Areeba Umair

Devising Artificial Intelligence Tools for Complex Data

Tutor: Prof. Elio Masciari

Cycle: XXXVI

Year: Second

My background

- MS in Computer Science (MSCS) from Pakistan
- Currently PhD student of the ITEE program
- Ph.D. started on 1st November 2020
- UNINA Scholarship

Research field of interest

My research field of interest is

- Data analytics

a) Sentimental analysis

b) Social media analysis

Research activity: Overview

- Problem: “COVID-19 sentimental analysis using social media data”

The social media has a vast amount of user-generated data.

Sentimental analysis is the field where people’s feeling are extracted.

COVID-19 pandemic has affected people’s lives all over the globe.

It caused the feelings of fear, anxiety, anger, depression and other issues.

Research activity # 1

PROBLEM: Sentimental analysis of COVID-19 Vaccines tweets using BERT+NBSVM model

Vaccine hesitancy means to show unwillingness for vaccine intake. The main causes of unwillingness are people's mistrust or misinformation. It is a hurdle in the control of COVID-19 in many countries. To understand the acceptance of vaccines in the population, is the important to analyze the people's sentiments. Therefore, it is necessary to know about the people's reactions and sentiments before the designing of vaccine policy and campaigns.

OBJECTIVES:

- Using freely available twitter data about COVID vaccines and categorize the text into different sentiment classes.
- To categorize the tweets based on their polarity values using python script.
- To propose BERT+NBSVM model for positive and negative tweet classification.

Research activity # 1

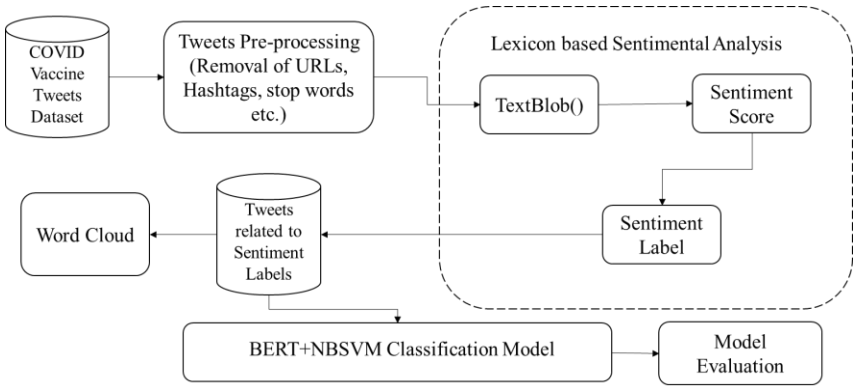
METHODOLOGY:

Phase 1: consists of collection of dataset and pre-processing steps.

Phase 2: Extraction of sentiments and their polarity value using TextBlob().

In **phase 3**, social network analysis is performed, which establishes relationships between two or more entities at a time and draw word clouds.

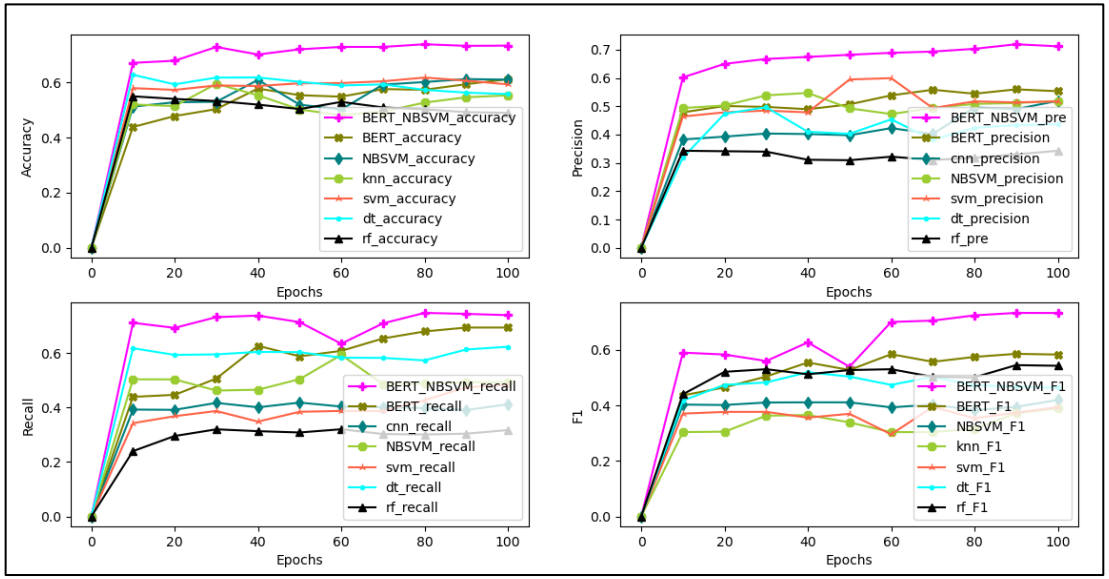
Phase 4: used polarity values and performed sentimental classification with the help of BERT+NBSVM model.

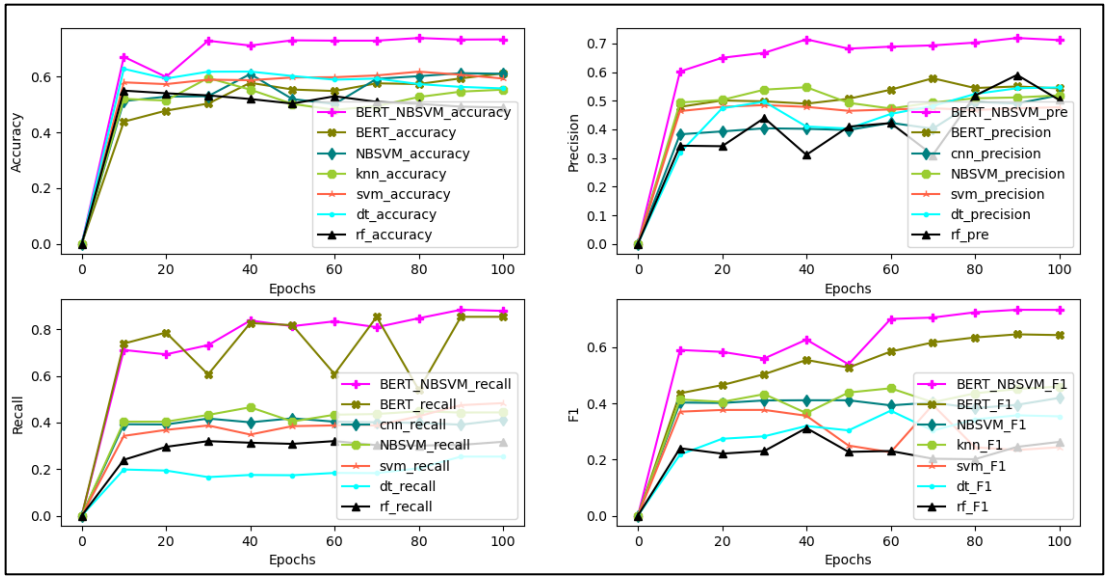


Research activity # 1

RESULTS:

The proposed BERT+NBSVM outperformed other models and achieved 73 % accuracy, 71 % precision, 88 % recall and 73 % F-measure for classification of positive sentiments while 73 % accuracy, 71 % precision, 74 % recall and 73 % F-measure for classification of negative sentiments respectively.





Umair, Areeba, et al. "Sentimental Analysis of COVID-19 Vaccine Tweets using BERT+ NBSVM." NFMCP 2022.

Umair, A., et al., Vaccine Sentimental Analysis using BERT+NBSVM and Geo-Spatial Approaches, Journal of supercomputing (2022). IF=2.557 (Under-review)

Research activity # 2

PROBLEM: Spatial Analysis of COVID-19 Vaccine Tweets

Today, GIS based modeling is performed using digital as well as electronic big data. Vaccine data sharing is the foremost step in preparing, controlling monitoring and recovery of disease. As infectious disease phenomena are greatly related to spatial and temporal factors. Web based GIS have provided opportunities to visualize disease control and vaccine information over maps. The web-based tools have caused a revolution in the history of disease mapping and controlling using - systems (GIS). The big electronic and print data can be visualized in interactive and real time dashboards, which can help to protect human lives.

OBJECTIVES:

- To geo-code the addresses of the twitter user
- To visualize the vaccine data on Map
- To perform the spatial analysis of vaccines related data using Geo-spatial approaches.

Research activity # 2

METHODOLOGY:

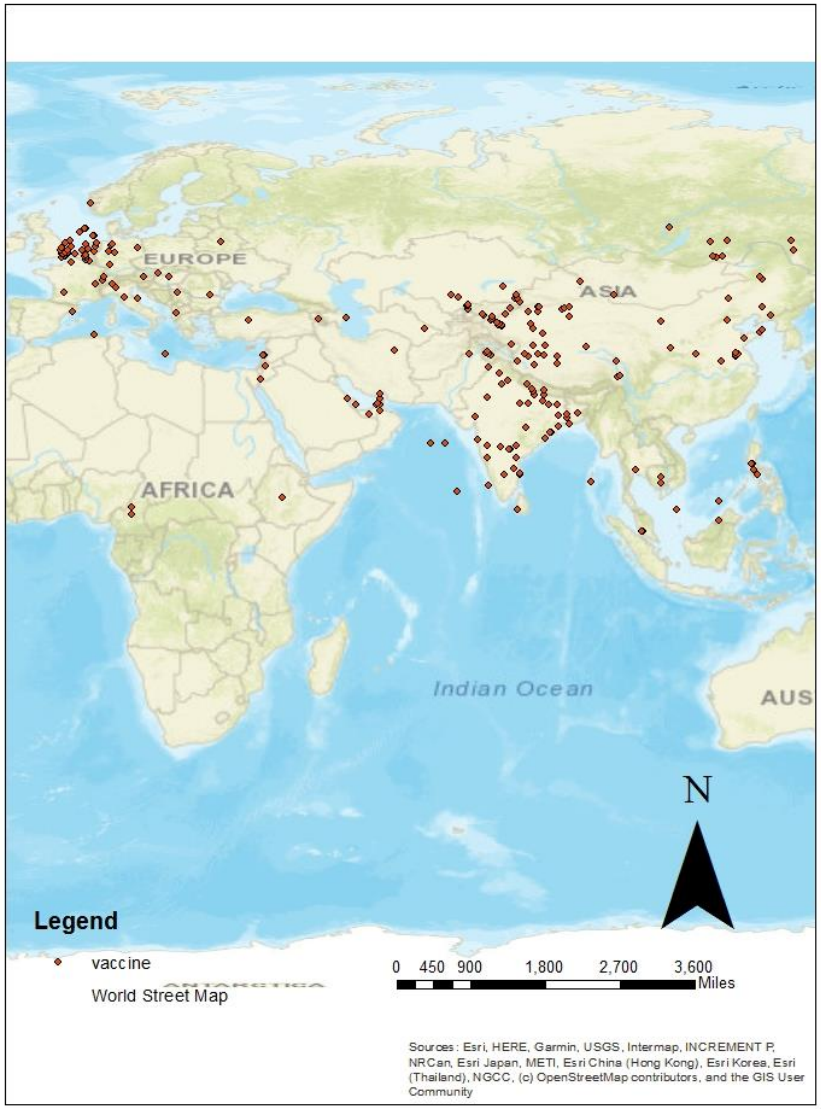
Methodology is the geo-coding of the data and the spatial analysis of the tweets considering the sentimental polarity of the tweets.

RESULTS:

1. **Geo-coding and visualization of data**

We performed Geo-coding using Geo-Py module of Python.

Vaccine tweets data was visualized using ArcGIS 10.5

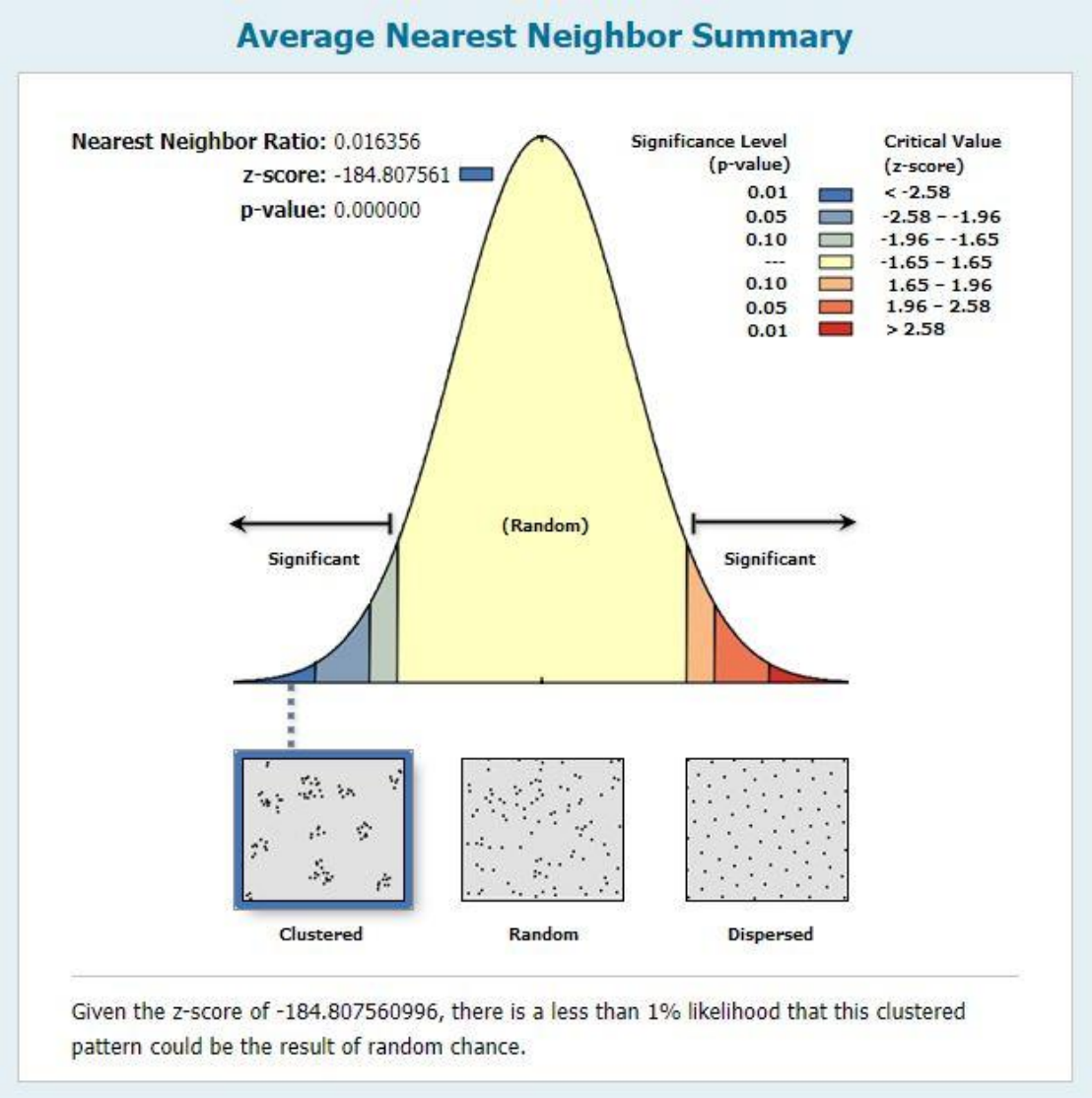


Research activity # 2

2. Analyzing the patterns using Average Nearest Neighbour

We have applied the ANN tool over vaccine tweet dataset to find if either feature set is geographically significant of not.

Figure shows that our tweets data is scientifically clustered which means that we can perform certain spatial operations on the data



Research activity # 2

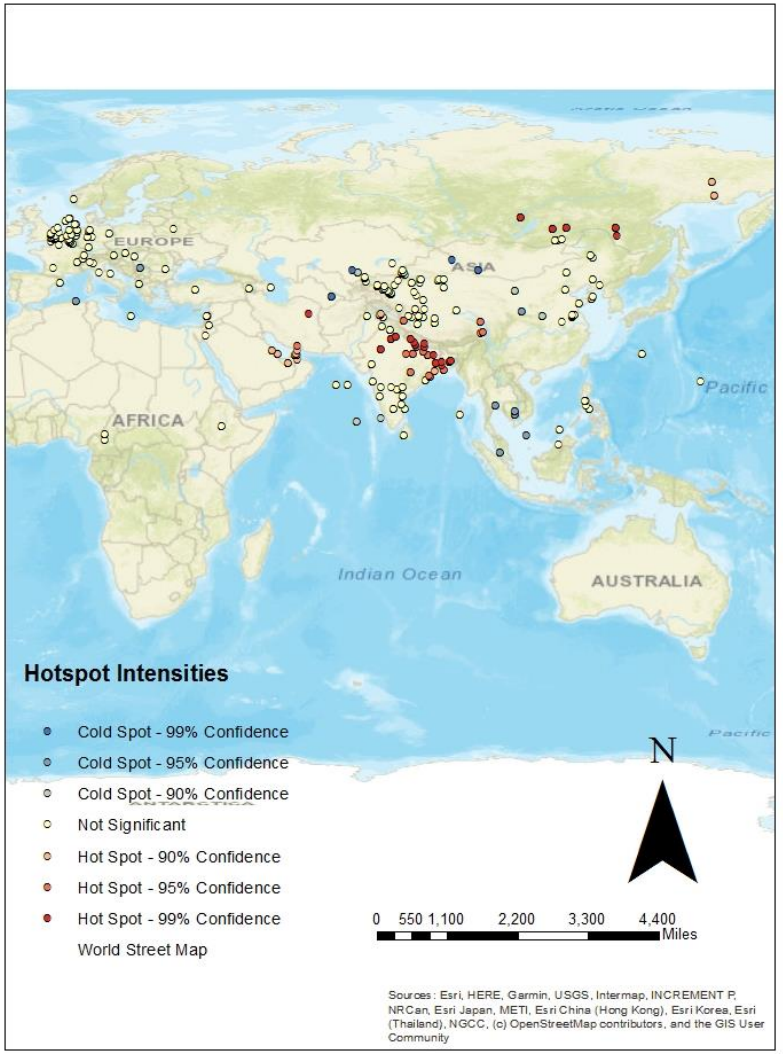
3. Hotspot analysis

Hotspot shows the geographical areas where the vaccine sentiment polarity is high in rate while cold spots shows the areas with less vaccine sentiment polarity.

Positive: India, Saudi

Neutral: Europe

Negative: China

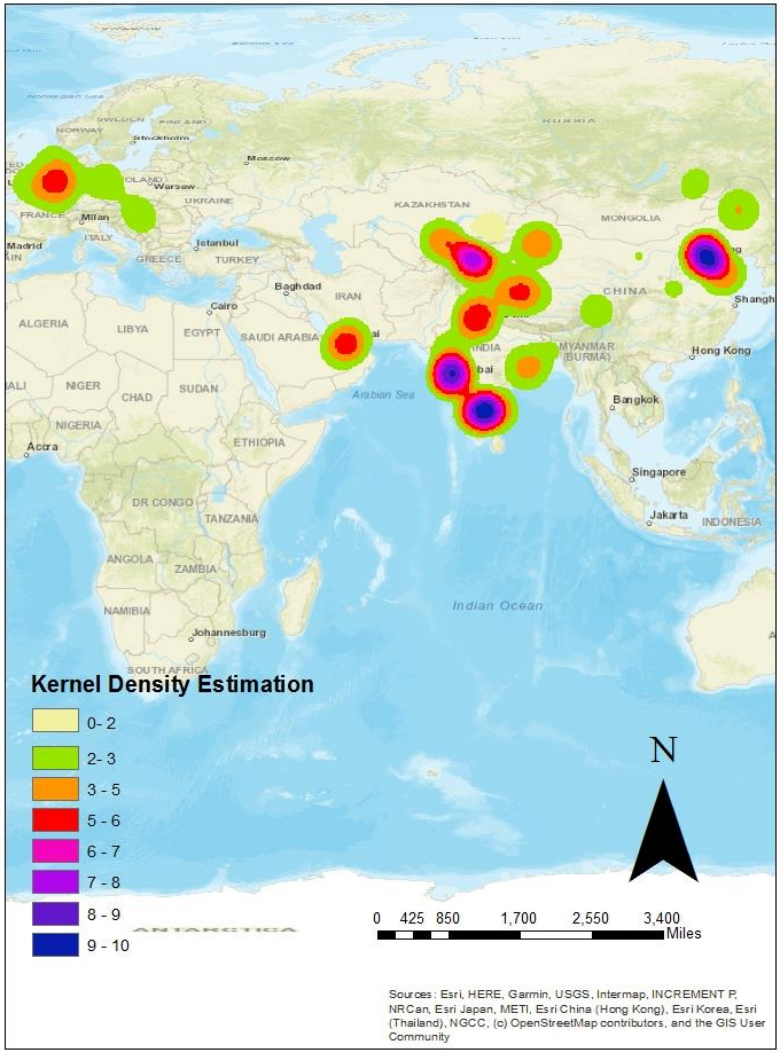


Research activity # 2

4. Analysis using kernel density estimation

It calculates the density of features in a neighborhood around those features.

It also shows that the more positive sentiment polarity is found in India and Europe behave neutrally in this context.

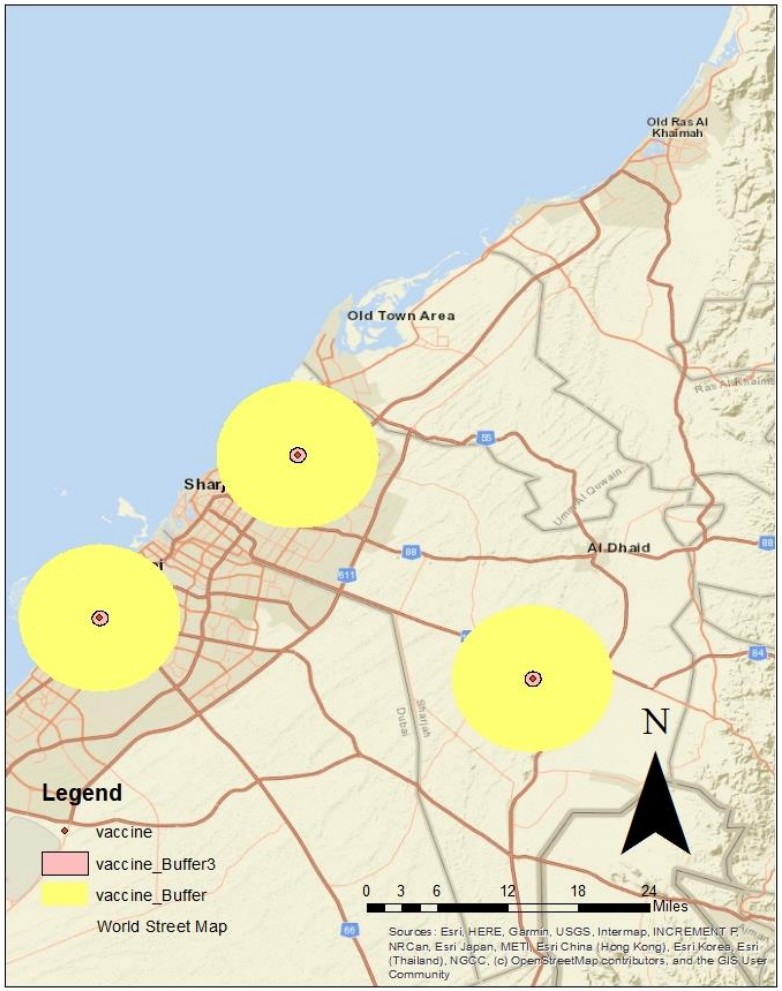


Research activity # 2

5. Buffering

One hurdle for vaccination may be the mobility. People are unable to reach the vaccination center. Buffering can be used to address allocation Problems.

Figure shows the suitability of people to move for vaccination. It is assumed that vaccination centers must be within 5 km of their residence address.



Umair, Areeba, et al. "Sentimental and spatial analysis of COVID-19 vaccines tweets." Journal of Intelligent Information Systems (2022): 1-21 (IF=2.504)

Research activity # 3

PROBLEM: Sentimental Analysis of Arabic Tweets related to COVID-19 using AraBERT model

Every country around the world is facing various challenges amid COVID-19 and trying to implement some measure in order to control the spread of COVID-19. Like people of all other counties and cultures, Arab people also shared their thoughts on social media during COVID-19. However, the Arabic language is rich in morphology and have large number of dialects. Arabic Sentimental Analysis has attracted many researchers but still the research work on Arabic Language is limited.

OBJECTIVES:

- Performing the pre-processing of Arabic Text, as it has complicated morphology.
- Create positive and negative wordcloud for Arabic Tweets.
- Sentiment Classification of Arabic Tweets using AraBERT model.

Research activity # 3

METHODOLOGY:

In this research, we used AraCOV-19 data, an Arabic language dataset related to COVID, and assigned the sentiments labels to each tweets using Textblob-ar, after doing necessary pre-processing. We drew two word-clouds i.e. positive and negative. We classify the negative and positive tweets using AraBERT model and compared the results with state-of-the-art models. The results showed that our model outperformed all other state-of-the-art models by achieving maximum accuracy.

RESULTS:

The results showed that our model outperformed all other state-of-the-art models by achieving 93% and 92% accuracy for positive and negative classification respectively. Hence, such methods of sentimental analysis are very helpful in devising new policies for the control and prevention of epidemics like COVID-19.

Umair Areeba et al., Sentimental Analysis of Arabic Tweets related to COVID-19 using AraBERT model, CICN 2022 **(Accepted)** (**SCOPUS indexed**).

Research activity # 4

PROBLEM: Applications of Majority Judgement for Winner Selection in Eurovision Song Contest.

The existence of big data, social media interactions, and digital globalization has changed the way people make decisions either in their life or those of collective importance.

In the last few decades, expert rating was used to select the winner in the contest or competition, that was, later, merged with crowd voting.

In Eurovision song context, the voting system was modified several times. In starting, the national juries used to select the winners. Later, in 1997, televoting was introduced. Then, from 2009, a hybrid system was introduced containing the popular vote and a jury vote to select the winner. In 2016, it was again revolutionized where the votes of jury and televoting were arranged separately.

OBJECTIVES:

Nevertheless, we believe that applying majority judgement in Eurovision song context can result in more fair voting.

Research activity # 4

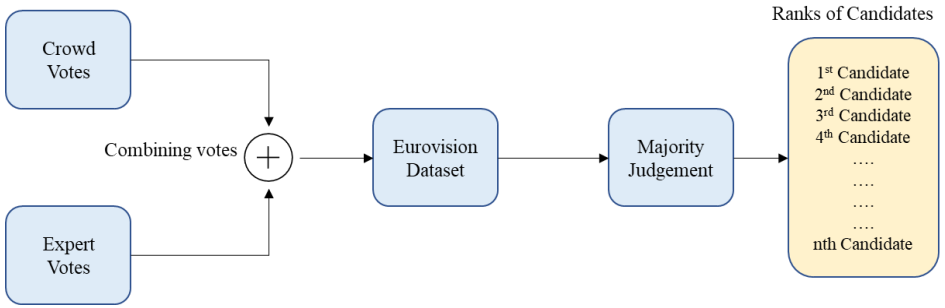
METHODOLOGY:

The overall methodology used in this research us given in figure. We first of all get the jury votes and expert votes and combine them manually.

A very careful approach was kept while adding the votes manually. Then, majority judgement was applied over the newly created dataset of Eurovision Context.

The majority judgement provided the ranks of all the participating countries.

The majority judgement is a new method of election where the jury or voters only judge the candidate instead of ranking them.



Umair, Areeba, et al. "Applications of Majority Judgement for Winner Selection in Eurovision Song Contest." Proceedings of the 26th International Database Engineered Applications Symposium. 2022 **(Published)**.

RESULTS:

The results of the experiments shows that Italy stands first, Croatia stood second and Australia got third position.

Products

	Journal Papers
[P1]	Umair, Areeba, and Elio Masciari. "Sentimental and spatial analysis of COVID-19 vaccines tweets." Journal of Intelligent Information Systems (2022): 1-21 (**IF=2.504, SCOPUS** and **ISI Web of Science indexed**) **(Published)**.
[P2]	Umair, A., Masciari, E. Habib Ullah, M. Vaccine Sentimental Analysis using BERT+NBSVM and Geo-Spatial Approaches, Journal of supercomputing (2022). **IF=2.557, SCOPUS** and **ISI Web of Science indexed**) (**Under-review**)
	Conference Papers
[P3]	Umair, Areeba, and Elio Masciari. "Artificial intelligence-based analysis of positive and negative tweets towards covid-19 vaccines." 2021 IEEE international conference on bioinformatics and biomedicine (BIBM). IEEE, 2021 **(Published)**.
[P4]	Umair, Areeba, and Elio Masciari. "A Survey of Sentimental Analysis Methods on COVID-19 Research." SEBD (2022) **(Published) (SCOPUS indexed)**.

Products

[P5]	Umair, Areeba, et al. "Applications of Majority Judgement for Winner Selection in Eurovision Song Contest." Proceedings of the 26th International Database Engineered Applications Symposium. 2022 **(Published)**.
[P6]	Areeba Umair and Elio Masciari, Giusi Madeo and Muhammad Habib Ullah, Sentimental Analysis of COVID-19 Vaccine Tweets using BERT+NBSVM, NFMCP 2022. (**Accepted**)
[P7]	Umair, Areeba, and Elio Masciari. "Using High Performance Approaches to Covid-19 Vaccines Sentiment Analysis." 2022 30th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP). IEEE, 2022 **(Published) SCOPUS** and **ISI Web of Science indexed)**.
[P8]	Umair Areeba et al., Sentimental Analysis of Arabic Tweets related to COVID-19 using AraBERT model, CICN 2022 **(Accepted)** (**SCOPUS indexed**).
[P9]	Umair, A., & Masciari, E. (2022). Human sentiments monitoring during COVID-19 using AI-based modeling. Procedia Computer Science, 203, 753-758 **(Published) (SCOPUS indexed)**.

Products

	Book Chapter
[P10]	Book chapter “Analyzing impact of COVID-19 on sentiments of people and its spatial trends using social media data”

Summary of study activities

Ad hoc PhD courses / schools

- Software Defined Radio Applications for Radar and Localization Systems
- Ultra High Field Magnetic Resonance Imaging
- Scuola Nazionale per Dottorandi “F. Gasparini”. XXIV Stage, Napoli

Courses attended borrowed from MSc curricula

- Data Visualization

Conferences / events attended

- 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), December 9, 2021, USA (Presented my paper online)
- 30th Euromicro International Conference on Parallel, Distributed and Network-Based Processing in Valladolid, Spain, March 9th - 11th, 2022 (Presented my paper Online)
- 30th Symposium on Advanced Database System - Tirrenia (Pisa), Italy - 19-22 June 2022 (Presented my paper in-person)
- ICIT 2022 International Conference on IT and Industrial Technologies, October 03-04, 2022

Summary of study activities

	Courses	Seminars	Research	Tutorship	Total
Bimonth 1	3	2.4	10		15.7
Bimonth 2	13	2.3	10		25.3
Bimonth 3	0	2.3	10		12.3
Bimonth 4	0	0.6	5		5.6
Bimonth 5	0	0	5		5
Bimonth 6	0	0	5		5
Total	16	7	45		68
Expected	10-20	5-10	30 - 45	0 – 4.8	

Next Year

- **Research activities:** Sentimental Analysis on the news articles related to Napoli Soccer.
- **International Conference**: In my third year of Ph.D., I have a plan to attend Computational Intelligence and Communication Networks (CICN 2022) conference, to be hosted by Prince Mohammad Bin Fahd University (PMU), Kingdom of Saudi Arabia (KSA).
- **Draft topic or title of the thesis**: Advanced artificial intelligence methods for sentimental analysis.

Thank you for the attention