

Perception Analysis: Pro- and Anti- Vaccine Classification with NLP and Machine Learning

Izunna Okpala Guillermo Romera Rodriguez Weibing Zheng Shane Halse Jess Kropczynski
 University of Cincinnati Penn State University University of Cincinnati University of Cincinnati University of Cincinnati
okpalaiu@mail.uc.edu gkr5144@psu.edu zhengwb@mail.uc.edu halsese@ucmail.uc.edu kropczjn@ucmail.uc.edu

Abstract

Online discussion of the ensuing pandemic exemplifies the extent and complexity of information required to understand human perception. Social media has proven to be a viable medium for identifying actionable data and analyzing public perception. As health sectors all over the world battled to obtain accurate information regarding COVID-19, this research focused on gauging public perceptions of the vaccine. The public reception of the vaccine can be determined by public perception. This study explores how to use machine learning to understand human perceptions in the context of the COVID-19 vaccine. Natural Language Processing (NLP) was employed to detect pro- and anti-vaccine tweets, while two machine learning classification models were used to study the patterns derived from the analysis. The study analyzed people's perceptions of the vaccine by presenting the results from a geographic region, while learning patterns that are likely to be associated with pro- or anti-vaccine perceptions.

1. Introduction

Local and regional health departments are actively developing public awareness strategies to encourage people to get the long-awaited COVID-19 vaccine [1] in places where vaccination rates might be low. Although social media shows promise in terms of better understanding human opinions and strategies for targeting these campaigns to areas that need them most, it is impossible for a human agent to analyze the pool of data from this medium with high accuracy and speed without the help of an automated system [2].

Previous works that used automated tools to identify human opinions heavily relied on keyword filtering approaches from within a particular discourse [3]. As such, social media platforms, especially Twitter, which receive millions of data points per second from people of diverse backgrounds, creating rich opportunities for

researchers, have been explored to identify clues to events. Each user on the platform contributes in the pool by sharing ideas, concepts, and perceptions about all aspects of life. These data in turn have been referred to as actionable data due to the derivatives from the content [4] such as concepts, business ideas, and disruption events. It shows the thoughts of people from different regions and different beliefs. In essence, these data are of particular interest to crisis informatics researchers especially when considering studies on trend analysis, situational awareness, and user perception.

The COVID-19 pandemic took the world by storm and disrupted the tech industry by way of a global shutdown, moving more human interaction online. Researchers and practitioners all over the world are working with various data across social spaces like Facebook, Twitter, and blogs to understand patterns of activity for the teeming populace, especially with regards to COVID-19 prevention measures such as mask-wearing, social distancing, and the vaccine. The long awaited vaccine was said to be ready in the fall of 2020 [1]. With this news, some citizens took to social media to describe resistance to obtaining the vaccine. Public health and community officials have worked to distribute accurate information about vaccines, but these efforts can be better targeted to areas where local perceptions of the vaccine are known. The specific aim of this research is to adopt a machine learning (ML) approach to provide officials with information about local perceptions of the COVID-19 vaccination. This study explores ML as an automated tool to gauge local perceptions in near real time from social media, and is combined with common methods like subjectivity and polarity in handling sentiments.

1.1. Research Questions

We utilize the following two research questions to guide the design of the aforementioned automated tool:

- **RQ1:** *How can an automated tool be used to understand pro- and anti- COVID-19 vaccine*

perceptions?

- **RQ2:** *How can the parameterized output from perception analysis be used to train an ML algorithm?*

2. Background Literature

A major issue that the field of information technology has grappled with is the influx of data without meaningful extrapolations from the data [5]. Additionally, it is seemingly impossible for a human agent to analyze millions of data points with high accuracy in a timely fashion [6], as is the case with social media data. The article *We are Drowning in Data* by Charlie Warzel in 2019 is a good illustration of this problem [7]. Making meaning out of data also entails contextualizing, formatting, and cleaning the data [8]. This paper, therefore, addresses contextualizing as well as extrapolating perceptions from a pool of Twitter data to make meaning out of it. In the case of the COVID-19 vaccination, scarce resources can be channeled to local areas that will make better use of them i.e., if we can tell intuitively that the chances of citizens accepting a vaccine are high, then local information campaigns may expend less effort in these areas. Conversely, when strong anti-vaccination perceptions persist, government officials can partner with health institutions to persuade a region to adopt COVID-19 preventive measures. The insights from this research can also be extended to contribute to other use cases. For example, analysts may look for ways to make predictions or tell stories in a visualized form with data about a particular phenomenon, be it a pandemic, industry trends, or simply experimenting with a decision support system.

2.1. COVID-19 Vaccination

Vaccinations are not new and have been around for many years [9], but the COVID-19 vaccination was met with controversy due to rapid production and misinformation shared online, raising safety concerns [10]. People from around the world have given their opinion on the subject matter via social platforms due to the unprecedented escalation of the pandemic and its effect on businesses, nations, and human lives [11, 12]. From the year 2020, when the outbreak hit on a large scale, medical professionals started experimenting with different medical solutions to douse the long term implications or impacts of the virus [13]. The COVID-19 vaccination, a proposal by major pharmaceutical industries in the US, was acknowledged as the best solution to reduce the virological effect of the virus [14]. We hence focused

our efforts on understanding the overall perception of the COVID-19 vaccination to provide more actionable data and help institutions and organizations prepare their for resistance.

2.2. Social Media Data as Actionable Data

Social media data has been identified by researchers as a relevant source of information for various domains, from crisis management [15] to social behavior, and more [16, 17]. Actionable data, as the name suggests, is data that contains important elements that can be acted upon or information that gives extra insights than conventional numeric data [18]. Nonetheless, a major problem is transforming social media data into actionable data that can be used by a major stakeholders during an ongoing event [19]. To overcome this obstacle, there are many different methodologies that have evolved over time [20]. In fact, data processing takes up different phases in data analysis - depending on the nature of the data and the attributes thereof. However, among all the different approaches, there are some common practices that are more or less shared. These common practices are generally applied during the pre-processing stage of data analysis. These are: *Tokenization, case conversion, removing punctuation, white spaces and lower-casing the text* [21]. While these steps are common in many domains, they are the first step in transforming noisy social media data into actionable information and data. Once social media data has been effectively pre-processed and cleaned, it offers an unparalleled benefit over other types of data, i.e., the data can be used on an almost real-time basis [22, 23]. After the data has been analyzed, its use and practicality are beyond what other types of data can offer [24]. Its use ranges from crisis management [25] to examining social behavior during times of unrest [26] as well as consumer behavior [27]. In this vein, examining public perception towards vaccination is not an exception [28]. However, we believe this unprecedented scenario can help us learn more about public attitudes and understanding of vaccination.

2.3. Fundamentals of Human Perceptions

Identifying perceptions about a particular discourse is often the most striking proof that some speculations are gaining momentum and can either be true or false [29], and the major source is social platforms. The derived evidence can be interpreted as either objective or subjective, making *social perception* tilt towards recognizing and using social cues from a sentence to make decisions about social views, associations, significance, or the attributes (e.g., trustworthiness) of

a phenomenon (vaccine) [30]. A scholarly definition of perception is the process by which individuals organize and interpret their sensory impressions in order to give meaning to their environment [31]. According to Faronbi et. al. [32], there are some similitudes between perceptions and attitudes. Whereas perceptions describe the organization and interpretation of a certain external stimulus through observations and word structure, attitudes are more of demonstrative constructs and behavioral - showing a clear distinction in terms of describing favorable or unfavorable scenarios through action. Our study considers text manipulation and not the physical behavioral translations through body movements and gesticulation. In this, observations are the metadata of social perception, based on a three-way interaction between people, circumstances, and attitudes [33]. Basic human perceptions are rooted in heuristics [34]. This informs the pattern of communication, thereby providing deeper meaning for data. It also involves the identification of words that have deeper meaning than others, especially those related to emotions. Nonetheless, for us researchers, identifying and analyzing all the different social perceptions and attitudes is not feasible without the use of natural language processing or training a machine learning model to understand patterns, and even then, this can be a difficult task [35]. The English text will be divided into parts called tokens where the individual words are analyzed, combined, and scripted as either pro or anti a phenomenon using natural language processing.

2.4. Natural Language Processing(NLP)

Building an intelligent system, as demonstrated in this study, requires interpreting a language via its natural means. NLP is a method of automatically manipulating natural languages to inform a decision [36]. This method relies on standard techniques for obtaining meaning from text data. As mentioned earlier, these techniques involve the process of data cleaning and pre-processing. However, to truly obtain meaning via a given text to inform a decision generally requires machine learning application [37]. One common usage of these NLP is in the detection of fake news and with evident success in this domain [38]. In addition, NLP has also been used to detect possible signs of radicalization within these social media platforms, and even detect hate speech [39]. In conclusion, there are a multitude of approaches to obtaining meaning from text data through NLP techniques. Some of them are more common and accessible, such as HateSonar, and LIWC [40] while others provide a state-of-the-art [41] approach. However, each one of these approaches has

their own limitations and constraints, and this is why creating a custom-made approach can be beneficial.

2.5. The Role of ML in Understanding Perceptions

Machine Learning has a long history, and modern advancements have made it more ubiquitous than ever, and it adopts the traditional style of learning. Conventional learning, as defined in Psychology, is the acquisition of knowledge through experience or study, thereby populating the human brain with the desired information [42]. Machine learning, therefore, is partly based on the model of brain cell interaction. The origin can be dated back to the 1940s when Donald Hebb published a book that presented his theories on neuron excitement and communication between neurons titled *The Organization of Behavior* [43]. In its early days, ML showed how a computer could play and outwit humans in the game of checkers [44]. This was done using a Perceptron program designed by Frank Rosenblatt in 1957 that combines the properties of Hebb's neuron functionalities with those of the algorithm in the Checkers game [45]. These days, those techniques are still useful and have proved very fruitful. That is why, after traversing through all its various phases, ML has gained momentum in many fields, specifically in the IT world. Since those early days, there have been many advances in the use of ML to solve both basic and complex human problems. One of such problems, that is related to the goal of our paper, is that of deciphering human perceptions of discourse. Substantial evidence has been shown to encourage the use of ML techniques and mechanisms to predict outcomes from social media. For example, Starbird, et. al., featured a Support Vector Machine (SVM) algorithm for identifying the likelihood that a Twitter user was present during a mass disruption event [46]. In addition, in 2015, Li et al. [47] carried out such research using the Naïve Bayes algorithm in order to prioritize tweets based on some defined classification criteria. These and other examples have shown the potential of ML in solving more human problems. Therefore, it is why our approach focuses on using ML to label unlabelled tweets based on different levels of objectivity and subjectivity, to then predict the likelihood of someone taking the vaccine, as opposed to its unlikelihood.

3. Methodology

This study addresses the two research questions by the use of Natural Language Processing (NLP) and machine learning classification methods. First,

evaluating the audience reaction to the COVID-19 vaccine, and second, what criteria can be used to detect vaccination-related perception. These two questions are aimed to be addressed thanks to advances in the use of NLP. Modern approaches in NLP allow for a greater understanding of people’s opinions either to detect attitudes, sentiments or personal opinion [48]. Within the context of our study, and our research questions, we are only interested in English texts as our NLP approach organizes tweets in relation to the language’s interjecting(linking) words like - and, but, although, etc. Additionally, given that some tweets can contain spelling errors, our study puts a check on how spelling errors would be corrected without human intervention.

3.1. Data Collection

Our primary data was sourced from Twitter using the new Academic Twitter API which requires personalized authentication and human verification to gain access. The researchers focused on tweets that were captured around the period that the COVID-19 vaccine topic gained momentum in the public space - September 2020 to May 2021. The bounding box coordinates of [-84.731712, 39.038386, -84.379807, 39.179046] used for collecting the tweets was derived using an online bounding box calculation program called *bbox finder* that joins all geolocations within a selected area as a point of interest - the tweets were curated by concentrating only on a 24miles radius within the Cincinnati region. The total number of the tweets collected was 876.

3.2. Data Pipeline

Considering the nature of the study and the deliverable expected, the authors divided the pipelining architecture into four steps.

- Data Ingestion
- Data Transformation and processing
- Data Categorization
- Training a Machine Learning Model

3.2.1. Data Ingestion This involves spooling the raw data from the collection funnel (Twitter) to our program for processing and analysis. In this, the raw data was converted to the UTF-8 Unicode type and presented via the JSON format to the python’s pandas package for easy manipulation.

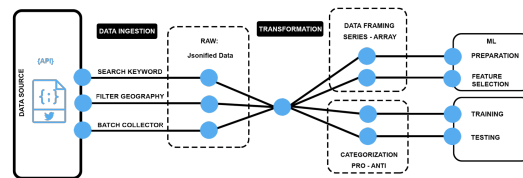


Figure 1. Data pipeline architecture

3.2.2. Data Transformation As the study intends to cycle through different perceptions within a single tweet by splitting individual tweets into parts based on linking (interjecting) words the data transformation stage involves several stages. The initial stage in the data transformation is text preprocessing on a minimal scale - removing the RT keyword and some characters like ”(” and ”)” that conflicts with the python program semantics and ensure the least alteration of our text. With this in place, the next stage ensured that the tweets are set for proper differentiation via splitting. Giving the desired result intended for the study, common words (stop words) in the English dictionary that has no perception connotations were removed as well to give room for fast computation of tweet perception. Because name of entities like cities and people contain no perception but can give a clue to the entities by which a particular attitude is directed towards, those that were discovered in individual splits were named and stored separately in a list for directionality and perception description. However, these were removed from the *tweet* to give an overall cleaner and accurate results. The second part of text cleaning consists on the removal of punctuation and special characters in the tweets. This is followed by the spelling correction that makes sure all words represented in individual tweets are captured with the correct spelling. To do so, we employ the *suggest* function within the Python’s *pattern* package which contains a dictionary of known English words and also identifies the relationship of the words in a text. Lastly, with the help of a well-scripted *max* function, our study was able to marry the words in a tweet by choosing the tokens with the maximum probability based on the suggestion made by the *suggest* function. Once we conducted this transformation our dataset was ready for data processing and analysis.

3.2.3. Data Categorization As per our RQ1, which involves the evaluation of an audience’s reaction towards tweets based on either a pro- or anti-vaccine perception. Like the data transformation stage, this stage is comprised of several steps in order to best decipher

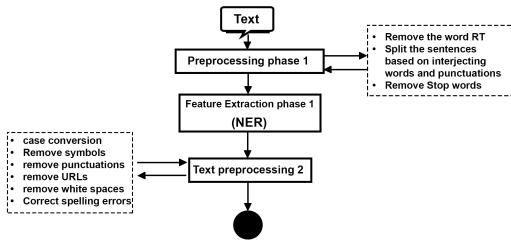


Figure 2. Data Transformation and Processing

the correct perception within the text. To do so, first we tokenized our transformed data with with the *NLTK* Python's library. Afterward the *TextBlob* analyzer was used to detect words that could suggest positive or negative attitudes towards the COVID-19 vaccine based on their polarity scores. In addition, this tool also enhances the study by collectively detecting personal opinion and facts based on word combinations and information that stresses more/less emotions.

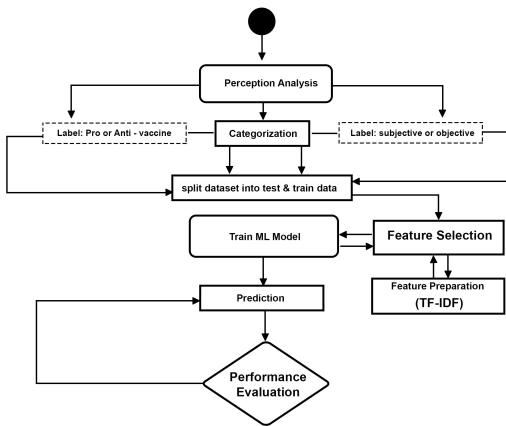


Figure 3. Data categorization and ML

3.2.4. Training a Machine Learning Model This stage utilizes the knowledge derived from the three stages above to train the *Naive Bayes* and *SVM* machine learning models, via the Python classification package called *scikit-learn*. The data input was based on the processed tweets from the transformation stage. This gives a correct mapping as per the exact English words that produce a particular derivative. The entire data is split into training and testing data via the 80:20 rule with a random state that helps the data maintain results even after splitting. These data are parsed for tokenization, i.e., creating a dictionary of features while transforming the data to feature vectors. To avoid potential discrepancies in the storage of the feature

vectors, our program made use of the *TfidfTransformer* function to compute the Inverse Document Frequency used in preventing memory lag and also for the smooth training and testing of our data. The accuracy of the SVM in the analysis was 70% while that of the Naive Bayes was 65%. It shows that a hyperplane adoption of the SVM is better than the probabilistic model of the Naive Bayes in this regard. This accuracy could change in the future if the training dataset is considerably larger. The step-by-step demo of the process is in Figure 3.

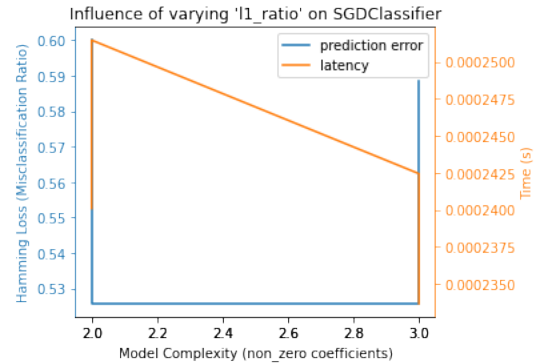


Figure 4. SVM Complexity Influence (a)

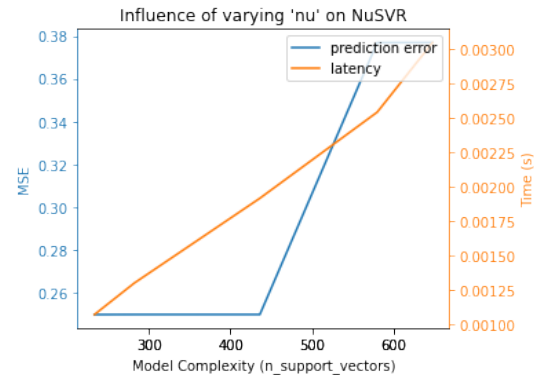


Figure 5. SVM Complexity Influence (b)

Due to a higher accuracy with the SVM algorithm, additional test to measure the latency and prediction error of the algorithm was conducted. For this to be done, the testing and training data were further converted from dataframe to a numpy array and fed into a parameterized SVM, NuSVR, and Gradient Boosting Regressor pipeline that contains the *vectorization* function and another *TfidfTransformer* function. For the data to fit well, the numpy array was reshaped and the texts were transformed to numbers using label encoder. The Figure 4-6 shows the complexity influence of the SVM algorithm and how all references were accommodated by the algorithm. The results are based

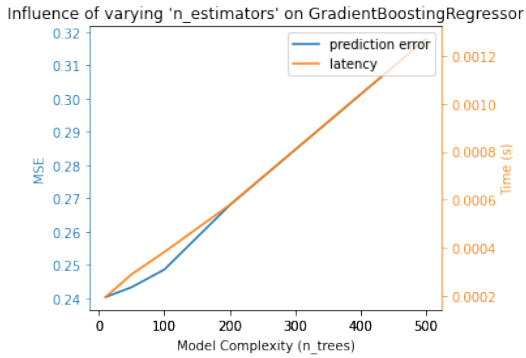


Figure 6. SVM Complexity Influence (c)

on non zero coefficients, number of support vectors and number trees(child) respectively. The SVM produced a slight decline in latency, but a consistent prediction error at the lowest point. The complexity of the model continually increases with the predictive power (MSE or Hamming loss) and the computational time in Figure 5 and Figure 6. Though the overall accuracy of the model was 70%, the major factor for this rate was the amount of data sent to the machine learning model.

4. Results

The result in Figure 7 shows an aggregation of the perception in the tweets collected from September 2020 to May 2021 within the Cincinnati region having the vaccine keyword. It shows that a greater percentage of the people in Cincinnati are pro-vaccine. Additionally, some of the tweets that were captured as neutral correlates with the tweets that were marked as subjective giving a clue that may lead to an inference. I.e., a sentence that is not one-sided can be based on facts and not human opinions. This is owing to the narrative that human opinions are frequently based on a biased perspective [49], but not all personal opinions are one-sided [50] - the reason why all subjective tweets were not reported as neutral.

Using a scatter plot to represent the relationship between perception and subjectivity as demonstrated in Figure 8, gives an abstract view of the non-aggregated data. From the plot, we can see that less than 10% of the data are based on facts. meaning that the majority of tweets are personal opinions. Additionally, the plot shows no particular digression in perception, though the large volumes were tilted towards the right.

To understand this better, a time series analysis of how the perception travels is demonstrated in Figure 9. This shows that perception is more concentrated on the positive axis as against the negative axis, confirming the aggregation in Figure 7. Additionally, the time series

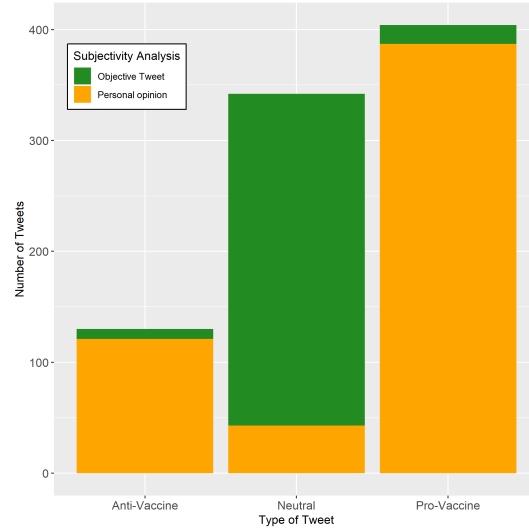


Figure 7. Perception VS Subjectivity Analysis

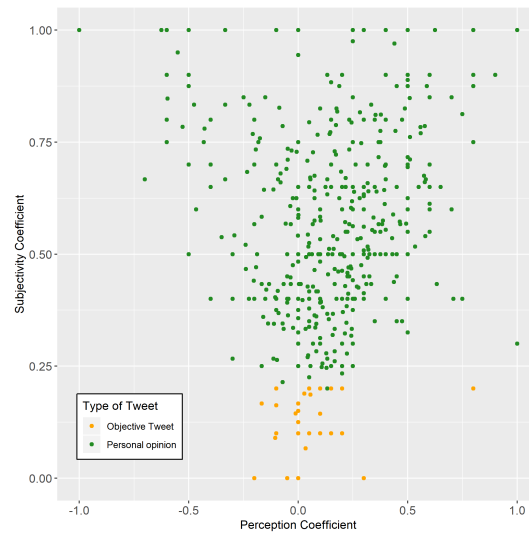


Figure 8. Non-aggregated scatter plot for the Perception Analysis

as shown in Figure 9 accommodating the perception index, demonstrates that there was scarcely a vaccine perception from the early days of the discussion (October 2020) about the vaccine. This could be due to level of awareness about the vaccine at the time. But from January through March, there were more people talking about the vaccine and its impact, because evidential proof from Figure 9 shows that the perception index kept increasing. This can be corroborated by the KFF COVID-19 Vaccine Monitor [51] showing the percentage number of people that have been vaccinated and the number of people that have okayed the vaccine.

Further consideration of the most popular words

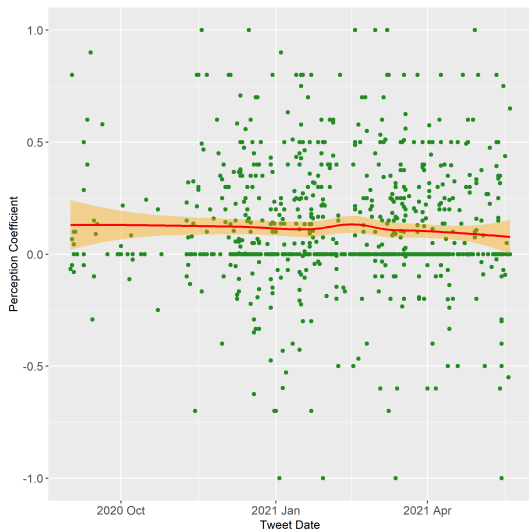


Figure 9. Time series representation of the perception index

in the collected tweets is presented as a word cloud in Figure 10. From the screenshot, the majority of the words captured are less of subjective information i.e., showing less emotional connotation. Though, a vast majority of the pro-vaccine tweets were marked as personal opinion from Figure 7, the neutral points were 98% objective.

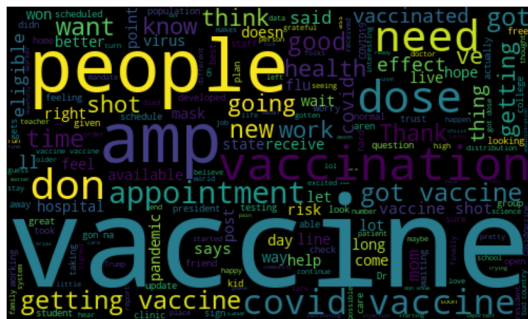


Figure 10. Word Cloud of popular mentions

5. Discussion

The two research questions were both addressed using a methodical approach demonstrated in the methodology section: evaluating an audience's reaction to the COVID-19 vaccine as well as the criteria that reveal vaccination-related perceptions in tweets were the critical areas. In deciphering perception in tweets, a high percentage of the tweets from the Cincinnati region results in a pro-vaccine stance rather than the anti-vaccine stance. It was also discovered that

some attributes like personal opinions influence peoples thoughts and perception. For example, the high level of unanimity in the collected tweets can be attributed to four different factors, which are: the situation on-site, personal opinions about the subject, and an individual's perspective. These perspectives can be influenced by internal beliefs and external environment.

5.1. Limitations

A number of reasons may have limited this research. First, we considered data from the twitter platform alone and the number of tweet locations we were able to get was limited by the Twitter API geolocation. Our sample only includes those who have their geolocation feature active - this is only true for a small percentage of the users. One possible option is to query the whole Twitter archive as against using a particular geography. Fetching the whole archive without limiting our search to a particular point of interest would pose difficulties and consume a significant amount of time, hardware usage, and memory space.

Another limitation of this study is insufficient data for training the machine learning model, and this was due to the limitation posed by the amount of data collected. Though our area of concentration is the Cincinnati area which limits generalizability, to better learn patterns across diverse viewpoints we would require large dataset and multiple geolocations.

5.2. Future Work

Due to the limitations discussed in the previous section, further research is needed to improve the accuracy and reliability of Twitter geolocation, which is an important component when collecting and analyzing social media data and can be useful to researchers. Furthermore, this research was limited to a single city. A larger study spanning numerous locations is needed to allow for better generalization.

Another aspect that was not addressed is the negating factors prevalent in the English vocabulary, which might be misunderstood by the machine, thereby exposing the aggregated data to a slightly overlooked fraction of errors. Also, though the analysis provided room for separating entities from texts, the named entities did not affect the analysis because they were extracted out to give room for faster analysis. These named entities will be valuable when addressing issues related to multiple entity annotation within a given sentence, and we're excited about investigating this in future research. The multiple steps in the data pipelining section can be improved further with the BERT model to reduce the time taken for classifying texts, making it possible to

separately measure the efficacy of vaccines based on the types.

6. Conclusion

This study looked at how social media can inform regional health agencies in charge of the vaccination about the public perception in their locality. The tweets collected centers around the Cincinnati area, and according to our result 46.1% of the tweets collected were pro-vaccine while 14.8% were against the vaccine and the others remaining neutral. A similar study by Sejal Dua [52], that adopted a kaggle dataset of all COVID-19 Vaccine tweets, presents a result of 14.8% negative tweets against 38.4% positive tweets; meaning that our study is right on track considering the phased approach that was demonstrated in the methodology section for higher accuracy in data analysis. Further analysis with the aggregated data provides us with ample evidence that a greater percentage of these perceptions were based on personal opinions while over 90% of the neutral tweets were objective - in simpler terms, people tend to be neutral when talking about facts than given personal opinion.

Training an SVM algorithm was promising to be able to avoid the rigorous steps involved in the NLP process, though the results for training the model were a product of the NLP. Overall, our model produced an accuracy of 70%, but the promising factor is the consistent degree to which the prediction error flattened across the training and testing process - suggesting that higher accuracy is guaranteed if more data was fed into the system.

Thus according to our findings, categorizing tweets based on either pro- or anti-vaccine can be valuable for investigating the connections between the availability of vaccines in a particular location and the efficacy of the vaccine. In summary, with the available data curated from September 2020 to May 2021, the study suggests that people are keen on having a solution to the dreaded coronavirus as opposed to a negative opinion towards exercise.

References

- [1] W. Brothers, "A Timeline of COVID-19 Vaccine Development," *BioSpace*, Dec 2020.
- [2] O. Sturman, L. von Ziegler, C. Schläppi, F. Akyol, M. Privitera, D. Slominski, C. Grimm, L. Thieren, V. Zerbi, B. Grewe, *et al.*, "Deep learning-based behavioral analysis reaches human accuracy and is capable of outperforming commercial solutions," *Neuropsychopharmacology*, vol. 45, no. 11, pp. 1942–1952, 2020.
- [3] S. Wiggins and J. Potter, "Attitudes and evaluative practices: Category vs. item and subjective vs. objective constructions in everyday food assessments," *British journal of social psychology*, vol. 42, no. 4, pp. 513–531, 2003.
- [4] P. Barnaghi, A. Sheth, and C. Henson, "From data to actionable knowledge: Big data challenges in the web of things [guest editors' introduction]," *IEEE Intelligent Systems*, vol. 28, no. 6, pp. 6–11, 2013.
- [5] S. Kaisler, F. Armour, J. A. Espinosa, and W. Money, "Big data: Issues and challenges moving forward," in *2013 46th Hawaii international conference on system sciences*, pp. 995–1004, IEEE, 2013.
- [6] J. Van Dijck, "Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology," *Surveillance & society*, vol. 12, no. 2, pp. 197–208, 2014.
- [7] C. Warzel, "Opinion | We Are Drowning in Data," *N.Y. Times*, Jun 2019.
- [8] R. Deveaud and F. Boudin, "Effective tweet contextualization with hashtags performance prediction and multi-document summarization," in *INitiative for the Evaluation of XML Retrieval (INEX)*, pp. n–a, 2013.
- [9] R. McMurray, F. M. Cheater, A. Weighall, C. Nelson, M. Schweiger, and S. Mukherjee, "Managing controversy through consultation: a qualitative study of communication and trust around mmr vaccination decisions," *British Journal of General Practice*, vol. 54, no. 504, pp. 520–525, 2004.
- [10] M. Thelwall, K. Kousha, and S. Thelwall, "Covid-19 vaccine hesitancy on english-language twitter," *Profesional de la información (EPI)*, vol. 30, no. 2, 2021.
- [11] X. Wu, R. C. Nethery, B. M. Sabath, D. Braun, and F. Dominici, "Exposure to air pollution and covid-19 mortality in the united states," *MedRxiv*, 2020.
- [12] D. Gruda and A. Ojo, "Inferring the relationship between anxiety and extraversion from tweets during covid19—a linguistic analytics approach," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, p. 2689, 2021.
- [13] K. Søreide, J. Hallet, J. B. Matthews, A. A. Schnitzbauer, P. D. Line, P. Lai, J. Otero, D. Callegaro, S. G. Warner, N. N. Baxter, *et al.*, "Immediate and long-term impact of the covid-19 pandemic on delivery of surgical services," *The British journal of surgery*, 2020.
- [14] M. Jeyanathan, S. Afkhami, F. Smaill, M. S. Miller, B. D. Lichty, and Z. Xing, "Immunological considerations for covid-19 vaccine strategies," *Nature Reviews Immunology*, vol. 20, no. 10, pp. 615–632, 2020.
- [15] C. Reuter, A. L. Hughes, and M.-A. Kauffhold, "Social media in crisis management: An evaluation and analysis of crisis informatics research," *International Journal of Human-Computer Interaction*, vol. 34, no. 4, pp. 280–294, 2018.
- [16] C. Li, L. J. Chen, X. Chen, M. Zhang, C. P. Pang, and H. Chen, "Retrospective analysis of the possibility of predicting the covid-19 outbreak from internet searches and social media data, china, 2020," *Eurosurveillance*, vol. 25, no. 10, p. 2000199, 2020.
- [17] W. He, X. Tian, Y. Chen, and D. Chong, "Actionable social media competitive analytics for understanding customer experiences," *Journal of Computer Information Systems*, vol. 56, no. 2, pp. 145–155, 2016.

- [18] J. L. Hellerstein, S. Ma, and C.-S. Perng, "Discovering actionable patterns in event data," *IBM Systems Journal*, vol. 41, no. 3, pp. 475–493, 2002.
- [19] J. L. Chan and H. Purohit, "Challenges to transforming unconventional social media data into actionable knowledge for public health systems during disasters," *Disaster medicine and public health preparedness*, vol. 14, no. 3, pp. 352–359, 2020.
- [20] P. Koukaras and C. Tjortjis, "Social media analytics, types and methodology," in *Machine Learning Paradigms*, pp. 401–427, Springer, 2019.
- [21] L. Derczynski, A. Ritter, S. Clark, and K. Bontcheva, "Twitter part-of-speech tagging for all: Overcoming sparse and noisy data," in *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*, pp. 198–206, 2013.
- [22] S. Shan, F. Zhao, Y. Wei, and M. Liu, "Disaster management 2.0: A real-time disaster damage assessment model based on mobile social media data—a case study of weibo (chinese twitter)," *Safety science*, vol. 115, pp. 393–413, 2019.
- [23] Y. Gu, Z. S. Qian, and F. Chen, "From twitter to detector: Real-time traffic incident detection using social media data," *Transportation research part C: emerging technologies*, vol. 67, pp. 321–342, 2016.
- [24] J. Oliverio, "A survey of social media, big data, data mining, and analytics," *Journal of Industrial Integration and Management*, vol. 3, no. 03, p. 1850003, 2018.
- [25] A. Saroj and S. Pal, "Use of social media in crisis management: A survey," *International Journal of Disaster Risk Reduction*, p. 101584, 2020.
- [26] R. Korolov, D. Lu, J. Wang, G. Zhou, C. Bonial, C. Voss, L. Kaplan, W. Wallace, J. Han, and H. Ji, "On predicting social unrest using social media," in *2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, pp. 89–95, IEEE, 2016.
- [27] C. F. Hofacker, E. C. Malthouse, and F. Sultan, "Big data and consumer behavior: Imminent opportunities," *Journal of consumer marketing*, 2016.
- [28] A. G. Dunn, J. Leask, X. Zhou, K. D. Mandl, and E. Coiera, "Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: an observational study," *Journal of medical Internet research*, vol. 17, no. 6, p. e144, 2015.
- [29] K. D. Landreville and C. Niles, "'and that's a fact!': The roles of political ideology, psrs, and perceived source credibility in estimating factual content in partisan news," *Journal of Broadcasting & Electronic Media*, vol. 63, no. 2, pp. 177–194, 2019.
- [30] J. J. Skowronski and D. E. Carlston, "Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases.," *Journal of personality and social psychology*, vol. 52, no. 4, p. 689, 1987.
- [31] R. Gilorkar, "Organisational Behaviour Perception," Jul 2019. [Online; accessed 10. Mar. 2021].
- [32] J. O. Faronbi, O. Adebawale, G. O. Faronbi, O. O. Musa, and S. J. Ayamolowo, "Perception knowledge and attitude of nursing students towards the care of older patients," *International journal of Africa nursing sciences*, vol. 7, pp. 37–42, 2017.
- [33] R. Li, J. Crowe, D. Leifer, L. Zou, and J. Schoof, "Beyond big data: Social media challenges and opportunities for understanding social perception of energy," *Energy Research & Social Science*, vol. 56, p. 101217, 2019.
- [34] G. Pogarsky, S. P. Roche, and J. T. Pickett, "Heuristics and biases, rational choice, and sanction perceptions," *Criminology*, vol. 55, no. 1, pp. 85–111, 2017.
- [35] J. Murphy, M. W. Link, J. H. Childs, C. L. Tesfaye, E. Dean, M. Stern, J. Pasek, J. Cohen, M. Callegaro, and P. Harwood, "Social media in public opinion research: Executive summary of the aapor task force on emerging technologies in public opinion research," *Public Opinion Quarterly*, vol. 78, no. 4, pp. 788–794, 2014.
- [36] S. Verma, S. Vieweg, W. Corvey, L. Palen, J. Martin, M. Palmer, A. Schram, and K. Anderson, "Natural language processing to the rescue? extracting "situational awareness" tweets during mass emergency," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 5, 2011.
- [37] E. Cambria and B. White, "Jumping nlp curves: A review of natural language processing research," *IEEE Computational intelligence magazine*, vol. 9, no. 2, pp. 48–57, 2014.
- [38] S. I. Manzoor, J. Singla, *et al.*, "Fake news detection using machine learning approaches: A systematic review," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 230–234, IEEE, 2019.
- [39] T. Davidson, D. Warmsley, M. Macy, and I. Weber, "Automated hate speech detection and the problem of offensive language," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 11, 2017.
- [40] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates*, vol. 71, no. 2001, p. 2001, 2001.
- [41] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [42] M. A. Conway, J. M. Gardiner, T. J. Perfect, S. J. Anderson, and G. M. Cohen, "Changes in memory awareness during learning: the acquisition of knowledge by psychology undergraduates.," *Journal of Experimental Psychology: General*, vol. 126, no. 4, p. 393, 1997.
- [43] F. K. D., "A brief history of machine learning."
- [44] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of research and development*, vol. 3, no. 3, pp. 210–229, 1959.
- [45] T. M. Khoshgoftaar, E. B. Allen, J. P. Hudepohl, and S. J. Aud, "Application of neural networks to software quality modeling of a very large telecommunications system," *IEEE Transactions on Neural Networks*, vol. 8, no. 4, pp. 902–909, 1997.
- [46] K. Starbird, G. Muzny, and L. Palen, "Learning from the crowd: Collaborative filtering techniques for identifying on-the-ground twitterers during mass disruptions.," in *ISCRAM*, Citeseer, 2012.
- [47] H. Li, N. Guevara, N. Herndon, D. Caragea, K. Neppalli, C. Caragea, A. C. Squicciarini, and A. H. Tapia, "Twitter mining for disaster response: A domain adaptation approach.," in *ISCRAM*, 2015.

- [48] B. Liu, *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press, 2020.
- [49] A. Das and D. Grover, "Biased decisions on it outsourcing: how vendor selection adds value," *Journal of Business Strategy*, 2018.
- [50] E. R. Stafford, "Let me show you i'm not biased! demonstrating non-prejudiced opinions while navigating the topic of race," 2017.
- [51] H. Liz, L. Lunna, S. Grace, S. Mellisha, and B. Mollyann, "KFF COVID-19 Vaccine Monitor: April 2021," Aug 2021. [Online; accessed 26. Aug. 2021].
- [52] S. Dua, "Sentiment Analysis of COVID-19 Vaccine Tweets - Towards Data Science," *Medium*, Mar 2021.