also contained tweets which combated a myth that the flu vaccine

Table II

BROADER THEMES WITH TOPICS UNDER THEM ALONG WITH THE WORDS AND REPRESENTATIVE TWEETS FOR EACH TOPIC

Broader theme 1: Importance of the flu vaccine			
S.no	Topic	Words	Representative tweets
#1	Flu vaccine reduces complications	today, vaccination, prevent, staff, get, annual, seasonal, reduce, complication, forget	WINTER IS COMING! Annual vaccination is the most effective way to protect against influenza infection and is advised for people with asthma. People with asthma are at higher risk of experiencing severe complications following the flu virus such as:
#2	Support for the flu vaccine	nasal, check, appointment, pandemic, here, healthy, life, support, vaccine, love	Anti-vac folks: Spanish Influenza became pandemic on three different occasions. Since the 'flu' managed to take as many as 40 million lives between 1918 to 1919, an intense need for medical progress and vaccination development followed as a result.
#3	Pro vaxxers	come, go, measles, yesterday, fight, always, mean, yearly, vaccine, watch	Holy shit, I want to put all the antivaxxers on an island and watch them all get the flu, measles, polio, etc. Have fun with 0% herd immunity you dumb fucks.
#4	Information on how the flu vaccine protects	season, important, parent, offer, protection, ready, vaccine, information, someone, almost	Important information regarding your yearly #fluvaccine
Broader Theme 2: Research and testing on the flu vaccine			
S.no	Topic	Words	Representative tweets
#1	Effectiveness of the flu vaccine	take, week, vaccine, reminder, report, effectiveness, make, number, case, little	New #influenza #vaccine with patented adjuvant from NovavaxInc shows improved effectiveness over egg-based #flu shot in recent trials
#2	Research on the flu vaccine	vaccine, disease, symptom, develop, research, serious, visit, find, common, illness	In recent flu seasons, 4 in 5 children who died of the flu hadn't been vaccinated — and that's almost certainly not a coincidence. Research shows vaccination helps reduce the risk of healthy children dying from the flu by two-thirds. #VaccinesWork
Broader Theme 3: Promotion of the flu vaccine			
S.no	Topic	Words	Representative tweets
#1	Availability of the flu shots at community pharmacies	start, begin, pharmacy, community, vaccine, especially, different, tetanus, prepare, surgery	Community pharmacies administer almost 250,000 more flu vaccines: Community pharmacy teams administered at least 240,000 more flu vaccinations during this winter than last. To read the whole article click on the headline.
#2	Protect your families against flu	protect, get, vaccinate, family, vaccine, available, everyone, people, local, other	Now is a great time to also begin thinking about flu season and vaccinations! Protect yourself and others by getting a flu vaccine! @GBMCHHealthcare Health Partners primary care practices expect to have the 2018 vaccine available by late August.
#3	Campaign for the flu vaccine	great, right, morning, provide, medical, campaign, share, encourage, vaccine, article	The number of people vaccinated under the nation-wide campaign to fight the influenza virus, launched by the Ministry of Public Health, has breached the 100,000 mark and is set to continue in the coming days. #Flu #Vaccination #QatarLiving
#4	Urge for public immunization	health, clinic, winter, public, immunization, offer, urge, county, deadly, effect	This #flu season is deadly. #Sepsis is responsible for most deaths from #flu. Know the signs of #sepsis. Please share this video. It could save a life. #Flu2018
Broader Theme 4: Nature of the flu virus			
S.no	Topic	Words	Representative tweets
#1	Flu season and different strains of flu virus	october, vaccine, better, universal, least, decade, quadrivalent, maybe, probably, nearly	My understanding is that there are many variants of the flu virus, so a universal vaccine is probably not possible
#2	Virulence of the flu virus	people, vaccine, student, thing, kill, american, tomorrow, eligible, official, there	Last year, influenza killed 80k Americans. 80,000/53% of us don't get the flu shot. Some of us have compromised immune systems and can't get the vaccine. If YOU get the vaccine, not only are you less likely to get sick, you're less likely to infect someone else.
#3	Virus strain changes yearly	vaccine, year, first, strain, virus, really, effective, hospital, change, cover	Getting vaccinated every year is important because the most common strains of the influenza virus change every year. So the vaccine that you get this year won't necessarily protect you against the different virus that may be circulating next year....
Broader Theme 5: Anti vaxxers			
S.no	Topic	Words	Representative tweets
#1	Flu shot is dangerous and is a money business	give, school, vaccine, reason, doctor, believe, importance, anyone, try, contain	Flu shot is a money business, to be exact 3 billion per year. People die and get injured by flu vaccine. They die of the flu despite having flu shots, so people should more concentrate at boosting their immune system, this will get them through the flu, not the flu shot!
#2	Skeptics and resistance towards the flu vaccine	vaccine, think, never, still, get, well, help, people, remember, healthcare	The flu shot never was introduced to cure you but only make you sicker the so called vaccine has caused more deaths and more sickness than healing the virus
Broader Theme 6: Vulnerable populations			
S.no	Topic	Words	Representative tweets
#1	Special facilities for children	thank, nurse, immune, friend, young, system, center, spray, child, vaccine	"It kills more children than meningococcal disease and the [best] prevention is vaccination." The flu vaccine available as of today for children aged between 6 months - 5 years in NSW
#2	Pregnant women	shot, vaccine, recommend, child, receive, would, pregnant, month, early, adult	Let's not forget how vulnerable pregnant women, and their babies, are to #Influenza. We health professionals can strongly influence knowledge and acceptance of #vaccination.
#3	Working age adults and elderly people	already, vaccine, immunity, office, death, elderly, feel, fluoride, issue, meningitis	Vaccines protect the elderly + young that may be more susceptible to the flu, not just yourself. I know a lot of you may already be aware, but still, get your Flu shot soon. You can often get free shots through employer, college, etc.

could be used to treat other diseases like measles as understood by the tweet: "Well it's a good thing ur not a researcher....the flu vaccine isn't given for measles and to quote your study "our findings do not demonstrate a causal role of vaccination in the pathoetiology of any of these conditions".

The topic like "Information on how the flu vaccine protects" contained tweets that dealt with providing relevant information about the flu vaccine by directing people to relevant news or articles. At the same, tweets like, "please stop spreading this dangerous and utterly false information about the flu vaccine. It is not "neurotoxic," it CAN'T sicken people, and it saves lives. The flu kills thousands even in a mild winter. Educate yourself or

STFU", showed that there is a lot of misinformation about the flu vaccine on social media.

- **Research and testing on the flu vaccine:** Every year, research and testing on the effectiveness of the flu vaccine is carried out and an interim report is published on the same annually [30]. The second theme in our corpus, which was "Research and testing on the flu vaccine", also discussed the same. It had the first topic as the effectiveness of the flu vaccine. The tweets in this topic talked about the effectiveness of the flu vaccine as studied by observational studies which was clearly understood in the tweet "Influenza vaccine. Positive mood on the day of influenza vaccination predicts vaccine effectiveness: A prospective observational cohort study. UK, general

practice. #vaccineswork”. The other topic within this theme talked about the ongoing research on the study and production of flu vaccines.

- **Promotion of the flu vaccine:** The flu vaccine is promoted every year and studies on the effective promotion of these vaccines have become quite important [31]. The third theme, “Promotion of the flu vaccine” contained tweets which talked about the availability of the vaccine at healthcare centres like “GP surgery” and asked people to protect their families through vaccination. Campaigns supporting the flu vaccine like “Get a Jab, Give a Jab” and “#FightFluTN” were also covered in this theme. This topic also contained tweets from health care centres which urged people to get vaccinated for #Flu2018.
- **Nature of the flu virus:** Researchers study various aspects of the flu virus including the diversity of strains and its ability to mutate [32]. The fourth theme in our data set, the “Nature of the flu virus”, mentioned the different strains of the flu virus and so it explained the reason as to why a universal flu vaccine might not be possible. There was an emphasis to get the flu vaccine by the end of October as clear from the tweet, “Patricia Schnabel Ruppert, MD, @Columbia assistant clinical professor of medicine, explains the benefits of getting your #flu shot by the end of October”. As the flu virus has different strains, there were tweets talking about the quadrivalent vaccine designed to protect against four different influenza strains; two influenza A strains and two influenza B strains. This theme had a topic which talked about the virulence of the flu virus as it kills many Americans every year [33]. The “antigenic drift” in the flu virus from one season to another season was also among the information discussed in this topic.
- **Anti vaxxers:** Anti vaxxers have been a topic of concern for public health officials and various researchers have addressed this issue [34]. In our corpus, this theme was also very interesting as it contained tweets which explained the other side of the coin as opposed to the pro vaxxers. These tweets were grouped into two topics. The first topic contained tweets that said that the flu vaccine is dangerous and it is predominantly a money business by doctors. Some people even called it a shady business. Under this topic, there were also tweets which spoke to conspiracy theories about the CDC doctor who went missing and was eventually found dead. Anti vaxxers claimed that the doctor was killed by the CDC after he raised his voice against the safety of flu vaccines as clear from the tweet, “So the CDC murdered the doctor who told the truth about the flu shot giving people the disease. We search your research: your vaccines are killing us”. Anti vaxxers in our corpus also considered the content of the vaccine like mercury in the Thimerosal flu vaccine and aluminum as dangerous, and also claimed that the flu vaccines were inefficient. The second topic under this theme depicted skepticism and resistance towards the flu vaccine. Some tweets mentioned the side effects of the flu

vaccine like headache and body aches while others cited that the flu shot causes Guillain Barre Syndrome as clear from the following tweet: “I have a friend that developed Gillian Barre from the flu shot a few years ago. In fact a group of people developed GB that year”. In total, this theme reflected both misconceptions and concerns about the flu vaccine, which were also cited in [35].

- **Vulnerable populations:** Health officials often recommend the flu vaccine to those highly susceptible to the flu virus like children [36], pregnant women [37] and elderly people [38]. Some tweets in our corpus also explained types of people who were at high risk of contracting and experiencing complications from influenza. These were explained under one broader theme as “Vulnerable populations”. Under this theme, Topic 1 explained the different facilities available for children for getting the flu shot. The second theme mentioned the importance of the flu vaccine to pregnant women as clear from tweets like “Let’s not forget how vulnerable pregnant women, and their babies, are to #Influenza. We health professionals can strongly influence knowledge and acceptance of #vaccination”. Topic 3 explained that flu vaccines are very important and relevant to working age adults and the elderly; however, due to the many accompanying health factors in the elderly, even flu shots are sometimes ineffective for preventing life-threatening complications. This was mentioned in the tweet, “Flu relevant. Vaccine fails to protect elderly from killer flu”.

B. Temporal trend of broader themes

The broader theme, “Importance of the flu vaccine” was more discussed from September to November, during which the public is encouraged to get vaccinated. According to the Centers for Disease Control and Prevention (CDC)³, flu activity increases during October and peaks between December and February. The CDC therefore recommends that flu vaccines should be taken by the end of October but not before September [39]. For the theme, “Research and testing on the flu vaccine”, more tweets talking about this were observed at the beginning of the month of August until the month of October, after which it started decreasing. This may be due to the fact that this was a time when the effectiveness of the flu vaccine was emphasized due to the upcoming flu season. As this theme also covered the tweets related to the research on flu vaccines, it may also explain the details of the types of flu vaccines available during that particular season after considering the nature of the flu virus that may prevalent that particular year.

The theme, “Promotion of the flu vaccine” was discussed the most during the months of May to July. As these months are not typically the months of flu season, during this time the planning for the campaigns for the next flu season might have been discussed but as the flu season approached, people started talking about the “Importance of the flu vaccine”. Thus,

³<https://www.cdc.gov/flu/about/season/flu-season.htm>

the trends of these two themes were opposite to each other. Percentage of tweets talking about the “Nature of the flu virus” and “Anti vaxxers” remained almost the same throughout the year. The “Nature of the flu virus” theme peaked slightly in August as the flu season was approaching but the theme “Anti vaxxers” did not show any specific temporal trend. This indicates that those opposed to the flu vaccine tend to express that sentiment year-round.

We observed that the trend of the “Vulnerable populations” theme was opposite to the trend of the “Research and testing of the flu vaccine”. This may be due to the fact that during the months of June to August, promotions for vaccinating vulnerable populations like children, pregnant women, working age adults, and elderly people began, and then as the flu season approached, the emphasis shifted towards the study of effectiveness of flu vaccines on these vulnerable populations. This kind of analysis could be useful to health and government officials to study public perceptions towards the flu vaccine. Both kinds of people, either supporting or rejecting the flu vaccine are present in a community and it is important to be aware of and to address their concerns. The temporal trends serve as a guide to the types of discussions around the flu vaccine that happen at different times during the year, and are therefore useful for understanding the types of information we should be providing the public during these times. For example, the finding that anti-vax discussions around the flu vaccine are happening year-round tells us that misinformation and concerns around the flu vaccination are discussed continually. The fact that discussions on the importance of flu vaccination are happening in early Autumn is promising in that they accompany the time when getting the annual vaccination is recommended.

C. Strengths and Limitations

The flu months in the Northern and Southern Hemispheres are different. Generally, the flu season in the Northern Hemisphere can begin as early as October and can last as late as April or May while in the temperate regions of the Southern Hemisphere, influenza activity typically occurs during April – September. In India, the flu season typically arrives during the monsoon time. We do not address this issue in our temporal data. Our temporal analysis is predominantly based on the flu season of the temperate regions in the Northern Hemisphere. This work focuses on discussions of flu vaccines for humans, but as flu is common in horses, cats, dogs, birds, and swine, we do not address the themes surrounding the tweets related to the perception of flu vaccines for animals.

The social media platform used in our study is Twitter and so we have not included the public discussions on the flu vaccine on other social media platforms like Facebook, Youtube and on blogs, news, and academic articles. These limitations should be considered when attempting to generalize our findings, and for planning of future research.

VI. CONCLUSION

The flu vaccination plays a major role in combating influenza. While many people accept that flu shots are effective, some people are resistant to these vaccines for various reasons. Data from social media platforms can play an important role to understand the major topics of discussion and perceptions around the flu vaccine. These discussions help us to understand the factors which influence the acceptance or rejection of the vaccine. The reach of various campaigns to promote the flu vaccine could also be inferred from this kind of analysis.

REFERENCES

- [1] E.-P. Chevalier-Cottin, H. Ashbaugh, N. Brooke, G. Gavazzi, M. Santillana, N. Burlet, and M. Tin Tin Htar, “Communicating benefits from vaccines beyond preventing infectious diseases,” *Infect Dis Ther*, Jun. 2020.
- [2] J. Leask, H. W. Willaby, and J. Kaufman, “The big picture in addressing vaccine hesitancy,” *Hum. Vaccin. Immunother.*, vol. 10, no. 9, pp. 2600–2602, Nov. 2014.
- [3] A. Llupia, G. Mena, V. Olivé, S. Quesada, M. Aldea, V. G. Sequera, J. Ríos, A. L. García-Basteiro, P. Varela, J. M. Bayas, and A. Trilla, “Evaluating influenza vaccination campaigns beyond coverage: a before-after study among health care workers,” *Am. J. Infect. Control*, vol. 41, no. 8, pp. 674–678, Aug. 2013.
- [4] G. Ada, “Vaccines and vaccination,” *N. Engl. J. Med.*, vol. 345, no. 14, pp. 1042–1053, Oct. 2001.
- [5] C. C. Blyth, P. Jacoby, P. V. Effler, H. Kelly, D. W. Smith, C. Robins, G. A. Willis, A. Levy, A. D. Keil, P. C. Richmond, and WAIVE Study Team, “Effectiveness of trivalent flu vaccine in healthy young children,” *Pediatrics*, vol. 133, no. 5, pp. e1218–25, May 2014.
- [6] G. A. Poland, “The 2009–2010 influenza pandemic: effects on pandemic and seasonal vaccine uptake and lessons learned for seasonal vaccination campaigns,” *Vaccine*, vol. 28, pp. D3–D13, Sep. 2010.
- [7] H. Kim, J. Y. Han, and Y. Seo, “Effects of facebook comments on attitude toward vaccines: The roles of perceived distributions of public opinion and perceived vaccine efficacy,” *J. Health Commun.*, vol. 25, no. 2, pp. 159–169, Feb. 2020.
- [8] A. M. Sadri, S. Hasan, S. V. Ukkusuri, and M. Cebrian, “Crisis communication patterns in social media during hurricane sandy,” *Transportation Research Record*, vol. 2672, no. 1, pp. 125–137, 2018. [Online]. Available: <https://doi.org/10.1177/0361198118773896>
- [9] D. Yates and S. Paquette, “Emergency knowledge management and social media technologies: A case study of the 2010 haitian earthquake,” *Int. J. Inf. Manage.*, vol. 31, no. 1, pp. 6–13, Feb. 2011.
- [10] C. Möller, J. Wang, and H. T. Nguyen, “#strongerthanwinston: Tourism and crisis communication through facebook following tropical cyclones in fiji,” *Tourism Manage.*, vol. 69, pp. 272–284, 2018.
- [11] M. Szomszor, P. Kostkova, and E. De Quincey, “# swineflu: Twitter predicts swine flu outbreak in 2009,” in *International conference on electronic healthcare*, 2010, pp. 18–26.
- [12] L. Mollema, I. A. Harmsen, E. Broekhuizen, R. Clijnk, H. De Melker, T. Paulussen, G. Kok, R. Ruiter, and E. Das, “Disease detection or public opinion reflection? content analysis of tweets, other social media, and online newspapers during the measles outbreak in the netherlands in 2013,” *J. Med. Internet Res.*, vol. 17, no. 5, p. e128, 2015.
- [13] L. Hossain, D. Kam, F. Kong, R. T. Wigand, and T. Bossomaier, “Social media in ebola outbreak,” *Epidemiol. Infect.*, vol. 144, no. 10, pp. 2136–2143, Jul. 2016.
- [14] S. F. McGough, J. S. Brownstein, J. B. Hawkins, and M. Santillana, “Forecasting zika incidence in the 2016 latin america outbreak combining traditional disease surveillance with search, social media, and news report data,” *PLoS Negl. Trop. Dis.*, vol. 11, no. 1, p. e0005295, Jan. 2017.
- [15] L. Li, Q. Zhang, X. Wang, J. Zhang, T. Wang, T. Gao, W. Duan, K. K. Tsoi, and F. Wang, “Characterizing the propagation of situational information in social media during COVID-19 epidemic: A case study on weibo,” *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 556–562, Apr. 2020.

- [16] X. Huang, M. C. Smith, M. J. Paul, D. Ryzhkov, S. C. Quinn, D. A. Broniatowski, and M. Dredze, "Examining patterns of influenza vaccination in social media," in *AAAI Workshops*, 2017.
- [17] N. Mekawie and A. Hany, "Understanding the factors driving consumers' purchase intention of over the counter medications using social media advertising in Egypt:(A facebook advertising application for cold and flu products)," *Procedia Comput. Sci.*, vol. 164, pp. 698–705, 2019.
- [18] S. K. Gudi, M. Chhabra, K. Undela, R. Venkataraman, U. V. Mateti, K. K. Tiwari, and S. Nyamagoud, "Knowledge and beliefs towards universal safety precautions during the coronavirus disease (COVID-19) pandemic among the indian public: a web-based cross-sectional survey," *Drugs & Therapy Perspectives*, vol. 36, no. 9, pp. 413–420, Sep. 2020.
- [19] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, pp. 8869–8879, 2017.
- [20] W. W. Chapman, J. N. Dowling, O. Ivanov, P. H. Gesteland, R. Olaszewski, J. U. Espino, and M. M. Wagner, "Evaluating natural language processing applications applied to outbreak and disease surveillance," in *Proceedings of 36th symposium on the interface: computing science and statistics*, 2004.
- [21] J. W. Uys, N. D. du Preez, and E. W. Uys, "Leveraging unstructured information using topic modelling," in *PICMET '08 - 2008 Portland International Conference on Management of Engineering Technology*, Jul. 2008, pp. 955–961.
- [22] S. K. Habibabadi and P. D. Haghighi, "Topic modelling for identification of vaccine reactions in twitter," in *Proceedings of the Australasian Computer Science Week Multiconference*, ser. ACSW 2019, no. Article 31. New York, NY, USA: Association for Computing Machinery, Jan. 2019, pp. 1–10.
- [23] M.-P. S. Chan, K. H. Jamieson, and D. Albarracin, "Prospective associations of regional social media messages with attitudes and actual vaccination: A big data and survey study of the influenza vaccine in the united states," *Vaccine*, Aug. 2020.
- [24] R. Mahajan, W. Romine, M. Miller, and T. Banerjee, "Analyzing public outlook towards vaccination using twitter," in *2019 IEEE International Conference on Big Data (Big Data)*, Dec. 2019, pp. 2763–2772.
- [25] Z. Xu, "Personal stories matter: topic evolution and popularity among pro- and anti-vaccine online articles," *Journal of Computational Social Science*, vol. 2, pp. 207–220, 04 2019.
- [26] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, no. Jan, pp. 993–1022, 2003.
- [27] R. J. Steele and A. E. Raftery, "Performance of bayesian model selection criteria for gaussian mixture models," *Frontiers of statistical decision making and bayesian analysis*, vol. 2, pp. 113–130, 2010.
- [28] D. W. K. Andrews, "Consistent moment selection procedures for generalized method of moments estimation," *Econometrica*, vol. 67, no. 3, pp. 543–563, 1999. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00036>
- [29] P. Carrillo-Santistevé, B. C. Ciancio, A. Nicoll, and P. Luigi Lopalco, "The importance of influenza prevention for public health," *Human vaccines & immunotherapeutics*, vol. 8, no. 1, pp. 89–95, 2012.
- [30] B. Flannery, J. R. Chung, E. A. Belongia, H. Q. McLean, M. Gaglani, K. Murthy, R. K. Zimmerman, M. P. Nowalk, M. L. Jackson, L. A. Jackson *et al.*, "Interim estimates of 2017–18 seasonal influenza vaccine effectiveness—united states, february 2018," *Morbidity and Mortality Weekly Report*, vol. 67, no. 6, p. 180, 2018.
- [31] A. Conte, R. Quattrin, E. Filiputti, R. Cocconi, L. Arnoldo, P. Tricarico, M. Delendi, and S. Brusaferrò, "Promotion of flu vaccination among healthcare workers in an italian academic hospital: An experience with tailored web tools," *Human vaccines & immunotherapeutics*, vol. 12, no. 10, pp. 2628–2633, 2016.
- [32] Y. Watanabe, M. S. Ibrahim, Y. Suzuki, and K. Ikuta, "The changing nature of avian influenza a virus (h5n1)," *Trends in microbiology*, vol. 20, no. 1, pp. 11–20, 2012.
- [33] P. Doshi, "Are us flu death figures more pr than science?" *Bmj*, vol. 331, no. 7529, p. 1412, 2005.
- [34] J. K. Ward, P. Guille-Escuret, and C. Alapetite, "“antivaxxers” as the embodiment of anti-science," *Deviance et Societe*, vol. 43, no. 2, pp. 221–251, 2019.
- [35] M. F. Daley, L. A. Crane, V. Chandramouli, B. L. Beaty, J. Barrow, N. Allred, S. Berman, and A. Kempe, "Misperceptions about influenza vaccination among parents of healthy young children," *Clinical pediatrics*, vol. 46, no. 5, pp. 408–417, 2007.
- [36] D. Campos-Outcalt, "Cdc recommendations expand vaccine indications: children ages 5-18 should get the flu vaccine. asthma patients and smokers should get the pneumococcal vaccine," *Journal of Family Practice*, vol. 58, no. 3, pp. 146–149, 2009.
- [37] K. E. Wiley, P. D. Massey, S. C. Cooper, N. J. Wood, J. Ho, H. E. Quinn, and J. Leask, "Uptake of influenza vaccine by pregnant women: a cross-sectional survey," *Medical Journal of Australia*, vol. 198, no. 7, pp. 373–375, 2013.
- [38] T. Jefferson, D. Rivetti, A. Rivetti, M. Rudin, C. Di Pietrantonj, and V. Demicheli, "Efficacy and effectiveness of influenza vaccines in elderly people: a systematic review," *The Lancet*, vol. 366, no. 9492, pp. 1165–1174, 2005.
- [39] L. A. Grohskopf, E. Alyanak, K. R. Broder, E. B. Walter, A. M. Fry, and D. B. Jernigan, "Prevention and control of seasonal influenza with vaccines: recommendations of the advisory committee on immunization practices—united states, 2019–20 influenza season," *MMWR Recommendations and reports*, vol. 68, no. 3, p. 1, 2019.