

Public Opinion on Vaccines during COVID19 pandemic through Twitter (324 words)


Himani Anil Deshpande

Research Question

What is the public opinion of population in USA about Vaccinations developed during the COVID19 pandemic?

What are the major topics or sentiments of population about these vaccinations?

Data

I will be using the **Twitter Developer API**  to gather at least 1000 random English tweets with geographical location of USA about COVID19 pandemic and COVID19 Vaccines from August 2020 to April 2021 using below collected hashtags and keywords.

#CovidVaccine, vaccination, vaccinations, vaccine, vaccines, immunization, vaccinate, vaccinated, #COVID-19, coronavirus, CORONA, covid-19 Vaccines, SARS-CoV-2, vaccination Refusal, #Pandemic, vaccination drive, immunity, immune, rollout, side effect, immunise, immunize, immunisation. [1,2,3,4]

I will be storing the text of the tweet, location, the timestamp, retweets and likes. Using this paper [4], I will be preprocessing the data to remove stopwords, URLs, usernames or user mentions and special character like emoticons as it shows that the data cleaning process helps in giving better results. However, I will not be cleaning data for hashtags as they are part of a sentence and provides information. [3]

Methods

I will use EMPATH [5] for the 200 built in normalized lexical categories it provides. Out of these 200 categories I will select the top 10 categories. These will be used for labelling each tweets that I have collected. I will also be using unsupervised classification model like LDA [6](Latent Dirichlet Allocation) to find the top 5 important topics that the dataset of tweets that I have collected. This model is freely available in sklearn [7] or gensim [8] library.

Another supervised classification model like SVM(Support vector Machine) [9], which uses the technique of separation of hyperplanes which is also available in sklearn [7] library will be evaluated. I will use the labels that I got from EMPATH [5] to train this model.

I will split the collected dataset into train and test sets and then compare the different models based on their scores and accuracy to find which performs better after tuning the hyperparameters like kernels [9] or alpha [8] to get the best performance from each model.

References

1. Lyu J, Han E, Luli G (2021, June 6) COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. J Med Internet Res 2021;23(6):e24435 [URL](#) DOI: 10.2196/24435
2. Ansari, Md Tarique Khan, Naseem. (2021). Worldwide COVID-19 Vaccines Sentiment Analysis Through Twitter Content. Electronic Journal of General Medicine. 18. em329. 10.29333/ejgm/11316. [URL](#)
3. Jang H, Rempel E, Roe I, Adu P, Carenini G, Janjua N (2022, March 29) Tracking Public Attitudes Toward COVID-19 Vaccination on Tweets in Canada: Using Aspect-Based Sentiment Analysis. J Med Internet Res 2022;24(3):e35016 [URL](#) DOI: 10.2196/35016

4. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, Ali A, Sheikh A (2021, April 5) Artificial Intelligence–Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study J Med Internet Res 2021;23(4):e26627 [URL](#) DOI: 10.2196/26627
5. Fast, E., Chen, B., Bernstein, M.S. (2016). Empath: Understanding Topic Signals in Large-Scale Text. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.
6. Blei, D. M., Ng, A. Y. Jordan, M. I. (2003). Latent dirichlet allocation. J. Mach. Learn. Res., 3, 993–1022. doi: <http://dx.doi.org/10.1162/jmlr.2003.3.4-5.993>
7. Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . others. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825–2830.
8. Řehůřek, R. (2019). Gensim: Topic modelling for humans. URL: <https://radimrehurek.com/gensim/>[accessed 2021-05-30].
9. Lilleberg J., Zhu Y. and Zhang Y., (2015) "Support vector machines and Word2vec for text classification with semantic features", IEEE 14th International Conference on Cognitive Informatics Cognitive Computing (ICCI*CC), pp. 136-140, 2015.