A special thanks to our Sponsor:

F5 is a security company committed to bringing a better digital world to life.

Program Chairs

Sudha Jamthe
Senior Program Chair
Technology Futurist & AI and
AV Instructor at Stanford

Susanna Raj
Program Chair
CEO & Founder of AI4Nomads
Cognitive Science & AI Ethics
Researcher

Yashaswani Viswanath
Program Chair
AI Researcher and AI Product
Engineer
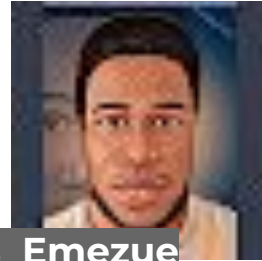Champion for Responsible AI
Author of "100 days of XAI"

Pariya Sarin
Program Chair
High School Student
passionate about Computer
Science

Peer Review: Area Chairs

Tahereh Saheb

Chris Emezue

Jamzia Henry

Krishna Srinivasan

Bonaventure Dossou

Building Technologies for African languages

Jazmia Henry

Operationalizing post production chatbots

Identifying COVID-19 Vaccine Hesitancy Hotspots in Nigeria: Analysis of Social Media Posts

Blessing Ogbuokiri[a], Ali Ahmadi[b], Zahra Movahedi[a], Bruce Mellado[c], Jiahong Wu[a], James Orbinski[a], Ali Asgary[a], and Jude Kong[a]

[a]York University, Toronto, Canada.
[b]K. N. Toosi University of Technology, Iran.
[c]University of the Witwatersrand, Johannesburg, South Africa

YORK UNIVERSITÉ UNIVERSITY U

1. Research Problem

Ⓞne of the major challenges faced by health policymakers in the fight against community-based infectious diseases, such as COVID-19, Malaria, Monkeypox, and Marburg, is vaccine hesitancy.

In Nigeria, Twitter is one of the social media platforms used to promote anti-vaccination posts. Anti-vaccination posts or reactions on Twitter can lead to a compromise of community confidence or lack of willingness in taking the vaccine during an outbreak.

Can sentiment analysis be conducted on vaccine related tweets to inform geolocation clustering for the identification of hotspots in Nigeria?

2. Method

Data Collection
10,000 historical tweets about vaccines and vaccination were collected using Twitter API + python script from December 2020 to February 2022.
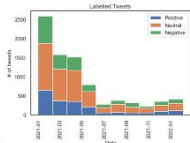
Data preprocessing
Dataset to dataframe – Pandas v1.2.4, Noise removal - tweets-preprocessor toolkit v0.6.0 and NLTK v3.6.2. Tokenization - Spacy2 toolkit v3.2.
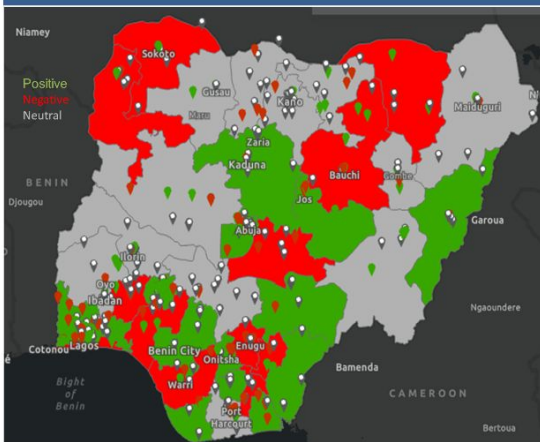
Sentiment Analysis
Tweet labelling- VADER v1.4.0, Manual labelling - (10%) Tweets, Classification

Visualization
Hotspot Analysis, Geocoordinate, ArcGis Online

Labelled Tweets

4. Hotspot locations

Positive
Negative
Neutral

Word Cloud

Positive
Negative
Neutral

The point-based location technique

$$S_c = \sum_{j=1}^{n} \sum_{i=1}^{m} X_{i,j}$$

Geo-coordinate calculation (bbox)

$$\text{Latitude:} \quad \frac{7.3045544 + 7.5436339}{2} = 7.424094$$

$$\text{Longitude:} \quad \frac{8.961887 + 9.1854407}{2} = 9.073064$$

5. Conclusions

The outcome of this research shows that social media data can be used to complement existing data in identifying hotspots during an outbreak. It can also be used to inform health policy in managing vaccine hesitancy.
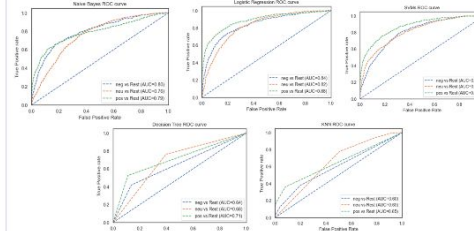
6. References

1. Thorakkattil, S. A., Abdulsalam, S, Karattuthodi, M. S., Unnikrishnan, M. K., Rashid, M, and Thunga, G., "COVID-19 Vaccine Hesitancy: The Perils of Peddling Science by Social Media and the Lay Press", *Vaccines*, Vol. 10, No 1059,pp1-7, 2022.

2. Honnibal, M., "spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural net- works and incremental parsing", Sentometrics Research, Vol. 1, No. 1, pp 2586–2593, 2017.

3. Results

Model Performance of Tweet Classification

SN	Machine Learning Classifiers	Accuracy (%)	Average F1-Score (%)	Average AUC (%)
1	NB	66	63	78
2	LR	71	68	85
3	SVMs	65	59	83
4	DT	61	58	67
5	KNN	56	41	63

ROC–AUC for different tweet sentiments classifiers

Limitation

Most NLTK for sentiment analysis techniques do not have the capacity to properly label figurative language, such as sarcasm. However, since the approach we used was able to label and score a large amount of the tweets in our dataset and was verified with the manual labeling of randomly selected (10%) of the tweets, in addition to the 71% accuracy achieved with the LR classifier, we assume it was able to deal with the noise generated by this obvious challenge.

Contact: Blessing Ogbuokiri
Department of Mathematics and Statistics, York University, Toronto, Canada
Email: blessogb@yorku.ca, Phone: 6473382025

IDRC·CRDI International Development Research Centre
Global South AI4COVID Program
Sida
Canada

Our Authors

Yashaswini Viswanath
Yine Nyika
Chinmayi Ramasubramanian
Chandra Sekhar Gupta
Nishant Kumar
Ali Hussein
Klinsman Nang'ore Agam
Smital Lunawat
Smrithi Suresh
Suresh Lokiah
Sanjana Melkote
Ricardo Alanis Tamez
Ivan Vladimir Meza Ruiz

Sundaraparipurnan Narayanan
T Pranav
Jihad Baiti
Rdouan Faizi
Youssef Hmamouche
Amal El Fallah-Seghrouchni
Amanuel Mersha
Muskan Mahajan
Blessing Ogbuokiri
Zahra Movahedi
Bruce Mellado
Jiahong Wu
James Orbinski
Jude Kong
Constance Muthuri