Paper ML12

Topic and Sentiment Analysis of Anti-Vaccine activism using Natural Language Processing and Large Language Models

Tamas Bosznay, Katalyze Data, Witney, United Kingdom

ABSTRACT

Anti-Vax activism and vaccine hesitancy have always been existing phenomena in modern society; however, they became more prevalent in recent years, mostly related to the COVID-19 pandemic. To understand the factors around and to improve vaccine uptake, researching the public sentiment and opinion on COVID-19 vaccines is a useful first step.

This study uses Social Media data (mainly datasets from X – formerly known as Twitter) to perform topic analysis around vaccine hesitancy. The relevant English language tweets are labelled against topics using various NLP techniques such as LDA and NMF. As an alternative method, various generic (GPT-4, Claude3, Llama3) and task-specific Large Language Models will be used. Following the visualisation of topics, sentiment analysis will also be performed both using more traditional sentiment classifiers and LLMs (BERT, VADER, TextBlob). The results specifically for the use case of anti-vaccine activism will be presented at the end of the paper.

1. INTRODUCTION

Anti-vaccine resistance has become one of the major challenges of modern public health in recent decades, especially during the COVID-19 pandemic. The phenomenon involves people, groups or movements who reject or are reluctant to use vaccines despite their proven efficacy and safety. Vaccine resistance manifests itself in many forms, from complete rejection to delayed vaccination. This attitude can be attributed to many factors, including scientific misunderstandings, mistrust, religious or cultural beliefs, and misconceptions spread on the Internet and social media.

Anti-vaccine is not a new phenomenon, it is almost as old as the history of vaccinations. In the early 1800s, when Edward Jenner discovered the smallpox vaccine, it already had its opponents. The fear that the vaccine could be dangerous and the concern about disrupting people's natural immune systems had already emerged at that time. Over the years, anti-vaccine arguments have gone through many waves, but they have always been opposed by public health data that have consistently demonstrated the effectiveness and importance of vaccines.

An important catalyst for modern anti-vaccination is the spread of misinformation. The Internet, especially social media, plays a significant role in giving people quick access to news and content that questions the effectiveness of vaccines. Andrew Wakefield's 1998, since retracted and widely condemned, study falsely linking the MMR vaccine (measles, mumps, rubella) to autism gave the movement a major boost. Although the scientific community has since thoroughly debunked this link, the study continues to have significant implications for vaccine confidence. There are several reasons behind anti-vaccine. Distrust of authorities is one of the potential reasons as many people are sceptical of government institutions or pharmaceutical companies and believe that vaccines are promoted for financial gain rather than for the public good. Misinformation and disinformation are also a potential reason as fake news about vaccines spreads quickly on social media, and many people accept it without verifying its authenticity. Religious and cultural beliefs can also contribute to anti-vaccine as certain religious or cultural groups have historically opposed medical interventions, including the use of vaccines. In these communities, the rejection of vaccination is often part of religious traditions and teachings. Personal freedom also can play a role here as the emphasis on individual freedoms also contributes to anti-vaccine. Some feel that being forced to vaccinate violates their personal rights and would prefer to make their own health decisions.

A direct consequence of anti-vaccination is that it increases the risk of the return of preventable diseases. For example, diseases such as measles, which were temporarily suppressed in many parts of the world, have re-emerged in low-vaccination areas. This puts not only unvaccinated individuals at risk, but also the most vulnerable members of society, such as the elderly and infants.

Anti-vaccine opinions can also undermine herd immunity, which develops when a large part of the population is vaccinated, thus preventing the spread of disease. When vaccination rates drop, community immunity weakens and the disease spreads more easily in the event of an outbreak.

In the age of internet, forums and social media, public opinion can be influenced by campaigns which can be initiated and run by many entities, such as companies, NGOs or governments. Public opinion can also be influenced by content produced and made viral via individuals and their networks, this content can be their posts, tweets and retweets or other content on social media or other forums. Also specifically, bad actors are also playing a role here to influence and bias the public opinion in their own interest to achieve direct or indirect objectives. Understanding and profiling public opinion can help fight negative opinions, such as vaccine hesitancy or anti-vaccine.

Social media platforms, especially Twitter, provide a vast amount of user-generated data that is a rich source of information for understanding public opinion. Due to the brevity and informal nature of Twitter posts, they are particularly suitable for the application of various methods of Natural Language Processing (NLP) and data mining. In this study, we analyse Twitter posts using various methods including natural language processing, topic modelling and sentiment analysis. We also include publicly available Large Language Models for topic modelling and sentiment analysis.

2. OBJECTIVES

In our analysis, we aim to research and answer the following main objectives:

- Can we use natural language processing to detect and profile public opinion (on Social Media)
- Can we use more traditional tools as well as LLMs?
- Use topic modelling techniques in uncovering opinions
- Analyse sentiment
- Try to categorise stance to see anti- or pro-vaccines
- Can we compare how public opinion has changed since COVID-19?

3. LITERATURE REVIEW

Studying public opinion was a well-researched area even before the age of Social media platforms, however these platforms provide data on people expressing their views and opinions in vast amounts. Especially in the years 2010 to 2020 a wide variety of studies used social media data and applied Natural Language Processing and text analytics methods combined with Machine Learning to uncover insights.

Resnik et al. (2015) studied the language used by people on Twitter in relation to whether they do or do not have depression. Their methodology both included unsupervised techniques such Latent Dirichlet Allocation (LDA) but also its supervised version (SLDA).

Mu et al. (2023) created VaxxHesitancy, a dataset for studying hesitancy towards COVID-19 vaccination. They have human annotated their dataset to flag whether texts are anti-vaccine, pro-vaccine, vaccine hesitant or neutral. This human annotated dataset can be used for creating supervised learning algorithms and classifiers on the subject.

Mandot et al. (2023) researched the side effects of antidepressants using Natural Language Processing and topic modelling on Twitter data. Their approach was mainly unsupervised in order to uncover patterns in tweets about people reporting the side effects they perceive. In their study they have compared their results with reported side effects of the same set of medicines from medical sources. This is a useful technique when data is available from a data source of a different nature to validate the analysis.

One particular aspect of vaccine hesitancy is misinformation. Various studies including Dodson et al. (2021) and Gisondi et al. (2022) researched this aspect, both on longitudinal datasets and more focused corpuses, in some cases in specific communities. Communities in Social Media and society in general are a very important factor in influencing people's opinions with a variety of options for research.

Research focuses not just on the English language but other languages to uncover potentially different specifics.

4. METHODOLOGY

In terms of the research and analysis methodology, we have followed a widely used process of preprocessing textual data. The main steps are discussed here below.

4.1. NATURAL LANGUAGE PROCESSING STEPS

Tokenization is a fundamental step in natural language processing, in which raw text is broken down into smaller units. These units or tokens can be words, stems, or even characters, depending on the level of analysis. The goal of tokenization is to make text easier for computers to process, allowing further analysis steps such as word counting, sentence interpretation, or text classification to be performed. One of the main challenges of NLP is that the text that people naturally use is often extremely complex in structure, full of complex grammatical rules, words that occur in different forms, and many different contextual meanings. With the help of tokenization, we break down this complicated structure into simpler units, so that it is easier to handle the text in further processing steps.

Removal of stop words is a fundamental step in natural language processing that aims to filter out frequent words that add little or no value to text analysis. Stop words are typically frequently occurring elements such as "the", "is", "and", "of". Although these words are important for the interpretation of sentences, they are often just noise in automated text analysis, so their removal can improve the accuracy and efficiency of analysis results.

Stemming is a natural language processing (NLP) technique that aims to identify the roots of words. When conjugating words, we simplify the words to their basic form, leaving out modifications resulting from conjugation, plurals, suffixes and other suffixes. The process helps to unify the text, as words that occur in different forms can be managed through a common root, which facilitates search, analysis and categorization. As opposed to stemming, lemmatization is a more sophisticated method which, after grammatical analysis, traces words back to their basic dictionary form (lemma).

4.2. FEATURE EXTRACTION TECHNIQUES

We have considered and used a couple of alternative techniques to extract textual features from the tweets analysed. These different methods are widely used and have different advantages and disadvantages when being used with further text analysis techniques. In our analysis we have done topic modelling and sentiment analysis as the further steps so the feature extraction techniques used can influence the results of the later analysis steps. In the following we list the different feature extraction techniques used in our analysis.

The Bag of Words (BoW) method is one of the simplest and most frequently used techniques in natural language processing for the numerical representation and analysis of texts. The essence of the Bag of Words method is to represent texts as a set of words (or tokens) occurring in a document, while ignoring the order and grammatical structure of the words. The resulting vectors or matrices basically reflect the frequency of words in the text, and these vectors can later be used in various machine learning algorithms. BoW starts with vocabulary generation after the tokenisation steps in order to identify a unique set of words within all documents. As a next step, BoW counts the words for each document, creating a vector containing the frequency of occurrence of the words in that document.

The unigram-bigram method is also a fundamental approach to the numerical representation and analysis of texts. The method breaks text into smaller units called n-grams, which are often used in various text analysis tasks such as language modelling, text classification, and information retrieval. A unigram is the simplest form of n-grams, where the value of "n" is 1, that is, it treats each word (or tokenized unit) separately. Using the simple unigram method, the words of the text are identified as separate tokens, and their occurrences are recorded. The resulting model does not consider the relationships or order between words, it only focuses on the presence and frequency of individual words. This simple approach works well for tasks where the occurrence of individual words is important, such as text classification or similarity measurement between documents. As a next step, the bigram method focuses on combinations of two words in the text. This approach takes into account the order and juxtaposition of words, which can help reveal relationships and connections between words. Increasing the context length for the n-grams (n) allows for more accurate results to be yielded, with more processing time.

The TF-IDF (Term Frequency-Inverse Document Frequency) method is a method that measures the importance of words used in text analysis in documents. This method is especially useful if we want to find out not only how many times a word occurs in a document (as in BoW), but also how important or characteristic the given word is for the document as a whole. Advantages of TF-IDF include filtering – as common words that occur in all documents (e.g., "a", "that") are given a low weight, while rarer but relevant words are given a higher weight, so they are more relevant to the model. Simplicity and efficiency are also an advantage as TF-IDF is easy to calculate and interpret, making it ideal for a variety of text analysis tasks. Limitations of TF-IDF include the ignorance of order as it does not consider the order of words in text, so you may lose context or the relationship

between words. In general TF-IDF is a simple but effective method of numerical representation of text, which helps to highlight words relevant to documents.

4.3. TOPIC MODELLING TECHNIQUES

Latent Dirichlet Allocation (LDA) is a popular topic modelling algorithm widely used in text analytics to identify hidden topics in text documents. LDA can be used to discover topics in large text corpora and automatically categorize documents according to dominant topics. One of the biggest advantages of LDA is its ability to highlight the underlying structures from texts that are difficult to identify manually. LDA is basically a probabilistic generative model. Its assumption is that every document is made up of several hidden topics, and each topic is characterized by certain words. LDA manages three important elements. Documents model each document as a mix of one or more topics. Each document may contain different proportions of these topics. Topics or themes are specific to a particular collection of words. Each word in the document comes from a topic. The assigned a probability of the word is then based on how typical the given word is for a given topic. Advantages of LDA include the automatic nature as it can discover hidden structures in documents without having to work with predefined categories. It is also flexible and so can be well applied to text corpora of different sizes and texts in different languages. Disadvantages include the sparsity problem as LDA can run into problems with rare words or rare topics, as it is difficult to build accurate models with such data. LDA is also sensitive to its parameters such as the number of topics to focus on and therefore applying it always takes some iterations to find the best match for the actual use case.

As an alternative approach to LDA we have used Non-negative Matrix Factorization (NMF) which is also a frequently used method. NMF is a linear algebra technique that approximates a document-term matrix as a product of two lower-rank matrices, one representing documents and topics, and the other representing topics and words. It decomposes the document-term matrix into non-negative factors. Unlike LDA, NMF is purely based on numerical optimization and does not involve probability distributions. One advantage of NMF on top of LDA that it tends to produce fewer overlapping words in topics, which may enable easier interpretation of the topics. This also can come as a disadvantage compared to LDA as in certain corpora there may be topics with slightly different themes/meanings but with more overlapping top words.

Further to the above mentioned traditional techniques, we have also used BERTopic as a modern topic modelling technique that uses deep learning-based language models such as BERT (Bidirectional Encoder Representations from Transformers) to identify topics inherent in texts. BERTopic is able to create deeper contextual representations of words. With BERTopic, the input texts are transformed into a low-dimensional vector space that represents the text content of the documents using BERT or similar models. These vectors carry the meaning and structure of the texts. Based on low-dimensional representations, BERTopic uses the HDBSCAN algorithm to identify topics. The algorithm groups them based on similarities and distances between documents, thus identifying the most important topics. BERTopic is able to generate keywords for each topic from the completed clusters, which characterize the given topic. BERTopic takes into account the contextual meaning of words, enabling the identification of deeper and more accurate topics. This is especially important for texts where the meaning of words can vary depending on the context. It automatically generates topics based on input data, so users do not need to manually define topics, which saves considerable time and resources.

As an addition, various generic and specific Large Language Models (GPTs in particular) can be used for categorising documents into topics. Given the wide variety of GPTs available online and offline, on-premise, in the cloud, with and without commercial implications, with many parameters and settings for many potential use cases, there are various considerations and options using GPTs in determining topics in corpora. The number of options available goes further beyond the scope of this paper. In our analysis we have used some distributions and versions of GPT Large Language Models from the GPT-4, Claude3 and Llama3 model families.

4.4. SENTIMENT ANALYSIS TECHNIQUES

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a popular and powerful sentiment analysis tool specifically designed to determine the emotional content of social media texts. VADER is based on a lexical approach and works with a predefined dictionary that contains words and phrases with emotions. Each word has a valence that indicates the positive or negative direction of the emotion expressed by the word. VADER analyses the words in the input text using its dictionary. By summing up the valence of each word, VADER calculates the overall emotional value of the text. The result of the analysis can be positive, negative or neutral, and it also provides an emotional score expressed on a numerical scale, which indicates the strength of the emotional content of the text. In terms of its advantages, VADER can process large amounts of text data quickly and efficiently. Adaptability is also an advantage as VADER performs well in different linguistic and cultural contexts and can handle the short and informal texts that are often typical in social media. VADER relies on a dictionary-based approach, so it is unable to identify more subtle, subjective or contextual aspects of emotion,

such as sarcasm or slang. VADER is primarily optimized for English texts, so its effectiveness may decrease when used in other languages.

As mentioned above, BERTopic includes a sentiment/valence analysis feature so as part of our analysis we have used that one as well.

We have also used TextBlob functionality to identify and analyse the sentiment polarity of the tweets in our analysis. TextBlob is an easy-to-use Python library that provides simple tools for basic natural language processing tasks, including sentiment analysis. TextBlob's sentiment analysis component uses a lexicon-based approach that assigns a predefined sentiment value to each word or phrase. This enables a quick and efficient analysis of the positive, negative or neutral emotional content of texts. Although limited in terms of contextual understanding and linguistic versatility, it is still a useful tool for quick and basic sentiment analysis in various application areas such as market research, social media analysis and political discourse mapping.

5. DATA

For the analysis, we have ended up using two main datasets. For the years 2020-2022, we have used a dehydrated Twitter dataset, containing message IDs, available online via Zenodo. Using the relevant X.com API, one can get the textual content back for analysis. The dataset had to be reduced as it contained some retweets and duplicates, also some of the tweets were not available any more based on their IDs according to the X.com API responses.

For the year 2024, we have used the X.com with an equivalent set of query keywords to search for tweets as in the 2020-2022 data. The set of query keywords ended up in the following list: #vaccine, #vaccines, #COVID, #COVID19, #COVID-19, #covidvaccine and #vaccination. We aimed to get a similar sized dataset for the two time periods to provide a balanced analysis.

Both datasets went through the same cleaning steps, including the removal of retweets, duplicate tweets and detecting the language of the tweets, all in order to make sure we are focusing on the English language posts.

After the removal of unwanted tweets, the 2020-2022 dataset contained 2711 unique tweets, whereas the 2024 dataset contained 2735 unique tweets. Other studies from the past include much more tweets, however the change in Twitter/X.com's general strategy a couple of years ago, this rich source of information got limited considerably with the rate and other limitations of X.com's APIs.

Further pre-processing steps were done to produce sanitised versions of the tweets' text. These steps included some of the following considerations:

- Stop words (English language) have been removed from the tweet texts
- Special characters, emojis and emoticons have also been removed from the corpus. Although emojis and emoticons may have a specific importance in determining sentiment and valence of the tweets, and in some cases, they can be used to identify factors like sarcasm, the volume of emojis make it more appropriate to remove them from our corpus prepared for topic modelling. The importance of emojis and the level of bias on sentiment of tweets may be analysed in some future work
- Lowercase transformation or the texts
- Stemming
- Lemmatisation
- Hyperlinks could have been considered to be removed, however the preliminary analysis and domain level knowledge dictated that the special characters used in the links are removed, however the domain of links is kept (in one version) of the text. This is because many tweets are sharing content via hyperlinks, and the frequent domains where links point to may have a specific meaning. As we will see in the topic modelling, this can prove useful in understanding and profiling the content.
- Removal of numbers. In some other corpora digits may have a more specific meaning, but in our case, we have decided to remove them from the texts.

6. ANALYSIS AND RESULTS

6.1. TOPIC MODELLING RESULTS

Once having our datasets cleaned, we started on the steps of natural language processing and topic modelling. Various approaches have been applied. Starting with different feature extraction techniques (bag of words, unigram-bigram, TF-IDF and transformers-based in BERT), slightly different document-term matrices or input representations have been produced. These inputs have then been used in various different topic modelling

techniques with different settings to produce the topics. In all cases, the two input datasets of the two time periods have been analysed separately.

First, we have created topics of the tweets using LDA. Different numbers have been used as the number of topics across the two datasets, 10 and 20 topics proving to provide the best interpretable and comparable results. Comparing the results with different number of topics by judging and profiling the text topics was the approach we have taken, accompanied by comparing coherence scores on the topics.

On the 2020-22 dataset, LDA has resulted in various topics that can be interpreted to an extent. Below is the word cloud of the top four topics.

Figure 1: Word cloud of the top 4 topics using LDA on 2020-2022 data

In these word clouds on the 2020-2022 data LDA topics the words related to term vaccine itself are the most prevalent, but the other words can be used to see the difference in the themes discussed in the tweets in each specific topic. Together with details information on the topics and their representative tweets, the topics can be understood better. The top-left topic on the word cloud is mostly about people's knowledge and feelings about the vaccines and that they want to know what is happening, and that the government needs to do better to inform them. The top-right topic is about herd immunity, and risk of vaccination, tweets in this topic seem to be sceptic about vaccines and emphasizing other modes of being immune against COVID-19, like herd immunity. The bottom-left topic is mostly about side effects and getting vaccinated on clinics and potential scepticism around this. The bottom-right topic on the word cloud is about whether the vaccines impact fertility and whether they are safe. Analysing the further LDA topics can reveal other topics, some in general showing more positive opinions on vaccines, whereas most topics are about some concern or types of concerns about the vaccines.

As an alternative to LDA, NMF topics have also been produced and although some of them seem more interpretable, the majority of the NMF topics in our case seems to be even less interpretable than the LDA generated topics.

As a next step, we have created topics using BERTopic on the 2020-2022 dataset.

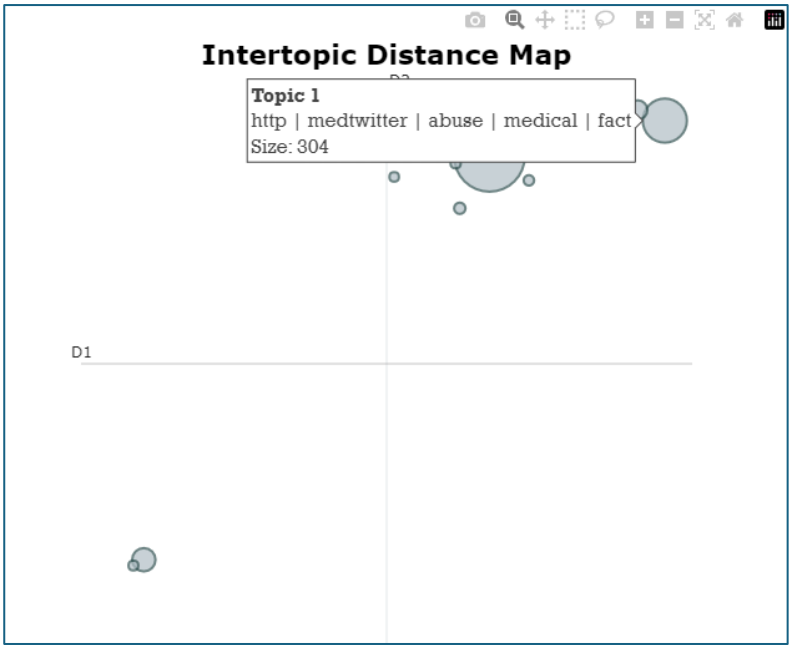

Figure 2: Topic distance map using BERTopic on 2020-2022 data

Figure 3: Word cloud of the top 4 topics using BERTopic on 2020-2022 data

In general, the topics generated by BERTopic in our case are easier to interpret and provide more sensible results when profiling their representative tweets.

Figure 2 above shows the intertopic distance map depicting the topics in the space of the 2 most important LDA dimensions.

Figure 3 above shows the word clouds generated for the top four BERTopic topics on the 2020-2022 dataset.

Topic 1 on Figure 2 (top-right on the word cloud in Figure 3) is a particular topic where the users post a link and/or make a reference to a specific group of users by the hashtag #Medtwitter. This group is an open source medical community sharing medical discoveries and findings. In the tweets in our case, they have posted several studies pro-vaccine, but tweets in the same topic have often reacted in a sceptic or negative way (such as mentioning abuse of information) on Medtwitter's posts, this way showing another form of vaccine hesitancy. The bottom-left topic on the word cloud is that is a similar to one the LDA topics mentioned: containing information about herd immunity and therefore showing vaccine scepticism in that form. Tweets in this topic are not necessarily all against vaccines but at least willing to discuss alternative methods of getting protected, such as with herd immunity.

On the 2024 dataset, LDA has resulted in various topics that can be interpreted, in this case slightly better than the 2020-22 LDA topics. Below is the word cloud of the top four topics.

Figure 4: Word cloud of the top 4 topics using LDA on 2024 data

The top-left topic on the word cloud is mostly about beliefs/misbeliefs in the area of vaccines killing people and some entities in the world specifically performing genocide with vaccines, specifically mRNA vaccines, or relation to perceived diseases such as turbocancer. The top-right topic is about misbeliefs/conspiracy theories around the Marburg syndrome being prevalent in African countries such as Rwanda, potentially being related to vaccines. The bottom-left topic is about the potential use of alternative medicine on COVID-129 and other diseases, these alternative medicine include Ivermectin, HCQS (Hydroxychloroquine), Fenbendazole and others. Tweets in this topic often have links where the users advertise opportunities to buy such alternative medicines, including home shipping. The bottom-right LDA topic on 2024 data is about certain figures or companies in the public and politics, and their related themes such as MAGA. Most tweets in this topic seem to be very anti-vaccine.

For the 2024 data, BERTopic topics have also been generated and they seem to be easy to interpret in our case.

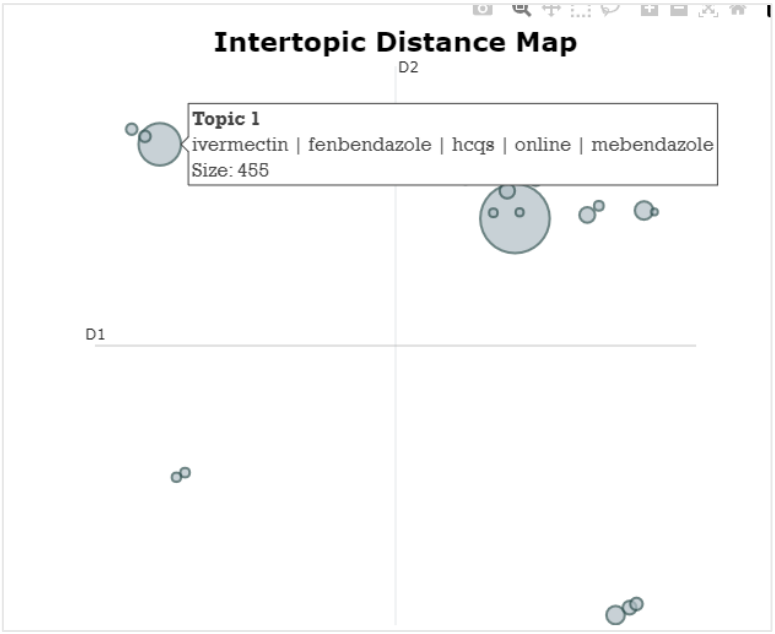

Figure 5: Topic distance map using BERTopic on 2024 data

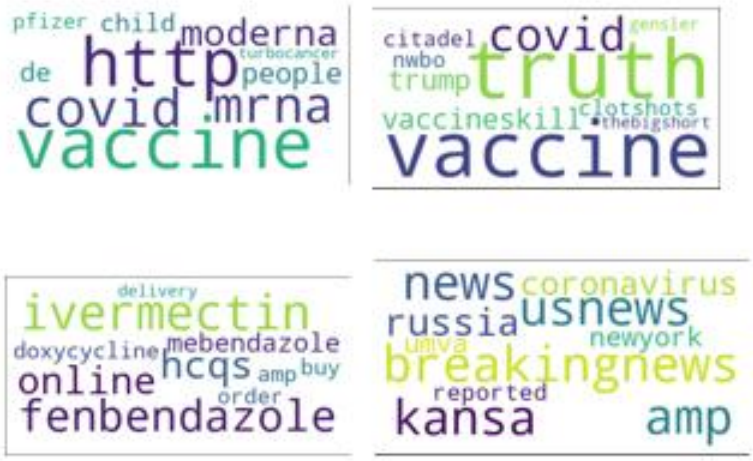

Figure 6: Word cloud of the top 4 topics using BERTopic on 2024 data

Figure 5 above shows the intertopic distance map depicting the topics in the space of the 2 most important LDA dimensions for the 2024 dataset.

Figure 6 above shows the word clouds generated for the top four BERTopic topics on the 2024 dataset.

Topic 1 on Figure 5 (bottom-left on the word cloud in Figure 6) is very similar to the one generated by LDA, where people share their views and commercial options around using alternative medicine such as Ivermectin to cure and prevent COVID-19 and other diseases. The top-left topic seems to be about the different main vaccines, let it be brands (Pfizer, Moderna) or just generally mRNA – sharing views around their risk or potential of causing side illnesses such as turbocancer. The top-right topic in the work cloud seems to be talking political actors and companies, mostly in relation with other beliefs/misbeliefs. Lastly, the bottom-right topic seems to be in relation with users posting various links from news sites or using hashtags around news or breaking news, showing another form of influence of news and politics on people's opinions about vaccines.

6.2. SENTIMENT ANALYSIS RESULTS

We have used three different techniques to categorise/flag sentiment in the tweets.

TextBlob has been used to create a general polarity and subjectivity score for each tweet.

VADER has been used to create a word-valence based sentiment categorisation for each tweet. For each tweet the word valences are being simply averaged going through the words in the tweet. This then results in one score each for the negativity, neutrality and positivity of the tweet, with an altogether compound score. At this stage, this yields a compound score for each of the tweets.

What we then went for is to create a word cloud coloured by the average VADER compound score of the tweets the word is used in the corpus.

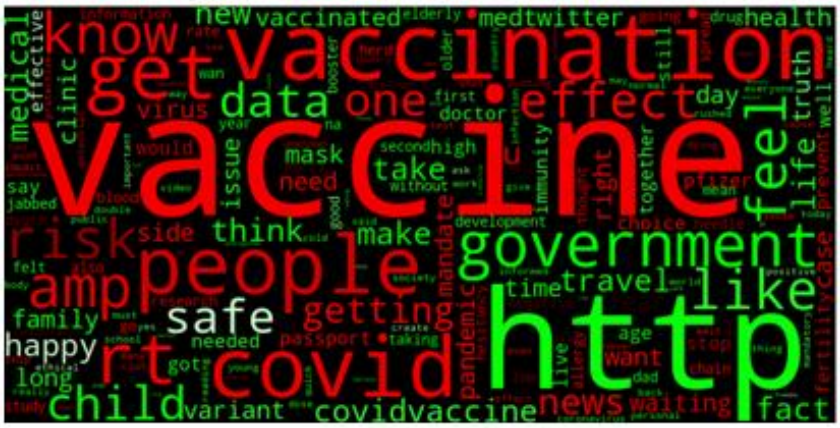

Figure 7: Word cloud by general tweet sentiment on the 2020-2022 data

This word cloud is show on Figure 7 above. The interpretation of this is that "what is average sentiment of the tweets that used this word", or, if you like, "what is the general sentiment of the users who used this word" (which is different to the individual valence of the word itself in VADER's dictionary. This way we can profile the general sentiment of the context each frequent word is being used in our corpus. Based on this, in 2020-2022, the terms vaccine and vaccination were used in negative terms, but other terms such as government, medical or truth were rather used in a more positive average way.

Whereas for 2024, which is shown on the word cloud below on Figure 8, colours have slightly changed. For example, the term http, which signs tweets where links were used is mostly positive in 2020-22 but more negative in 2024. Similarly, the word truth is used in an average more negative way than in 2020-22. The general colour scheme seems to have also changed a bit in the direction of the negatives. Also worth noting the smaller prevalence of the neutral (coloured closer to white) terms or words.

The above analysis seems to be showing public opinions around and related to vaccines on Twitter getting more negative and also more polarised between 2020-22 and 2024.

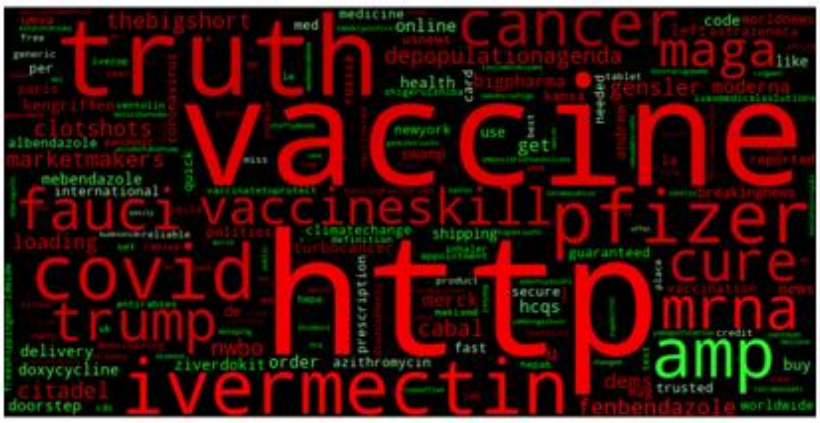

Figure 8: Word cloud by general tweet sentiment on the 2024 data

The same shift towards polarisation and negativity is detectable when visualising the general average sentiment of the tweets based on BERT. Although both shifted, 2024 data seems to be more negative compared to 2020-22 (Figures 9 and 10 below).

Figure 9: BERT-based average sentiment of tweets in the 2020-2022 dataset

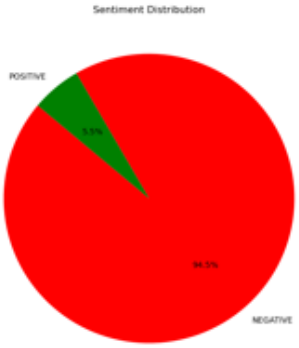

Figure 10: BERT-based average sentiment of tweets in the 2024 dataset

After analysing the sentiment of the tweets in various ways, we have used various publicly available GPTs to detect and label the stance of the tweets against the vaccines, in general to get the GPT respond with whether a certain tweet is pro- or against vaccines. Valuable results have only been yielded by using some task specific LLMs from Huggingface and also via using some of the GPT-4 model family from OpenAI. Even to create meaningful results, this latter analysis needs to remain future work but will show a general stance distribution of tweets in the two corpuses pro or against vaccines.

7. CONCLUSIONS AND FUTURE WORK

In general, topic modelling shows interesting results. Among the different topic modelling techniques, BERTopic proves to be more useful than LDA or NMF approaches as being more interpretable. This is likely due to the nature of these techniques and shows the value of the transformer-based approaches like BERT, however LDA remains a popular technique in uncovering underlying themes and topics in text corpuses.

Various themes and beliefs can be uncovered analysing tweets. Themes include information sharing about vaccines, general hesitancy, misbeliefs and conspiracy theories. Misbeliefs can be analysed using the shown techniques and this can provide valuable insight on ways misbeliefs spread across society. Approaches identifying influencers via network analysis techniques can be useful in uncovering potential ways of fighting misbeliefs and conspiracy theories. Comparing 2020-2022 to 2024, conspiration theories seem to get more frequent and polarisation of sentiments are higher.

It is useful to understand the general sentiment of tweets. To do this, various alternatives are available, some that just considers the words but some also that assumes more of the semantics of the texts, These techniques can be used in conjunction with each other to yield comprehensive results. For sentiment analysis, various custom classifiers and task-specific LLMs are available via Huggingface and other ways. Analysis can be augmented with these, depending on use cases.

Although sentiments of texts or tweets are useful to analyse, they by nature do not show whether the general stance against a phenomenon or an entity is of what direction. Stance analysis, however, can help in this. As an

example, detailed stance analysis can uncover more on pro- or anti-vaccine opinions in texts in general or specifically in Social Media. General LLMs/GPTs can help in this.

As potential future work following this study, we are considering some of the following:

- Stance analysis using LLMs/GPTs and other approaches
- Focus on non-English languages using appropriate tools
- Use human annotated datasets to train and evaluate pro/anti vaccine classifiers
- Combine human annotation with LLM (GPT) annotation on pro/anti vaccine classifiers
- Apply certain feature extraction techniques to uncover more on the nature of public opinion
- Research more into the likely increase in polarisation of opinions on vaccines

REFERENCES

Resnik, P. , Armstrong, W., Claudino, L., Nguyen, T., Nguyen, V., Boyd-Graber, J.: Beyond LDA: Exploring Supervised Topic Modeling for Depression-Related Language in Twitter. In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 2015

Mu, Y., Jin, M., Grimshaw, C., Scarton, C., Bontcheva, K., Song, X.: VaxxHesitancy: A Dataset for Studying Hesitancy towards COVID-19 Vaccination on Twitter. Department of Computer Science, The University of Sheffield, 2023

Mandot, S., Dridi, A., Wanyoni, Z., Vakaj, E.: Leverage Social Media Analysis and NLP to Study the Effectiveness of Antidepressants. PHUSE EU Connect 2023, 2023

Dodson, K.; Mason, J.; and Smith, R. 2021. Covid-19 vaccine misinformation and narratives surrounding Black communities in Social Media. First Draft, 2021

Gisondi, M. A., Barber, R., Faust, J. S., Raja, A., Strehlow, M. C., Westafer, L. M., Gottlieb, M.: A deadly infodemic: social media and the power of COVID-19 misinformation. Journal of Medical Internet Research, 2022

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author name: Tamas Bosznay

Company: Katalyze Data

Address: The Old School Hall OX28 6ZJ Witney Oxfordshire United Kingdom

Email: tamas.bosznay@katalyzedata.com

Website: www.katalyzedata.com