Clustering-driven Sentiment analysis for COVID-19 vaccination in Tunisia

Imen Hamed, Wala Rebhi, Narjes Bellamine Ben Saoud

Proceedings of the 38th Pacific Asia Conference on
Language, Information and Computation (PACLIC 38)

Nathaniel Oco, Shirley N. Dita, Ariane Macalinga Borlongan, Jong-Bok Kim (eds.)

2024

Clustering-Driven Sentiment Analysis for COVID-19 Vaccination in Tunisia

Imen Hamed and **Wala Rebhi** and **Narjes Bellamine Ben Saoud**
RIADI laboratory
National school of computer science
University of Manouba, Manouba, Tunisia

Abstract

The rapid development of vaccines for the infectious coronavirus disease-19 (COVID-19) has been a crucial solution to combat the global impact of the virus. As a result, understanding the sentiments responses of individuals towards vaccination has become a significant issue since it could provide valuable insights into the public sentiment landscape and help inform targeted strategies for addressing concerns, increasing vaccine acceptance, and tailoring communication efforts. However, while sentiment analysis has matured for widely spoken languages like English, addressing dialects, such as the Tunisian dialect, remains a challenging task. In this context, this paper aims to propose a clustering-based approach for analyzing sentiments related to the COVID-19 vaccine using social media data, specifically focusing on Tunisian Facebook users. This approach combines the k-means clustering algorithm with the Naive Bayes classification model in order to classify Tunisians' opinions towards vaccination. Compared with the pre-trained Arabert model, the proposed approach gives better results proving its effectiveness for Tunisian opinions classification.

1 Introduction

The COVID-19 pandemic has radically affected the overall wellness and health of the entire world (Catapang and Cleofas, 2022). Indeed, it has changed our lives, not only in the health care area, but also in many aspects of human life such as education, transportation, politics, supply chain, etc. (Pham et al., 2020). As of March 11, 2020, there were 118,326 confirmed cases and 4,292 deaths, according to the World Health Organization who declared the COVID-19 as a pandemic on the same day (Ge et al., 2020).
Furthermore, the Covid-19 has been considered much more dangerous and easily spread than other Coronavirus families because it has become highly efficient in human-to-human transmissions (Pham et al., 2020). In Tunisia, for example, there have been 1,01 Million confirmed cases and 27922 deaths since March 2020 until March 2022 according to the Tunisian public health ministry.

Although COVID-19 preventive behaviors such as mask wearing and social distancing have been shown to be effective in curbing the spread of the virus, long-term control of the COVID-19 pandemic will hinge on the development and uptake of a preventive vaccine (Chou and Budenz, 2020). Therefore, the development of vaccines against COVID-19 made rapid progress in the last three years and to date, different vaccines showed good efficacy against COVID-19 (Bendau et al., 2021).

In this context, understanding the sentiments or emotional responses of individuals towards vaccination has become a crucial and relevant issue. This enables the identification of specific sentiments prevalent within different communities, such as vaccine hesitancy, vaccine confidence, or concerns about vaccine safety and efficacy. Analyzing these sentiments can provide valuable insights to address the unique concerns and needs of each community, ultimately promoting informed decision-making and increasing vaccine acceptance and uptake.

People nowadays rely mainly on social media to express their feelings, thoughts and opinions on different kind of events. Facebook, twitter and Instagram are considered as preferred platforms with millions of users. Thanks to the rapid information dissemination, social media platforms become the first source of information for many individuals. Therefore, collecting data and analyzing it may provide insights about different viewpoints about Covid-19 vaccine. There are different ways to analyze social media content, sentiment analysis is prominent among them. Two main methodologies can be used to perform sentiment analysis: knowledge-based systems based upon linguis-

tic rules and statistical machine learning models (Samaras et al., 2023). Since Twitter is considered very active platform, many researchers employ the huge number of tweets produced on a daily basis in different languages such as: Portuguese (Garcia and Berton, 2021), Spanish (Turón et al., 2023), Greece (Samaras et al., 2023) and especially English to analyze sentiments.

Despite the presence of many efforts on analyzing social media platforms with different languages, there is still lack of works on Arabic dialects mainly Tunisian dialect. Moreover, to the best of our knowledge, there are no studies on Tunisian sentiment analysis regarding COVID-19 vaccination. Thus, in this paper we focus on Facebook comments written in Tunisian dialect to analyze and extract meaningful insights by proposing a hybrid clustering-based approach for textual sentiment detection.

Therefore, our main contributions are:
-Proposing a new hybrid approach for Tunisian sentiment analysis by combining unsupervised clustering with supervised classifier model.
-Analyzing and classifying Tunisian comments to get meaningful insights about Tunisians opinions towards the Covid-19 vaccine.

The remainder of this paper is as follows: Section 1 is dedicated to present related work about sentiment analysis and covid-19 vaccination. We detail the proposed approach in Section 2. Section 3 is devoted to showcase the experimental results. We evaluate the proposed approach and discuss the retrieved results in Section 4. Finally, we conclude the paper and present future research directions.

2 Related work: Sentiment Analysis and Covid-19 Vaccination

Many recent studies have investigated peoples opinions regarding the COVID-19 vaccine (Antoun et al., 2020). Indeed, studying peoples perceptions on social media to understand their sentiment presents a powerful medium for researchers to identify the causes of vaccine hesitancy and therefore develop appropriate public health messages and interventions (Alamoodi et al., 2021).

For example, the authors in (Hussain et al., 2020) develop and apply an artificial intelligence (AI)-based approach to analyze social-media public sentiment in the United Kingdom (UK) and the United States (US) towards COVID-19 vaccinations, to better understand public attitude and identify topics of concern.

Likewise, in (Kwok et al., 2021) the authors use machine learning methods to extract topics and sentiments relating to COVID-19 vaccination on Twitter. To do this, they collected 31,100 English tweets containing COVID-19 vaccinerelated keywords between January and October 2020 from Australian Twitter users. Specifically, they analyzed tweets by visualizing high-frequency word clouds and correlations between word tokens. They built a latent Dirichlet allocation (LDA) topic model to identify commonly discussed topics in a large sample of tweets. They also performed sentiment analysis to understand the overall sentiments and emotions related to COVID-19 vaccination in Australia (Kwok et al., 2021).

The authors in (Turón et al., 2023) propose to analyze Spanish tweets through the combination of sentiment analysis techniques mainly lexicons and multivariate statistical methods to track the evolution of social mood. They recognized different emotions during the four phases of the vaccination process and show the interconnections and clustering of the community of tweeters around interest groups. As for authors in (Garcia and Berton, 2021), they focus on Brazil and USA since these countries had a large number of COVID cases. They explore English and Portuguese tweets to detect dominant sentiments related to Covid-19 discussions. They mainly find out fear is the most dominant feeling. They also recognize ten different topics in the conversations. They mainly rely on existing classifiers, they combined recent embedded models to extract features. Another study (Lin et al., 2023) uses multilingual tweets to distinguish different opinions about COVID 19 vaccines. They compared machine learning models such as Random Forest (RF) and Support Vector Machine (SVM) with different deep learning models to find up that deep learning methods outperform machine learning ones in tweets classification.

Now addressing the Tunisian Covid-19 scenario, we may find very few works on the sentiment analysis during the pandemic. For example, the work in (Shahriar et al., 2022) performs sentiment analysis on Tunisian comments to analyze public perceptions on the Covid-19 pandemic. The problem was considered as text classification issues: multi-class classification for sentiment analysis (optimist, pessimist or neutral) and binary

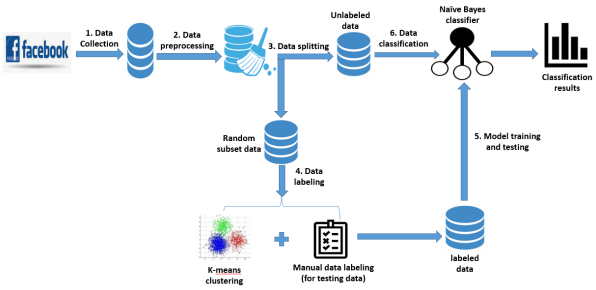

Figure 1: The clustering-driven sentiment analysis approach.

classification for sarcasm detection. Then, the authors compared machine learning models and deep learning ones on the studied data set. They find up that deep learning models outperform the machine learning models. As for authors in (Mekki et al., 2022), they refer to deep learning model Bi-LSTM to analyze sentiments of Tunisians during the pandemic. So, they introduce a deep Bi-LSTM network to improve the sentiment analysis task. The proposed model outperforms machine learning models and standard deep learning models.

Despite the existence of many studies addressing sentiment analysis regarding COVID-19 vaccination, it is noteworthy that these studies are mainly related to texts written in English. Only few studies have addressed sentiment analysis for Tunisian dialect. Moreover, even this limited number of studies have approached the Covid-19 pandemic in general rather than vaccination specifically. Furthermore, for the Tunisian dialect, another issue concerns the lack of data for training classification models. Most works resort to manual data annotation, which is relevant but resource-intensive in terms of time. This is why, in this work, we propose a hybrid approach for analyzing Tunisian Covid-19 vaccination opinions, which presents a solution for manual data annotation.

This approach will be detailed in the next section.

3 Proposed approach: a clustering-driven sentiment analysis approach for Tunisian COVID-19 Vaccination

In order to analyze sentiment responses of Tunisian individuals towards Covid-19 vaccination, we propose a hybrid approach, as illustrated by Fig. 1. This approach, by combining unsupervised and supervised methods, contains six

Before cleaning: masa7eche eli mla9a7 ma ya3diche 😷😔

After cleaning: ['masaheche', 'eli', 'mlakah', 'yaadiche', 'yebki', 'mghachech']

Figure 2: Example of data before and after cleaning.

phases:

1. Data Collection: The first step undertaken involved data collection. We embarked on gathering comments from Tunisians posts on Facebook regarding the COVID-19 vaccination campaign from August 2021 until June 2022.

For this purpose, we opt to use the "ExportComments" platform[1], which allowed us to extract relevant comments from the URL of each post and save them in Excel format.

The process took place in two steps: first, we identified the links to the posts originating from the Tunisian Ministry of Health on Facebook; then, we provided these links successively to the "ExportComments" platform, which merged the resulting Excel files into a single dataset file.

Thus, we obtained 50k comments written in Tunisian dialect in Arabic and Latin letters from posts related to Covid-19 vaccination.

2. Data pre-processing: After data extraction, the next step is data preprocessing which relies mainly on removing noise and irrelevant information so that the effectiveness of subsequent model learning could be optimized. This involves cleaning the data by removing duplicate comments and unnecessary symbols such as stop words, punctuation marks and any URLs or mentions of other Facebook users. Likewise, emojis were converted into words. Finally, the data has been tokenized into words and normalized by transforming number into letters as Tunisian dialect uses a lot of numbers instead of letters. For example: 2 -> "a"; 3 -> "a"; 4 -> "gh"; 5 ->"kh"; 7 -> "h"; 8 -> "ch"; 9 -> "k"; etc.

Fig. 2 shows before and after cleaning a sentence from the collected data. The cleaning process consists in removing punctuation, emojis and stop words. Stop words in Arabic are not very informative so we processed to remove it. Another very important step involves vectorization, which refers to the process of converting textual data into numerical representations. In this setting, there are several vectorization techniques such as Continuous Bag of Words (CBOW), Skip-gram bag of words, Term FrequencyInverse Document Frequency (TF-IDF), and Distributed Memory of

[1]https://exportcomments.com/

Paragraph Vector (DM-PV) (Dey et al., 2016). In this work, we have chosen CBOW (Continuous Bag-of-Words) model as part of the Word2Vec method, which learns the embedding by predicting the current word based on its context (Shahriar et al., 2022). This technique has proven its effectiveness compared to other techniques (Dey et al., 2016), particularly for the Tunisian dialect (Shahriar et al., 2022).

3. Data splitting: After cleaning our data, we split it into two sets: unlabeled and labeled data. Indeed, a random subset of 5K comments (10% of the total comments) was extracted to undergo annotation for training and testing the model.

4. Data labeling: K-means clustering + manual labeling: This step concerns only the random subset of 5K comments and it aims to annotate data for our model training and testing. As the annotation process is crucial and time-consuming manually, we resorted to an unsupervised method which is the k-means clustering. This choice was motivated by the fact that this method has been used in several studies and has yielded significant results for English text analyzing (Lin et al., 2023) and (Chen et al., 2022). Furthermore, it has been listed among the top 10 clustering algorithms for data analysis (Shahriar et al., 2022). Indeed, K-Means clustering is a popular clustering algorithm based on the partition of data. Data that have the same characteristics are grouped into one cluster, whereas data that have different characteristics are grouped into other clusters (Chen et al., 2022). Steps for K-Means clustering are as follows (Chen et al., 2022) and (Mekki et al., 2022):

1. Decide the number of cluster K
2. Initialization of the cluster center (centroid). It can be conducted by using various ways. However, the most frequent way is by using random way. Clusters centers are assigned by random numbers.
3. Allocate all data/objects to the closest cluster. Determination of closeness of two objects is determined based on the distance of two objects. For calculating the distance of all data to each centroid point, Euclidean Distance theory is used, which is formulated as given in Equation (1):

$$D(i,j)=\sqrt{(X_{i1}-X_{j1})^2+(X_{i2}-X_{j2})^2+\cdots+(X_{ip}-X_{jp})^2} \quad (1)$$

Where:
$D(i,j)$ = distance of i^{th} data to cluster center j
$X_{ik} = i^{\text{th}}$ on the k^{th} data attribute
X_{ij} = j^{th} center point on the k^{th} data attribute

4. Recalculate centroid with current cluster membership. Centroid is an average (mean) of all data/objects within particular cluster. If desired, the median of this cluster can also be used.
5. Reassign each object by using new cluster center, if the cluster doesnt change, then clustering process finished otherwise repeat step 3 until there is no change for each cluster.

In K-Means, the number of clusters is determined. For this work, the number of clusters is fixed at **three (3)** since we noticed the presence of three opinions in favor, against, and neutral.

Thus, as a result, we obtained three clusters: Cluster 1 with 1966 comments, Cluster 2 with 451 comments and Cluster 3 with 2583 comments. Fig. 3 shows an example of comment from each cluster.

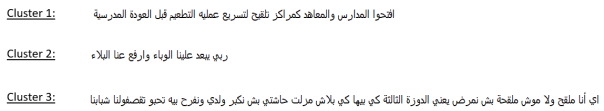

Figure 3: Example of comment from each cluster.

Then, we extracted approximately 20% comments from each cluster which are in total 1000 comments. These comments were manually annotated in order to discover firstly which cluster represents which sentiment and to evaluate the relevance of the K-means clustering as shown in Table 1. Indeed, we calculated the precision obtained for each cluster, which was higher than 90% for the three clusters.
Then, this 1000 comments will be considered as a testing set of data for the classification model later.

Therefore, at this stage, we managed, by applying the k-means clustering, to obtain training and test datasets.

5. Model training and testing: For this step, two models were trained and tested: the Naive Bayes classifier and k-nearest neighbour classifier. Naïve Bayes algorithm is a simple probabilistic

Table 1: K-means clustering Evaluation.

Cluster	Number of comments	Number of correct comments	Precision of each cluster
Cluster 1 (positive comments)	392	364	0.92
Cluster 2 (neutral comments)	92	88	0.95
Cluster 3 (negative comments)	516	467	0.90

classifier that applies the Bayes theorem that calculates a set of probabilities by calculating the frequency and the combination of values of the given data set (Jaballi et al., 2023) and (Chen et al., 2022). The reason behind this choice is that the Naive Bayes is fast and accurate and widely used for classification problems.

As for the k-nearest neighbour model, it is one of the most fundamental and simple classification methods and it is commonly based on the Euclidean distance between a test sample and the specified training samples (Peterson, 2009) and (Cunningham and Delany, 2021).

Using the training dataset, we first trained each model. Then, we evaluated them using the test data and based on the three evaluated metrics: P: Precision, R: recall and A: accuracy. This evaluation, as given in Table 2, shows the effectiveness of Naive Bayes compared to k-nearest neighbour.

Table 2: Comparison of classification models.

Classifier Model	Precision	Recall	Accuracy
Naive Bayes	0.79	0.744	0.8
k-nearest neighbour	0.72	0.69	0.73

This is why, in this work we propose to choose Naive Bayes to classify our datasets.

6. Data classification: The last step is the classification of the unlabelled data which is of 45K comments. Indeed, once trained and tested, we used our model to classify the data. So, we obtain a set of comments annotated with either 1 (positive), 0 (neutral), or -1 (negative).

4 Evaluation and results

In this section, we first evaluate our approach, which consists of combining a clustering method with a classification model in order to determine the sentiment from a comment. Then, we provide an analysis of the results obtained within the context of COVID-19 vaccination in Tunisia.

4.1 Evaluation

In order to evaluate the proposed approach, we propose to apply the pretrained Arabert to the annotated test set data. Arabert is a pretarained model based on the BERT transformer model (Devlin, 2018) for the Arabic language (Antoun et al., 2020). Indeed, our goal is to see the impact of using an unsupervised method for constructing the training data. This explains why we choose to deal with a pretrained model.

Table 3: Proposed approach evaluation.

Model	Precision	Recall	Accuracy
Proposed approach	0.79	0.744	0.8
Arabert	0.757	0.743	0.753

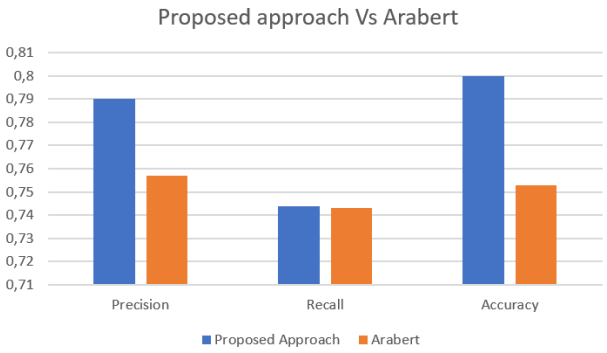

Figure 4: Proposed approach evaluation.

The evaluation results (Precision (P), Recall (R) and Accuracy (A)) detailed in Table 3 and are graphically visualized in Fig. 4, indicate that the proposed approach is more efficient than the pretrained model, particularly in terms of precision and accuracy.. This could be attributed to the fact that Arabert is pretrained for the Arabic language, whereas the Tunisian dialect incorporates more specificities such as the combination of various languages such as French and English. Moreover, the discussed topic, concerning COVID-19 vaccination within the comments, is notably specific, and viewpoints diverge from typical subjects that rely on familiar terminology.

4.2 Results analysis

After evaluating the proposed approach, we propose in this section to use the classification re-

sults to analyze Tunisians' sentiments towards the COVID-19 vaccination.

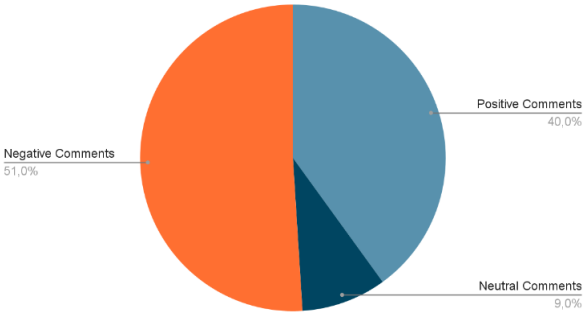

Figure 5: Comments statistics.

To begin with, as shown in Fig. 5, we notice that generally negative comments carry more weight which is 51% than positive and neutral comments. To better understand these results, we suggest tracking the evolution of these comments over time. Thus, considering the timestamps of each comment, we observe from Fig. 6 that the number of negative comments increased while positive comments decreased over time. As for neutral comments, their volume seems to remain relatively stable throughout. This trend logically corresponds to the vaccine's rollout date. Indeed, based on this variation, it is possible to distinguish two phases:

- Phase 1: August 2021 until November 2021: characterized by a very high number of positive comments. This phase corresponds in reality to the launch of the vaccine in Tunisia and the decrease in COVID-19 cases following this vaccination.
- Phase 2: December 2021 until June 2022: characterized by a decrease in positive comments with a remarkable increase in negative comments. This can be explained by the fact that many individuals were still affected by COVID-19 despite vaccination, as well as the appearance of negative effects of the vaccine on several individuals.

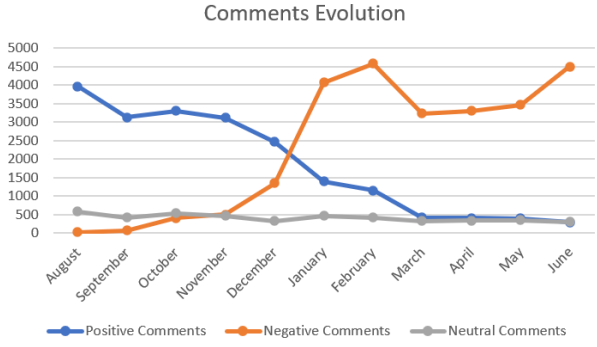

Figure 6: Evolution of Tunisian sentiments towards COVID-19 vaccination.

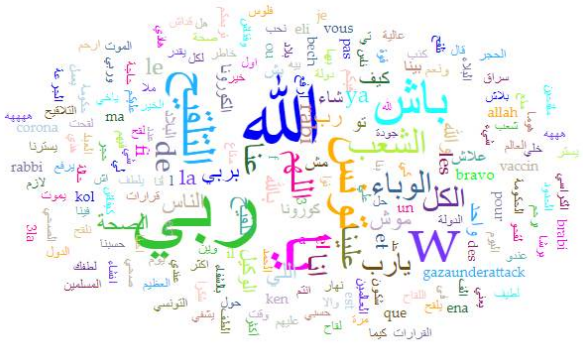

Figure 7: Word Cloud representing the most commonly occurring words in the gathered dataset.

Finally, in order to know which words were most frequently used during the Covid-19 vaccination, we create the word cloud as shwon in Fig. 7. We observed that terms such as the names of God (Rabi, Allahomma, Allah), prayers (Ya rab, Yostorna) and terms related to COVID-19 vaccination (TalKih, Joraa, Sahha, wabaa) exhibited the highest frequency in the word cloud generated from the collected dataset. This could be explained by Tunisian citizens' habit of expressing their various positive, negative, and especially neutral sentiments through prayer to God.

4.3 Discussion

This paper introduces a hybrid approach that integrates both supervised and unsupervised methods to analyze the sentiments of Tunisian people towards COVID-19 vaccination, particularly focusing on their expressions on the Facebook platform. The proposed approach has shown notable effectiveness by achieving high precision and recall rates, and it also excels in terms of speed because it automates the manual annotation process without compromising the quality of the results.

Moreover, the dataset generated from this study can serve as a valuable reference for future sentiment analysis on various other diseases affecting the Tunisian population. This data can provide insights and benchmarks for researchers and policymakers.

Additionally, the insights derived from this study act as a crucial alert for the Ministry of

Health, underscoring the importance of increasing efforts and preparedness when introducing new vaccines. The analysis emphasizes the need for proactive measures to address public concerns and improve vaccination campaigns.

Finally, this research has the potential to be extended further to investigate the negative side effects associated with different vaccines. This is particularly important given the documented issues with certain vaccines that have been approved and used in various countries. Such extensions could provide critical information for improving vaccine safety and public health strategies.

5 Conclusion

This paper introduced a clustering-based approach to examine sentiments surrounding the COVID-19 vaccine via social media posts, focusing particularly on Tunisian Facebook users. The technique combines the k-means clustering algorithm with the Naive Bayes classification model to sort Tunisian perspectives on vaccination.
Compared to the pre-trained Arabert model, the proposed method yields better results and demonstrates its efficacy in classifying Tunisian viewpoints.
Additionally, a detailed analysis of the findings is provided to gain insights into Tunisian attitudes towards COVID-19 vaccination.

As future work, we aim to enrich the obtained datasets with additional comments about other pandemics to establish a baseline for the Tunisian dialect that could be used to support the understanding of pandemics. Moreover, we intend to add another crucial facet of the comments analysis: detecting sarcasm, which can be very informative when studying social behavior and expressed emotions.

References

Abdullah Hussein Alamoodi, BB Zaidan, Maimonah Al-Masawa, Sahar M Taresh, Sarah Noman, Ibraheem YY Ahmaro, Salem Garfan, Juliana Chen, Mohamed Aktham Ahmed, AA Zaidan, et al. 2021. Multi-perspectives systematic review on the applications of sentiment analysis for vaccine hesitancy. *Computers in Biology and Medicine*, 139:104957.

Wissam Antoun, Fady Baly, and Hazem Hajj. 2020. Arabert: Transformer-based model for arabic language understanding. *arXiv preprint arXiv:2003.00104*.

Antonia Bendau, Jens Plag, Moritz Bruno Petzold, and Andreas Ströhle. 2021. Covid-19 vaccine hesitancy and related fears and anxiety. *International immunopharmacology*, 97:107724.

Jasper Kyle Catapang and Jerome V Cleofas. 2022. Topic modeling, clade-assisted sentiment analysis, and vaccine brand reputation analysis of covid-19 vaccine-related facebook comments in the philippines. In *2022 IEEE 16th International Conference on Semantic Computing (ICSC)*, pages 123–130. IEEE.

Ninghan Chen, Xihui Chen, and Jun Pang. 2022. A multilingual dataset of covid-19 vaccination attitudes on twitter. *Data in Brief*, 44:108503.

Wen-Ying Sylvia Chou and Alexandra Budenz. 2020. Considering emotion in covid-19 vaccine communication: addressing vaccine hesitancy and fostering vaccine confidence. *Health communication*, 35(14):1718–1722.

Padraig Cunningham and Sarah Jane Delany. 2021. K-nearest neighbour classifiers-a tutorial. *ACM computing surveys (CSUR)*, 54(6):1–25.

Jacob Devlin. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Lopamudra Dey, Sanjay Chakraborty, Anuraag Biswas, Beepa Bose, and Sweta Tiwari. 2016. Sentiment analysis of review datasets using naive bayes and k-nn classifier. *arXiv preprint arXiv:1610.09982*.

Klaifer Garcia and Lilian Berton. 2021. Topic detection and sentiment analysis in twitter content related to covid-19 from brazil and the usa. *Applied soft computing*, 101:107057.

Yiyue Ge, Tingzhong Tian, Suling Huang, Fangping Wan, Jingxin Li, Shuya Li, Hui Yang, Lixiang Hong, Nian Wu, Enming Yuan, et al. 2020. A data-driven drug repositioning framework discovered a potential therapeutic agent targeting covid-19. *BioRxiv*, pages 2020–03.

Amir Hussain, Ahsen Tahir, Zain Hussain, Zakariya Sheikh, Mandar Gogate, Kia Dashtipour, Azhar Ali, and Aziz Sheikh. 2020. Artificial intelligence-enabled analysis of uk and us public attitudes on facebook and twitter towards covid-19 vaccinations. *medRxiv*, pages 2020–12.

Samawel Jaballi, Manar Joundy Hazar, Salah Zrigui, Henri Nicolas, and Mounir Zrigui. 2023. Deep bidirectional lstm network learning-based sentiment analysis for tunisian dialectical facebook content during the spread of the coronavirus pandemic. In *International Conference on Computational Collective Intelligence*, pages 96–109. Springer.

Stephen Wai Hang Kwok, Sai Kumar Vadde, and Guanjin Wang. 2021. Tweet topics and sentiments relating to covid-19 vaccination among australian twitter

users: machine learning analysis. *Journal of medical Internet research*, 23(5):e26953.

Bor-Shen Lin et al. 2023. Visualizing change and correlation of topics with lda and agglomerative clustering on covid-19 vaccine tweets. *IEEE Access*, 11:51647–51656.

Asma Mekki, Inès Zribi, Mariem Ellouze, and Lamia Hadrich Belguith. 2022. A tunisian benchmark social media data set for covid-19 sentiment analysis and sarcasm detection.

Leif E Peterson. 2009. K-nearest neighbor. *Scholarpedia*, 4(2):1883.

Quoc-Viet Pham, Dinh C Nguyen, Thien Huynh-The, Won-Joo Hwang, and Pubudu N Pathirana. 2020. Artificial intelligence (ai) and big data for coronavirus (covid-19) pandemic: a survey on the state-of-the-arts. *IEEE access*, 8:130820–130839.

Loukas Samaras, Elena García-Barriocanal, and Miguel-Angel Sicilia. 2023. Sentiment analysis of covid-19 cases in greece using twitter data. *Expert Systems with Applications*, 230:120577.

Khandaker Tayef Shahriar, Muhammad Nazrul Islam, Md Musfique Anwar, and Iqbal H Sarker. 2022. Covid-19 analytics: Towards the effect of vaccine brands through analyzing public sentiment of tweets. *Informatics in medicine unlocked*, 31:100969.

A Turón, A Altuzarra, JM Moreno-Jiménez, and J Navarro. 2023. Evolution of social mood in spain throughout the covid-19 vaccination process: a machine learning approach to tweets analysis. *Public health*, 215:83–90.