Book Chapter

An Analysis of French-Language Tweets About COVID-19 Vaccines: Supervised Learning Approach

Romy Sauvayre[1,2]*, Jessica Vernier[1] and Cédric Chauvière[2,3]

[1]Laboratoire de Psychologie Sociale et Cognitive, Université Clermont Auvergne, CNRS, France
[2]Polytech Clermont, Clermont Auvergne INP, France
[3]Laboratoire de Mathématiques Blaise Pascal, Université Clermont Auvergne, CNRS, France

***Corresponding Author:** Romy Sauvayre, Laboratoire de Psychologie Sociale et Cognitive, Université Clermont Auvergne, CNRS, Clermont-Ferrand, France

Published **October 10, 2023**

This Book Chapter is a republication of an article published by Romy Sauvayre, et al. at JMIR Medical Informatics in May 2022. (Sauvayre R, Vernier J, Chauvière C An Analysis of French-Language Tweets About COVID-19 Vaccines: Supervised Learning Approach JMIR Med Inform 2022;10(5): e37831. URL: https://medinform.jmir.org/2022/5/e37831
DOI: 10.2196/37831)

How to cite this book chapter: Romy Sauvayre, Jessica Vernier, Cédric Chauvière. An Analysis of French-Language Tweets About COVID-19 Vaccines: Supervised Learning Approach. In: Nicollas Nunes Rabelo, editor. Prime Archives in Medicine: 5th Edition. Hyderabad, India: Vide Leaf. 2023.

Conflicts of Interest: The authors declare no competing interests.

Abstract

Background: As the pandemic progressed, disinformation, fake news and conspiracy spread through many parts of society. However, the disinformation spreading through social media is, according to the literature, one of the causes of increased COVID-19 vaccine hesitancy. In this context, the analysis of social media is particularly important, but the large amount of data exchanged on social networks requires specific methods. This is why machine learning and natural language processing (NLP) models are increasingly applied to social media data.

Objective: The aim of this study is to examine the capability of the CamemBERT French language model to faithfully predict elaborated categories, with the knowledge that tweets about vaccination are often ambiguous, sarcastic or irrelevant to the studied topic.

Methods: A total of 901,908 unique French tweets related to vaccination published between July 12, 2021, and August 11, 2021, were extracted using the Twitter API v2. Approximately 2,000 randomly selected tweets were labeled with two types of categorization: (1) arguments for ("pros") or against ("cons") vaccination (sanitary measures included) and (2) the type of content of tweets ("scientific", "political", "social", or "vaccination status"). The CamemBERT model was fine-tuned and tested for the classification of French tweets. The model performance was assessed by computing the F1-score, and confusion matrices were obtained.

Results: The accuracy of the applied machine learning reached up to 70.6% for the first classification ("pros" and "cons" tweets) and up to 90.0% for the second classification ("scientific" and "political" tweets). Furthermore, a tweet was 1.86 times more likely to be incorrectly classified by the model if it contained fewer than 170 characters (odds ratio = 1.86; 1.20 < 95% confidence interval < 2.86).

Conclusions: The accuracy is affected by the classification chosen and the topic of the message examined. When the vaccine debate is jostled by contested political decisions, tweet content becomes so heterogeneous that the accuracy of the models drops for less differentiated classes. However, our tests showed that it is possible to improve the accuracy of the model by selecting tweets using a new method based on tweet size.

Keywords

Social Media; Natural Language Processing; Public Health; Vaccine; Machine Learning; CamemBERT Language Model; Method; Epistemology

Abbreviations

Adam: Adaptive Moment Estimation; CCNet: Criss-Cross Network; BERT: Bidirectional Encoder Representations from Transformers; NLP: Natural Language Processing

Introduction

Background

The COVID-19 pandemic has profoundly affected our society and social activity on the planet. A part of this activity is perceptible through messages exchanged on social media, specifically on the vaccination topic. Since the apparition of the MMR (measles, mumps and rubella) vaccine controversy in 1998 [1], vaccine hesitancy has grown on the internet [2,3] and subsequently on social media such as Facebook and Twitter [4,5]. In the same way, as the pandemic progressed, disinformation, fake news and conspiracy spread [6] through many parts of society. However, the disinformation spreading through social media is, according to the literature, “potentially dangerous” [7] and is one of the causes of increased COVID-19 vaccine hesitancy [8,9]. Another cause mentioned focuses on the loss of confidence in science [10].

In this context, social media analysis is particularly important, but the large amount of data exchanged over social networks

requires specific methods. This is why machine learning and natural language processing (NLP) models are increasingly popular for studying social media data. The most used and "most promising method" [11] is sentiment analysis. For example, sentiment analysis was conducted on messages posted on Twitter (tweets) to measure the opinion of Americans regarding vaccines [12] or to evaluate the rate of hate tweets among Arabic people [13]. Additionally, another method, opinion mining, is used and has obtained an equal level of maturity [14]. Both methods attempt to identify and categorize subjective content in text, but it is not an easy task to correctly identify such concepts (opinion, rumor, idea, claim, argument, emotion, sentiment, affect). Psychology and philosophy have extensively studied these concepts but have raised the difficulty of defining their boundaries. This is why stance detection raising has grown to be considered "a subproblem of sentiment analysis" [15]. In addition, according to Visweswaran et al. [16], sentiment analysis of tweets is a challenge because tweets contains short text (280 character limitation), abbreviations and slang terms. However, few studies focus on the difficulties encountered by a neural network according to the chosen categories [17]. The aim of this article is to provide additional methodological reflection.

Objective

The aim of this study is to examine the capability of the CamemBERT model to faithfully predict elaborated categories while considering that tweets about vaccination are often ambiguous, sarcastic or irrelevant to the studied topic. Based on the resulting analysis, this article aims to provide a methodological and epistemological specific reflection about the analysis of French tweets related to vaccination.

A State-of-the-Art French Language Model

The CamemBERT model was launched in 2020 and is considered one of the state-of-the-art French language models [18] (together with its close cousin flauBERT [19]). It makes use of the RoBERTa (robustly optimized BERT pretraining approach) architecture of Liu et al. [20], which is an

improved variant of the famous BERT (bidirectional encoder representations from transformers) architecture of Devlin et al. [21]. The BERT family of models are general multipurpose pre-trained models that may be used for different NLP tasks: classification, question answering, translation, and so on. They rely heavily upon transformers, which have radically changed the performance of NLP tasks since their introduction by Google researchers in 2017 [22]. They have been pre-trained on large corpus ranging from gigabits to terabits of data using considerable computing resources.

Although multilingual models are plentiful, they usually lag behind their monolingual counterparts. This is why, in this study, we chose to employ a monolingual model to classify French tweets. As far as we are concerned, CamemBERT comes in six different flavors, ranging from small models with 110 million parameters trained on 4 GB of text up to middle size models having 335 million parameters and trained on 135 GB of text. After testing them, we found that better results were obtained with the largest size model that was pre-trained on the CCNet corpus.

All these models require fine-tuning on specific data to achieve their full potential. Fine-tuning or transfer learning have been common and successful practices in computer vision for a long time, but it is only in the last three years or so that the same approaches have become effective for solving NLP problems on specific data. This approach can be summarized in the following three steps:

- A model language such as BERT is built in an unsupervised manner using a large database, removing the need to label data;
- A specific head (such as dense neural network layers) is added to the previous model to make it task specific;
- The new model is trained in its entirety with a small learning rate on specific data.

The first step is usually performed by large companies, such as Google or Facebook, or public research centers that make their

model freely available on internet platforms. The second and third steps are a process that is generally referred to as "fine-tuning", and this is what we will do in this study.

Methods

Data Collection

French tweets published between July 12, 2021, and August 11, 2021, were extracted using the Twitter API v2 (Figure 1) with a Python script request (vaccin lang: fr), and several elements (tweet content, tweet id, author id, creation date) were stored in a document-oriented database (MongoDB). As queries can only contain a limited number of terms (1024 characters), it was more relevant to carry out a search for the word "vaccin" (i.e. vaccine) knowing that related terms were included by the Twitter API v2 search tools since November 15, 2021, rather than selecting a nonexhaustive keyword list. Indeed, Twitter's query tool collected all words containing the base word "vaccine" in French (i.e. *vaccin, vaccins, vaccination, vaccinations, vaccinat°, vacciner, vaccinés, vaccinées, vaccinerait, vaccineraient, pro-vaccin, anti-vaccin, #vaccin,* #vaccinationobligatoire, etc.). The goal of this approach was to collect all tweets containing the base word "vaccin" to explore their content in a bottom-up approach without additional inclusion/exclusion criteria. A total of 1,782,176 tweets were obtained, including 901,908 unique tweets (29,094 tweets per day) published by 231,373 unique users. To fully test CamemBERT, only unique tweets were included in the analysis. When dealing with the analysis of text (such as tweets), it is important to keep a great variability (vocabulary, syntax, length...) to strengthen deep learning algorithms. This variability will guarantee the power of model generalization. This is why, in this article, the 1,851 tweets that compose the dataset were drawn randomly from a set of 901,908 unique tweets.

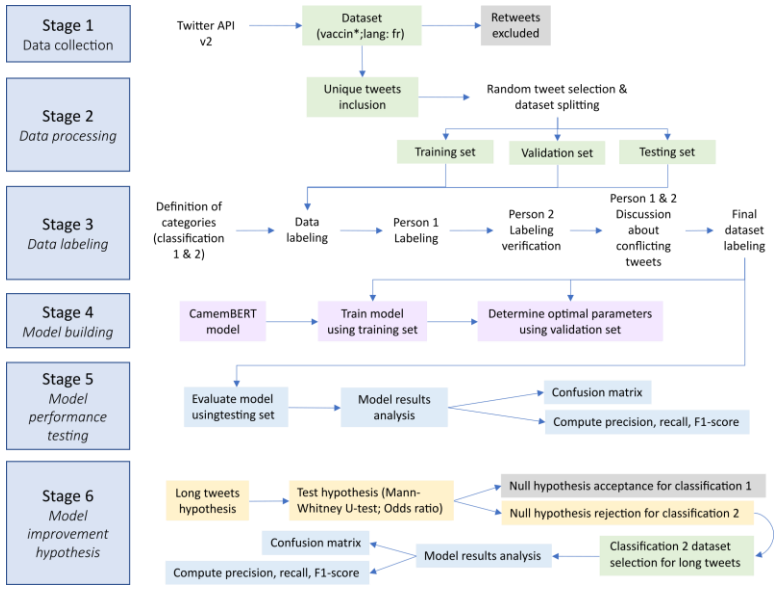

Figure 1: Flow chart of methodology steps.

Labeling

A total of 1,851 unique tweets were randomly selected and manually labeled by two people (1,451 for training and validation and 400 for testing). When doubt arose about labeling, which occurred in 87 of the 1,851 tweets (4.67%), a discussion occurred to determine the relevant label for each litigious tweet (see examples in Multimedia Appendix 1). Note that no duplicates were identified by the automated verification performed.

Two classifications were developed to examine arguments for (“pros”) or against (“cons”) vaccination (sanitary measures included) and to examine the type of tweet content (“scientific”, “political”, “social” or “vaccination status”). The classifications and definitions used to label tweets are provided below (Table 1) with translated examples of tweets for each label. In accordance with Twitter's terms of use under the European General Data Protection Regulation, original tweets cannot be shared [28].

Therefore, the translations have been adjusted to ensure the anonymity of Twitter users.

Table 1: Classification criteria and definitions.

Type of tweet		Definition	Translated examples (French to English)
Classification 1			
	Unclassifiable	Unclassifiable or irrelevant for the topic "vaccination" or "sanitary measures"	The Emmanuel Macron effect
	Noncommittal	Neutral or without explicit opinion on vaccination and/or on the sanitary pass	I have to ask my doctor for the vaccine
	Pros	Arguments in favor and/or on the sanitary pass Arguments in favor of the benefits of the COVID-19 vaccine and/or on the sanitary pass (efficiency, safety, relevance)	Personally, I am vaccinated so nothing to fear, on the other hand, good luck to all the anti-vaccine, you will not have the choice now??
	Cons	Arguments against vaccination or doubts about the effectiveness of the COVID-19 vaccine, fear of side effects and refusal of the sanitary pass	I am against the vaccine I am not afraid of the virus but I am afraid of the vaccine
Classification 2			
	Unclassifiable	Irrelevant or unclassifiable	A vaccine
	Scientific	Scientific or pseudoscientific content that using true beliefs or false information	The vaccine is 95% efficient, a little less in fragile people. The risk is not zero, but a vaccinated person has much less chance of transmitting the virus.
	Political	Comments on a legal or political decisions about vaccination or sanitary measures	Basically the vaccine is mandatory, shameful LMAO

	Social	Comments, debates or gives an opinion on the report to other members of society	"Pro vaccine" you have to also understand that there is people who does not want to be vaccinated.
	Vaccination status	Explicit tweet about the vaccination status of the tweeter's users Comments on the symptoms experienced after COVID-19 vaccine injection Explicit refusing a COVID-19 vaccine	Ex.1: I am very glad to have already done my 2 doses of the vaccine, fudge Ex.2: I don't want to get vaccinated. Why? Well, you know, we don't know what's in this vaccine, it can be dangerous.

Classification Method

This study followed the general methodology of machine learning to guarantee a rigorous building of the model. To ensure that the model did not overfit/underfit the dataset, the steps below were observed:

- the dataset was divided into training (N=1,306), validation (N=145) and testing (N=400) dataset;
- the training loss was represented as a function of the number of epoch to monitor the correct learning of the model and select its optimal value;
- the validation accuracy is represented as a function of epoch to ensure that the model was not overfitting/underfitting the data;
- the final model was evaluated on a testing dataset that had not been

Previously used to build or validate the model.

Two fully connected dense neural network layers with 1,024 and 4 neurons (for problem classification 1) or 5 neurons (for problem classification 2) were added to the head of the CamemBERT model, adding an additional 1.6 million parameters. Furthermore, to prevent overfitting, a 10% dropout was applied between those two layers. A small learning rate of $2x10^{-5}$ was used for the fine-tuning, and Adam with a decoupled

weight decay regularization [23] was chosen for the optimizer (see full codes used on GitHub [29]). The parameters were adjusted by minimizing the cross-entropy loss, which is a common choice when dealing with a classification problem. Fine-tuning was performed on a dataset consisting of the 1,451 labeled French tweets: 90% for training and the remaining 10% for validation. Once the model was built, it was tested on a new set of 400 labeled tweets from which a statistical analysis was made. Two classification models were built from the same dataset: one with four labels ("unclassifiable", "neutral", "positive" or "negative") related to a Twitter user's opinion about vaccination and one with five labels related to the type of tweet content ("unclassifiable", "scientific"," political", "social", "vaccination status" or "symptoms"). The proportion of each label for those two problems is given in Table 2 below. We see that the dataset is imbalanced, but it remains in a reasonable proportion. As such, it does not require a special treatment.

Table 2: Proportion of each class in the dataset for classification problems 1 and 2 (N=1,451).

Type of problem		**Number of tweets**
Problem classification 1, n (%)		
	Unclassifiable	189 (13.0)
	Neutral	354 (24.4)
	Positive	392 (27.0)
	Negative	516 (35.6)
Problem classification 2, n (%)		
	Unclassifiable	226 (15.6)
	Scientific	441 (30.4)
	Political	316 (21.8)
	Social	353 (24.3)
	Vaccination status	115 (7.9)

One of the main hyperparameters to be tuned for the training of the model is the number of epochs. As a rule of thumb, to prevent overfitting, the number of epochs is usually chosen when the abruptness of the slope of the loss changes while maintaining a low rate of misclassification on the validation dataset. Figure 2 shows that 7 epochs should lead to the best result.

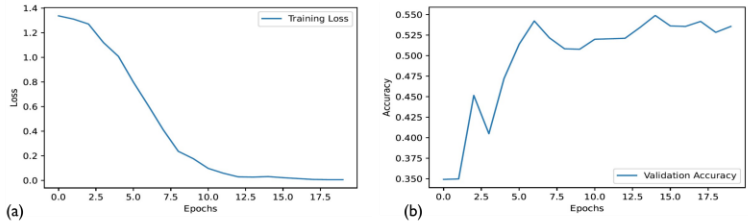

Figure 2: Training loss (a) and validation accuracy (b) of the model over 20 epochs for classification problem 1.

This was confirmed by computing the precision, recall and F1-score at three different epochs (7, 15 and 20), as shown in Table 3. Reported results were computed on the test dataset with 400 tweets. The average results over the classes were weighted to account for imbalanced classes in the dataset. As expected, the highest score was obtained with 7 epochs, however, not by a wide margin (Table 3).

Table 3: Classification performance of the model for classification problem 1

Epochs, n	Precision, %	Recall, %	F1-score, %
7	59.0%	55.3%	55.3%
15	56.6%	53.0%	53.2%
20	56.9%	54.5%	55.2%

A similar study for the second classification problem determined that 6 epochs was enough to prevent overfitting. The performance of the model was also measured by computing the weighted precision, recall and F1-score, as shown in Table 4.

Table 4: Classification performance of the model for classification problem 2

Epochs, n	Precision, %	Recall, %	F1-score, %
6	67.6%	64.5%	62.9%
15	62.7%	62.8%	61.3%
20	60.6%	59.5%	56.5%

The size of the datasets is quite similar to that of Kummervold et al [17] (N=1,633 tweets for training, and N=544 for testing) and Benítez-Andrades et al [26] (N=1,400 for training and N=600 for

testing). Furthermore, the benefit of using a pre-trained model such as CamemBERT is that a large dataset in not required to obtain good results. We also tried to build a neural network model from scratch with the same dataset, but the classification performance of the model was significantly below the results presented in this article with the CamemBERT model. On classification problem 1, we reached an accuracy of 33% (versus 59% with the pre-trained model) and on classification problem 2, we reached an accuracy of 40% (versus 67.6% with the pre-trained model).

Results

Statistical Analysis

From the results of the previous section, we see that it is significantly more difficult to build a performant classifier based on the four vaccine sentiment labels ("unclassifiable", "noncommittal", "pros", "cons"), with the maximum F1-score reaching 55.3% in that case. On the other hand, the classifier built from the same tweets but with five different labels based on content types ("unclassifiable", "scientific", "political", "social", "vaccination status" or "symptoms") reached a much higher F1-score (62.9%).

To analyze the strength and weakness of a model more finely, it is always instructive to represent it using a confusion matrix [24], as shown in Figure 3.

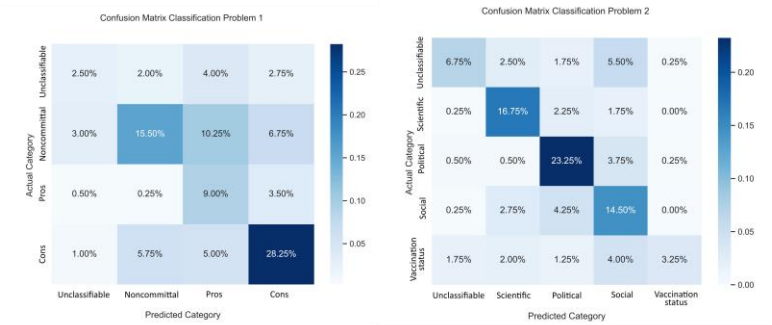

Figure 3: Confusion matrix for classification problems 1 and 2 (N=400).

Since the values in these matrices are percentages, their interpretation requires some care. For the first problem, summing figures line-by-line in the matrix shows that out of 100 tweets from the test dataset, on average, 11.25 are "unclassifiable", 35.50 are "noncommittal", 13.25 are "pros" and 40.00 are "cons". It is then possible to compute the proportion of tweets correctly classified by the model, label-by-label. The results are shown in Table 5 below. We see that the model can accurately classify the "pros" and "cons" tweets. It misclassifies a large part of the "unclassifiable" tweets and, to a lesser extent, the "noncommittal" tweets. Looking back to the confusion matrix, for the last two labels, we observe that the model tends to classify the tweets as being "pros".

Table 5: Correct classification for each label of classification problems 1 and 2 (N=400).

Type of problem		**Number of tweets**
Problem classification 1, n (%)		
	Unclassifiable	10 (22.2)
	Noncommittal	62 (43.7)
	Pros	36 (67.9)
	Cons	113 (70.6)
Problem classification 2, n (%)		
	Unclassifiable	27 (40.3)
	Scientific	67 (79.8)
	Political	93 (82.3)
	Social	58 (66.7)
	Vaccination status	13 (26.5)

For the second problem, as expected, in coherence with the higher F1-score found in the previous section, the model achieves much better classification performance. It excels at classifying "scientific" and "political" tweets and is also good for social tweets. It still has some difficulties classifying "unclassifiable" tweets and, in a larger proportion, the "vaccination status" tweets. Looking back to the confusion matrix, for the last two labels, we observe that the model tends to classify them as being "social" tweets.

Text Size Analysis

To improve the performance of the fine-tuned CamemBERT model, a hypothesis of the influence of the tweet size on the accuracy was tested. A Mann–Whitney U-test was statistically significant on classification problem 2 (U=21,202; *p*=.004) and not on classification problem 1 (U=19,284; *p*=.79). As Figure 4 shows, the correctly predicted tweets are significantly longer in classification problem 2. A second analysis carried out on a dichotomous variable created from the tweet text size (greater than or less than 170 characters) confirmed this significance for classification problem 2. A tweet was 1.86 times more likely to be incorrectly predicted by the model if it contained less than 170 characters (odds ratio = 1.86; 1.20 < 95% confidence interval < 2.86). Therefore, the significance obtained using these two analyzes (Mann-Whitney U-test and Odds ratio) allows us to rigorously validate [30] our hypothesis.

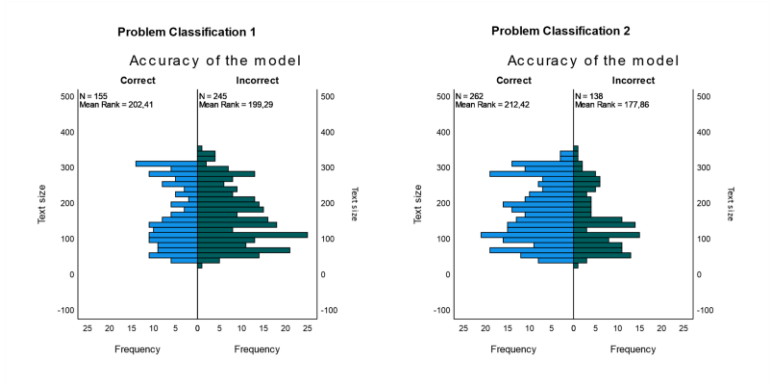

Figure 4: Tweet text size as a function of the accuracy of the fine-tuned CamemBERT model conducted on classification problems 1 and 2 (Mann–Whitney U-test).

Long Tweet Test

The finding of the previous section is further supported after carrying out the following experiment. Tweets with more than 170 characters were selected from the 400 tweets dataset.

Classification model 2 was then tested with these 168 tweets to see if its accuracy performance increased.

As shown in Table 6, the accuracy improved from 64.5% to 73.2% (an 8.7% increase), confirming our hypothesis. The F1-score also increased by approximately the same amount.

Table 6: Classification performance of the model for classification problem 2 limited to long tweets (170 or more characters).

Classification problem 2	Precision, %	Recall, %	F1-score, %
	72.6%	73.2%	72.4%

The confusion matrix generated from the comparison between the model-classified and the manually-classified 168 long tweets is shown in Figure 5. From this matrix, it is possible to compute the percentage of correct classification for each label, the results of which are shown in Table 7. The increase in accuracy performance is significant for the "vaccination status" label (+9.2%), followed by the "political" label (+7.7%) and the "unclassifiable" label (+6%).

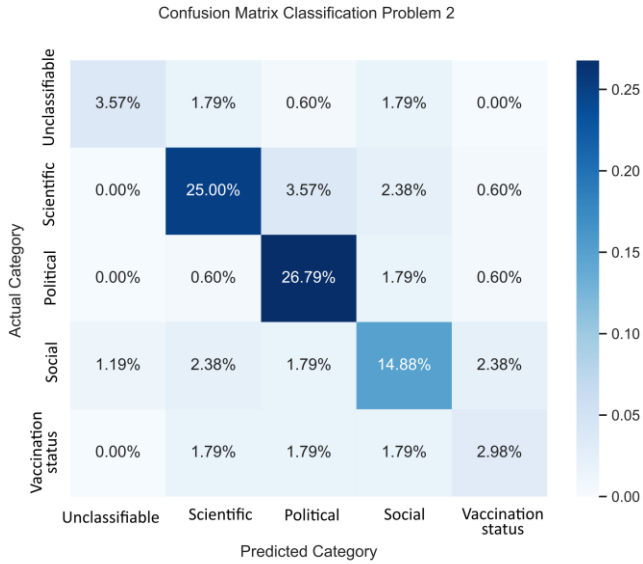

Figure 5: Confusion matrix for classification problem 2 limited to long tweets (N=168).

As already pointed out using the Mann–Whitney U-test and Odds ratio test, the model for the second problem has much better classification performances with long tweets. It should be noted that the rate of good classification of “political” labels reached an impressive 90% (45/50).

Table 6: Correct classification for each label of classification problem 2 limited to long tweets (170 or more characters) (N=168).

Type of problem		**Number of tweets**
Problem classification 2, n (%)		
	Unclassifiable	6 (46.3)
	Scientific	42 (79.2)
	Political	45 (90.0)
	Social	25 (65.8)
	Vaccination status	5 (35.7)

Discussion

Principal Findings

Two types of classification were examined. The accuracy of the model was better, with the second classification (67.6%; F1-score 62.9%) than the first classification (59.0%; F1-score 55.3%). This accuracy is slightly higher than that obtained by BERT for the same topic (vaccine) [17] and in the same range of previous findings [16,25]. However, CamemBERT obtained a better accuracy (78.7-87.8%) in a study using dichotomous labels on eating disorders and using a preprocessing step reducing the initial number of tweets by two [26]. However, by limiting the analysis to long tweets (170 or more characters in accordance with the statistical analysis conducted on the performance of the model), the accuracy of classification model 2 improved significantly (from 62.9% to 72.4% for the F1-score).

Therefore, as shown by Kummervold et al. [17], the classification choices have a significant influence on the accuracy of a model. As in other research areas, the vaccine hesitancy debate crystallizes the opposition. The "pros" and "cons" sides argued using tweets after the announcement by the French president of the implementation of a health pass during

the month of July 2021. The mobilized arguments were scientific or pseudoscientific to justify or contest this political decision. Several Twitter users participated in the debate to convince antivaxxers to get vaccinated. Another group of users participated to joke about or ironize the positions of each.

Consequently, tweet content is so varied that it remains difficult to manually categorize, and this has been reflected in the model predictions. On the one hand, considering classification problem 1, tweets containing characteristic terms of the antivax position, such as "5G", "freedom", "phase of testing", "side effect" and "#passdelahonte" ("shameful pass"), were found to be easier to label and predict. However, because antivaxxers spread disinformation more widely on social media [27], the position of the provaxxers is less polarized [7], which reduces the model precision because the terms are less singular. On the other hand, considering classification problem 2, the classes were more distinctive since their lexical fields did not overlap. Indeed, when Twitter users commented on political decisions, the terminology used was different from that used to mobilize scientific or pseudoscientific arguments. Moreover, the "scientific" and "political" labels were best predicted by the model (79.8% (67/84) and 82.3% (93/113), respectively).

Finally, relevant tweets for a topic may be rare in a dataset. In some studies, the corpus is halved [13], while in others, only 0.5% (4,000/810,600) of downloaded tweets were included in the analysis [16]. It would be interesting to find an objective method to improve model predictions without drastically reducing the dataset. The approach of limiting tweet size can be an option, as we have demonstrated in this article.

Limitations

Several limitations can be highlighted: (1) the data was only provided from a unique social media platform (Twitter); (2) all tweets containing the term "vaccine" and its derivatives were included without preselection; (3) several categorization classes were unbalanced; (4) a larger training set could provide contrasting results; (5) the categorization choices could affect the

performance of CamemBERT, as seen in the confusion matrix; and (6) the suggestions provided (limiting the number of tweet characters) may only apply to vaccination tweets, so further studies are needed to confirm the relevance of our conclusions.

Conclusions

In this study, we tested the accuracy of a model (CamemBERT) without preselecting tweets and elaborate an epistemological reflection for future research. When the vaccine debate is jostled by contested political decisions, tweet content becomes so heterogeneous that the accuracy of the model decreases for the less differentiating classes. In summary, our analysis shows that epistemological choices (type of classes) can affect the accuracy of machine learning models. However, our tests also showed that it is possible to improve the model accuracy by using an objective method based on tweet size selection. Other possible avenues for improvement remain to be tested, such as the addition of features provided by Twitter (conservation ID, number of Twitter users following or followers, user public metrics listed count or user public metrics tweet count user ID).

References

1. Sauvayre R. Ethics of belief, trust and epistemic value. The case of the scientific controversy surrounding the measles vaccine. J. Leadersh Account Ethics. 2021; 18: 24-34.
2. Kata A. A postmodern Pandora's box: Anti-vaccination misinformation on the Internet. Vaccine. 2010; 28: 1709-1716.
3. Kata A. Anti-vaccine activists, Web 2.0, and the postmodern paradigm – An overview of tactics and tropes used online by the anti-vaccination movement. Vaccine. 2012; 30: 3778-3789.
4. Aquino F, Donzelli G, De Franco E, Privitera G, Lopalco PL, et al. The web and public confidence in MMR vaccination in Italy. Vaccine. 2017; 35: 4494-4498.
5. Deiner MS, Fathy C, Kim J. Facebook and Twitter vaccine sentiment in response to measles outbreaks. Health Informatics J. 2019; 25: 1116-1132.

6. Bavel JJV, Baicker K, Boggio PS. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav. 2020; 4: 460-471.
7. Sear RF, Velásquez N, Leahy R. Quantifying COVID-19 Content in the Online Health Opinion War Using Machine Learning. IEEE Access. 2020; 8: 91886-91893.
8. Kanozia R, Arya R. "Fake news", religion, and COVID-19 vaccine hesitancy in India, Pakistan, and Bangladesh. Media Asia. 2021; 48: 313-321.
9. Puri N, Coomes EA, Haghbayan H, Gunaratne K. social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccines Immunother. 2020; 16: 2586-2593.
10. Edwards B, Biddle N, Gray M, Sollis K. COVID-19 vaccine hesitancy and resistance: Correlates in a nationally representative longitudinal survey of the Australian population. PLoS One. 2021; 16: e0248892.
11. Antonakaki D, Fragopoulou P, Ioannidis S. A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks. Expert Syst Appl. 2021; 164: 114006.
12. Hu T, Wang S, Luo W. Revealing Public Opinion Towards COVID-19 Vaccines With Twitter Data in the United States: Spatiotemporal Perspective. J Med Internet Res. 2021; 23: e30854.
13. Alshalan R, Al-Khalifa H, Alsaeed D, Al-Baity H, Alshalan S. Detection of Hate Speech in COVID-19–Related Tweets in the Arab Region: Deep Learning and Topic Modeling Approach. J Med Internet Res. 2020; 22: e22609.
14. D'Aniello G, Gaeta M, La Rocca I. KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis. Artif Intell Rev. 2022.
15. Küçük D, Can F. Stance Detection: A Survey. ACM Comput Surv. 2020; 53: 1-12:37.
16. Visweswaran S, Colditz JB, O'Halloran P. Machine Learning Classifiers for Twitter Surveillance of Vaping: Comparative Machine Learning Study. J Med Internet Res. 2020; 22: e17478.
17. Kummervold PE, Martin S, Dada S. Categorizing Vaccine Confidence With a Transformer-Based Machine Learning

Model: Analysis of Nuances of Vaccine Sentiment in Twitter Discourse. JMIR Med Inform. 2021; 9: e29584.

18. Martin L, Muller B, Suárez PJO. CamemBERT: a Tasty French Language Model. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020 Jul 5-10; Online. Association for Computational Linguistics. 2020; 7203-7219.
19. Le H, Vial L, Frej J. FlauBERT: Unsupervised Language Model Pre-training for French. Proceedings of the 12th Language Resources and Evaluation Conference. 2020 May 13-15. Marseille, France. European Language Resources Association. 2020 ; 2479-2490. Available online at : https://aclanthology.org/2020.lrec-1.302
20. Liu Y, Ott M, Goyal N. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv. 2019. Available online at: http://arxiv.org/abs/1907.11692
21. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019 June 2-7. Minneapolis, USA. Association for Computational Linguistics. 2019; 4171-4186.
22. Vaswani A, Shazeer N, Parmar N. Attention Is All You Need. Proceedings of 31st Conference on Neural Information Processing System. 2017 Dec 4-9. Long Beach, USA: Curran Associates Inc. 2017; 6000-6010. Available online at: https://dl.acm.org/doi/pdf/10.5555/3295222.3295349
23. Loshchilov I, Hutter F. Decoupled Weight Decay Regularization. Proceedings of the 7th International Conference on Learning Representations. 2019 May 6-9. New Orleans, Louisiana, USA. 2019. Available online at: https://openreview.net/pdf?id=Bkg6RiCqY7
24. Shin D, Kam HJ, Jeon MS, Kim HY. Automatic Classification of Thyroid Findings Using Static and Contextualized Ensemble Natural Language Processing Systems: Development Study. JMIR Med Inform. 2021; 9: e30223.

25. Kumar A, Singh JP, Dwivedi YK, Rana NP. A deep multi-modal neural network for informative Twitter content classification during emergencies. Ann Oper Res. 2020.
26. Benítez-Andrades JA, Alija-Pérez JM, Vidal ME, Pastor-Vargas R, García-Ordás MT. Traditional Machine Learning Models and Bidirectional Encoder Representations From Transformer (BERT)–Based Automatic Classification of Tweets About Eating Disorders: Algorithm Development and Validation Study. JMIR Med Inform. 2022; 10: e34492.
27. Germani F, Biller-Andorno N. The anti-vaccination infodemic on social media: A behavioral analysis. PLoS One. 2021; 16: e0247642.
28. Twitter Controller-to-Controller Data Protection Addendum. GDPR Twitter. Available online at: https://gdpr.twitter.com/en/controller-to-controller-transfers.html
29. NLP-French-model-for-vaccine-tweets. Github. 2022. Available online at: https://github.com/cdchauvi/NLP-French-model-for-vaccine-tweets
30. Significant debate. Nature. 2019; 567.

Multimedia Appendix 1

Examples of conflicting labeling.

4.7% of tweets (87/1851) have been conflictual and needed to be discussed. In accordance with Twitter's terms of use under the European General Data Protection Regulation, these original tweets cannot be shared. However, several types of problems during labeling have been identified and are presented below with modified and selected examples to ensure the anonymity of Twitter users.

Type 1 : the tweets lack information to identify the label

It is a common type encountered. Lack of information conducted us, after verification, to label tweets as unclassifiable or noncommittal.

Example 1:

If the vaccine is so effective why not vaccinate only person's risk of severe illness.

Example 2:

A scary vaccine! So the variants do not scare you!

Type2: the "cons" arguments unidentified by one of the two labelers

It is a common type encountered. The sharing of knowledge on the subject and the verification of the labeling criteria allowed these tweets to be classified.

Example 3: freedom is a strong argument among people against vaccination or health measures

Liberty! We are against blackmail, threats, restrictions on our private life.

Example 4: minimizing the effects of COVID and the request of

treatment is a strong argument among people against vaccination or health measures.

No one is asking for a vaccine. We demand treatment. Let's stop saying that the virus kills: the vast majority do well.

Example 5: the spread of the virus among vaccinated people is a strong argument among people against vaccination or health measures

The vaccine does not prevent transmission of the virus.

Type 3 : the "pro" arguments unidentified by one of the two labelers

The sharing of knowledge on the subject and the verification of the labeling criteria allowed these tweets to be classified.

Example 6: arguments in favor of vaccination, but which is close to the arguments used by the "cons"

But the virus has not been eradicated and continues to spread with a mutation's risk. for less than 80% of vaccinated people.

Type 4 : tweets recalling the measures taken during the announcement of the President of the French Republic (health pass, vaccination obligation) or commenting on what the French people are doing.

This type of tweets conducted us, after verification, to label tweets as unclassifiable or noncommital Example 7:

Will need the vaccine from August to travel by plane, train and bus.

Example 8:

Reinforced health pass, but the young people will obtain special conditions.