PhD Student: Areeba Umair

Cycle: XXXVI

Training and Research Activities Report

Academic year: 2021-22 - PhD Year: Second

Tutor: prof. Elio Masciari

tutor signature

Date: October 30, 2022

1. Information:

- **PhD student: AREEBA UMAIR**   **PhD Cycle: XXXVI**
- **DR number: DR995146**
- **Date of birth: 07/01/1996**
- **Master Science degree: MSCS**   **University: NUCES, Pakistan**
- **Scholarship type: UNINA**
- **Tutor: Prof. Elio Masciari**
- **Co-tutor:**

2. Study and training activities:

Activity	Type[1]	Hours	Credits	Dates	Organizer	Certificate[2]
Data Visualization	Master's Course		6	28/1/2022	Dr. ROBERTO PIETRANTUONO	Y
Software Defined Radio Applications for Radar and Localization Systems	Ad hoc course		3	22-23-24-25-26-29-30 / 11/2021	Proff. Antonio De Maio, Augusto Aubry, Dr. Vincenzo Carotenuto	Y
Ultra High Field Magnetic Resonance Imaging	Ad hoc course		3	17-18-20-25-26-31 / 01/2022	Dr. Arturo Brunetti, Dr. Christopher Collins, Dr. Riccardo Lattanzi, Dr. Rita Massa, Dr. Giuseppe Ruello	Y
Scuola Nazionale per Dottorandi "F. Gasparini". XXIV Stage, Napoli	Ph.D. school		4	24-28 January 2022	Oriano Bottauscio, Stefano Squartini	Y
Cyber security in Akka Technologies	Seminar	2	0.4	03/11/2021	Dr. Luigi Villa- Sara Belluccini - Matteo Pracchia	Y
Single cell omics leverage Machine Learning to dissect tumor	Seminar	2.5	0.5	12/2/2021	Dr. Raoul J.P. Bonnal	Y

microenvironment and cancer immune editing						
Possible Quantum Machine Learning Approaches in HEP	Seminar	2	0.4	11/12/2021	Dr. Michele Grossi	Y
Connecting the dots: Investigating an APT campaign using Splunk	Seminar	2	0.4	26/11/2021	Dr. Antonio Forzieri	Y
Complexity and the City: transitioning towards the smart cities of the future	Seminar	1.5	0.3	11/23/2021	Dr. Mary Avery, Luigi Amodio and Luis Bettencourt	Y
Threat Hunting Essentials	Seminar	2	0.4	12/3/2021	Group-IB	Y
The learning landscape in deep neural networks and its exploitation by learning algorithms	Seminar	1	0.2	1/21/2022	Prof. Riccardo Zecchina	Y
Torsional instability in suspension bridges: a new mathematical explanation of the Tacoma Narrows Bridge collapse	Seminar	1	0.2	1/14/2022	Prof. Filippo Gazzola	Y
The Spatial structure of Bi-photon States	Seminar	1	0.2	1/11/2022	Dr. Alessio D'Errico	Y
Computational analysis of cancer genomes	Seminar	1.5	0.3	16/2/2022	Dr. Nuria Lopez-Bigas	Y
Artificial intelligence and autonomous weapon systems	Seminar	2	0.4	1/19/2022	Fosca Giannotti, Guglielmo Tamburrini	Y
RAILS MID-TERM WORKSHOP	Seminar	5	1	25/2/2022	Valeria Vittorini, Lorenzo De Donato,	Y

					Francesco Flammini, Denis Miglianico, Autilia Vitiello, Ruifan Tang	
Explainable Natural Language Inference	Seminar	1.5	0.3	04/13/2022	Dr. Marco Valentino	Y
Bench to Bytes to Bedside: Converting genomic data into Healthcare tools	Seminar	1	0.2	04/3/2022	Serena Nik-Zainal	Y
Ethics and Philosophy of Artificial Intelligence	Seminar	1	0.2	04/11/2022	Mark Coekelberg	N
Can a Text-to-Speech Engine Generate Human Sentiments?	Seminar	1	0.2	28/02/2022	prof. Vijay K. Gurbani	N
Da profughi a rifugiati: i corridoi umanitari in Europa	Seminar	2	0.4	04/13/2022	Francesco Dondolo, Amedeo Arena, Flora Di Donato	Y
Using delays for control	Seminar	1	0.2	04/21/2022	Prof. Emilia Fridman	Y
Ciberconflitti e minacce per la pace e la stabilità internazionale	Seminar	2	0.4	04/05/2022	Simon Pietro Romano, Guglielmo Tamburrini	Y
On using simple optimization techniques for tuning of UAVs	Seminar	2	0.4	04/27/2022	Ph.D., D.Sc. (eng.) Dariusz Horla	Y
Probing and infusing biomedical knowledge for pre-trained language models	Seminar	2	0.4	06/07/2022	Dr. Zaiqiao Meng	Y
Towards AI-Driven Cancer Precision Medicine	Seminar	1	0.2	22/04/2022	Olivier Elemento	Y
Published two conference papers	Research		10			
Published one	Research		10			

journal paper						
Start working on computational social choice. Worked on song completion using majority judgement	Research		10			
One paper accepted at MobiSens4H workshop. One conference paper submission in Ideas2022 conference One Journal paper submission in Journal of Supercomputing	Research		10			
One conference paper Accepted in Ideas2022 conference One conference paper Accepted in NFMCP2022 conference One Journal paper under-review in Journal of Supercomputing. Submission of one book chapter.	Research		10			
One Conference paper Accepted in CICN 2022	Research		10			

1)   Courses, Seminar, Doctoral School, Research, Tutorship
2)   Choose: Y or N

2.1. Study and training activities - credits earned

	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	

3. Research activity:

Research Topic:
My research topic is "Devising Artificial Intelligence Tools for Complex Data". During my second year of Ph.D., I performed four research activities listed below:
1. Sentimental analysis of COVID-19 Vaccines tweets using BERT+NBSVM model.
2. Sentimental Analysis of Arabic Tweets using AraBERT model.
3. Spatial Analysis of Tweets related to COVID-19
4. Applications of Majority Judgement for Winner Selection in Eurovision Song Contest.

3.1. Sentimental analysis of COVID-19 Vaccines tweets using BERT+NBSVM model

Topic: 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.

Methodology: The overall framework of my proposed methodology is given in figure 1 below:
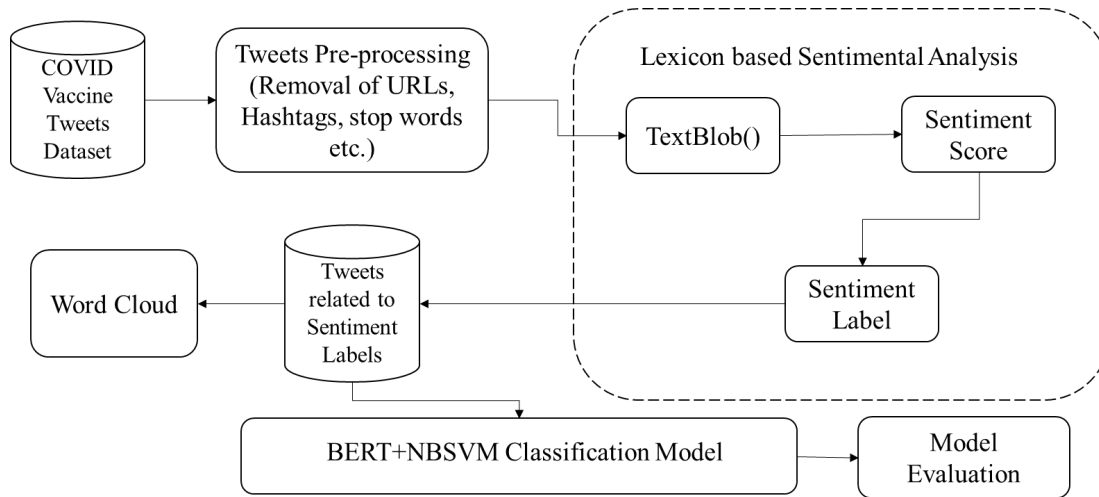

Figure 1 Preposed Framework of Sentimental Analysis of Vaccine Tweets

Phase 1: consists of collection of dataset and pre-processing steps. Tweets data is in unstructured form and it is necessary to make it in structured form therefore, pre-processing is applied for gaining the maximum efficiency. In **phase 2**, social network analysis is performed, which establishes relationships between two or more entities at a time. **Phase 3:** leads to sentimental classification in the phase four by extracting the sentiments and their polarity value. **Phase 4:** used polarity values and performed sentimental classification with the help of BERT+NBSVM model.

Results:

To evaluate and compare the results of our proposed model, we designed the experiments with state-of-the-art algorithms which have been mentioned in previous studies. We performed the comparison with KNN (K-nearest neighbor) algorithm, SVM (Support Vector Machine) algorithm, RF (Random Forest) algorithm, NB (Naive Bayes) algorithm and DT (Decision Tree) algorithm because of their being mostly used in literature.

Figures 2 and 3 show the sub-graphs depicting the classification accuracy, precision, recall and F1 score of our proposed BERT+NBSVM model in comparison with BERT, NBSVM, Decision tree, KNN, random forest and SVM for the classification of positive sentiments. The classification results of positive and negative tweets classification show that our proposed approach outperformed all other state of the art models. The proposed BERT+NBSVM showed the best accuracy for both positive and negative classification.
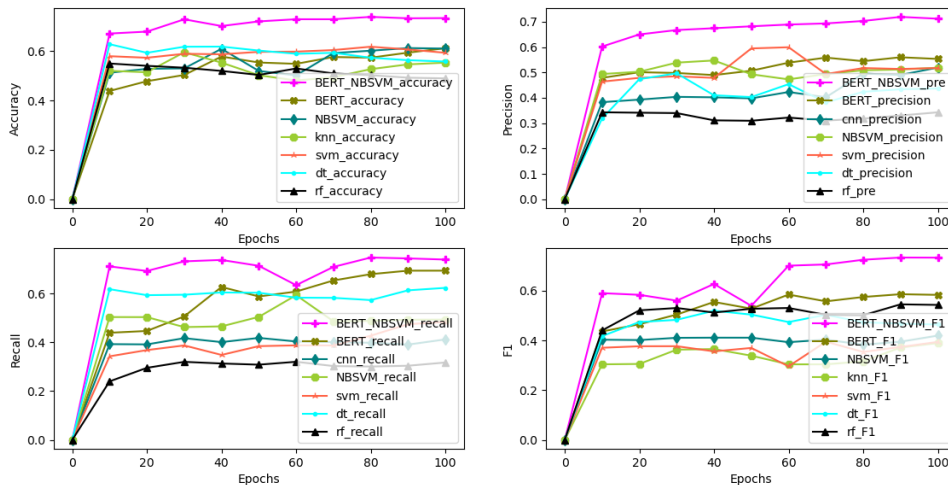
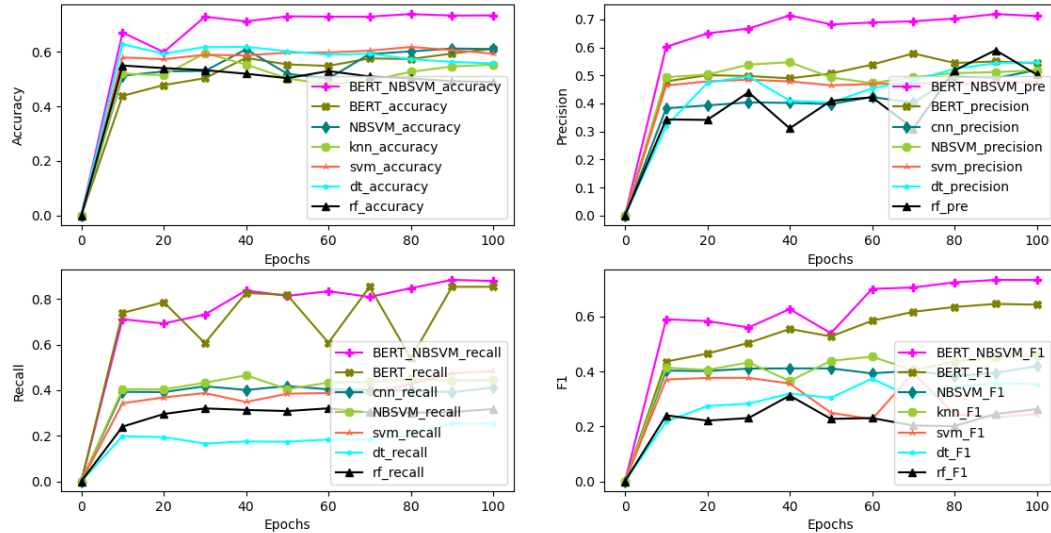*Figure 2 Results of Positive Tweets classification*

Figure 3 Results of Negative tweets classification

3.2. Sentimental Analysis of Arabic Tweets using AraBERT model.

Topic: COVID-19 has caused depressions, anxiety, stress, fear among the people. 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. Now-a-days, social media is a powerful tool and people use it for sharing their feeling and thoughts. Arab people also shared their thoughts on social media during COVID-19. However, the Arabic language is rich in morphology and has large number of dialects. Arabic Sentimental Analysis has attracted many researchers but still the research work on Arabic Language is limited.

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 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.

Training and Research Activities Report
PhD in Information Technology and Electrical Engineering
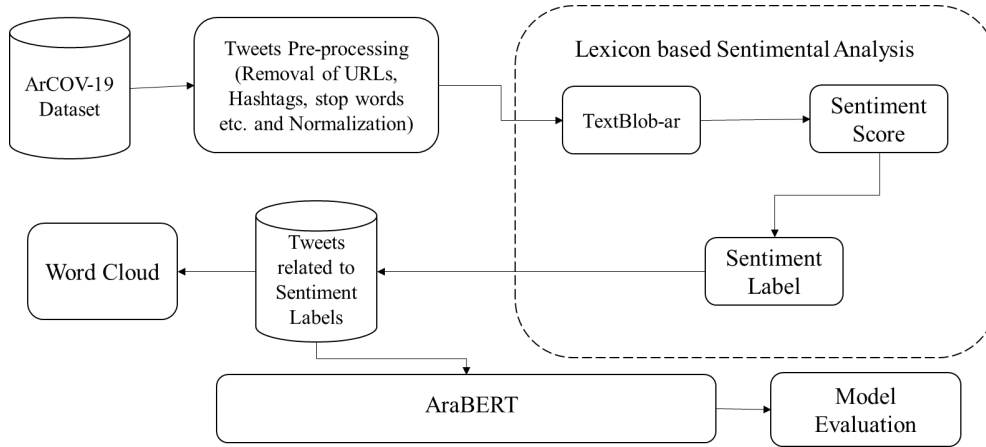**Cycle: XXXVI**   **Author: AREEBA UMAIR**

Figure 4 Sentimental Analysis of Arabic Tweets using AraBERT Model.
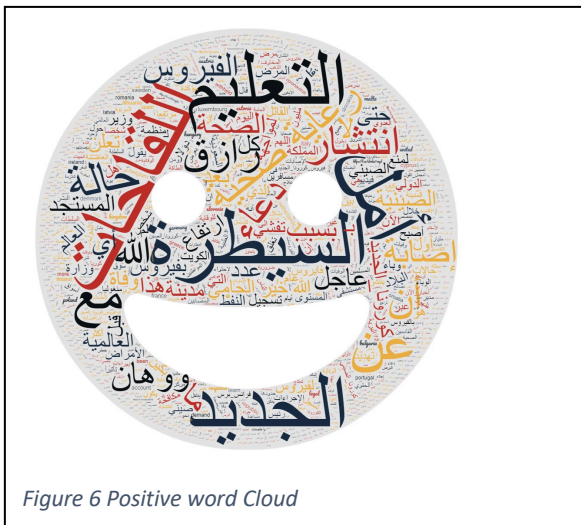
Results:

Word Cloud:
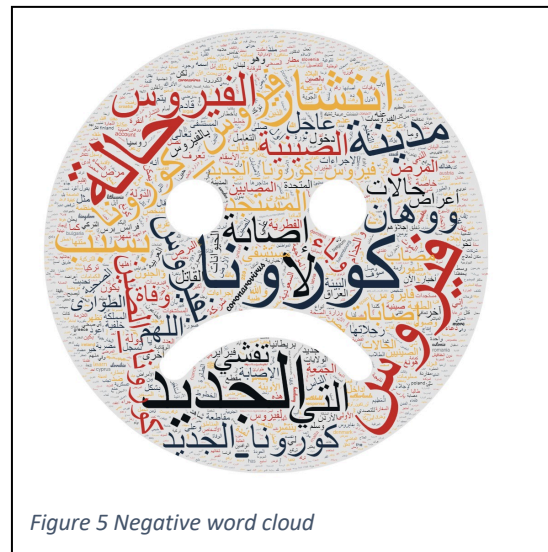

Figure 6 Positive word Cloud

Figure 5 Negative word cloud

The Arabic words like {السيطرة} control, {اللقاحات} vaccines, {رعاية صحية} healthcare, {التعليم} education, {أمل} hopeful, {دعاء} prayer, {الله خير الحامي} God is the best protector, {رازق} sustenance provider, {الجديد} new {كورونا_الجديد} {انتشار} spread, show the positive behavior of people during COVID-19. While the words like new coronavirus, {حالة} condition, {بفيروس} virus, {إصابة} infection shows the negative thoughts of the Arabic people during COVID-19.

Classification Results:

The results of Arabic sentiment classification can be seen in Figure 7 and Figure 8. We can see from the figures that our proposed AraBERT model outperformed all other state-of-the-art models
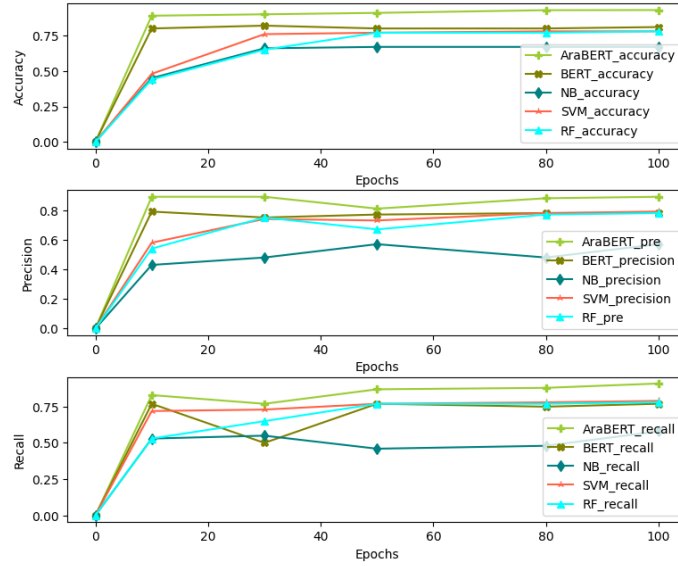
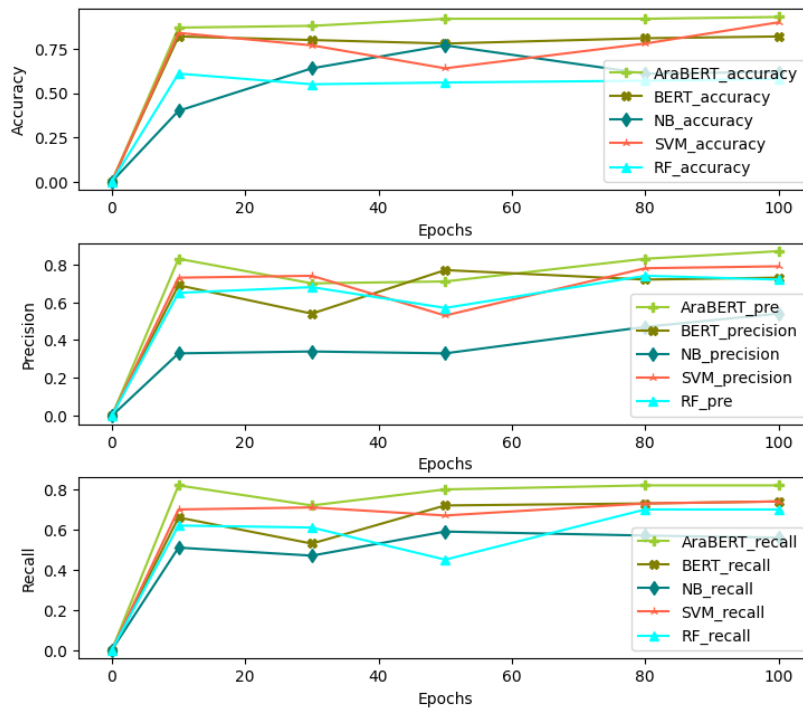*Figure 7 Positive tweets classification*

Figure 8 Negative tweets classification

Training and Research Activities Report
PhD in Information Technology and Electrical Engineering

Cycle: XXXVI   **Author: AREEBA UMAIR**
_____

3.3. Spatial Analysis of Tweets related to COVID-19

Mapping is considered a critical tool in establishing the relationship between infectious disease control and their modern environment. Today, GIS based modeling is performed using digital as well as electronic big data.
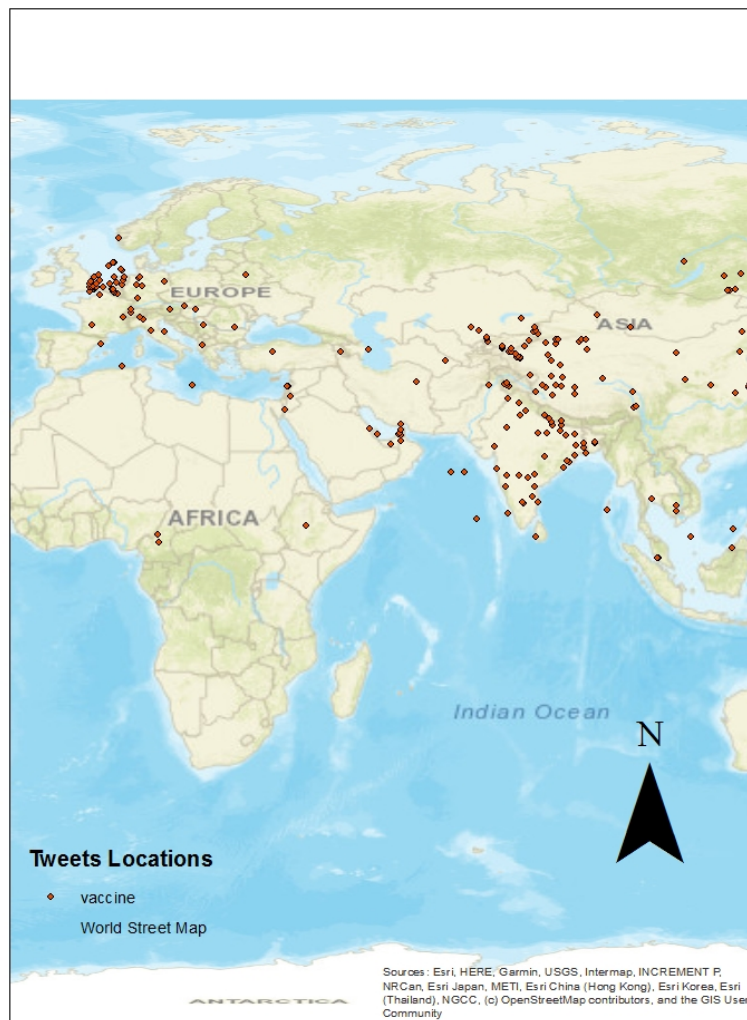
Geo-Coding and Visualization of Data:

Figure 9 Visualization of data on Map

Analyzing the patterns using Average Nearest Neighbor (ANN):

There exist patterns in data if the data falsified the null hypothesis (i.e. the features are in complete state of randomness). On the other hand, it exhibits the relationship of either clustered or dispersed. Clustering relationship is important because it shows the high geographical associativity between the features
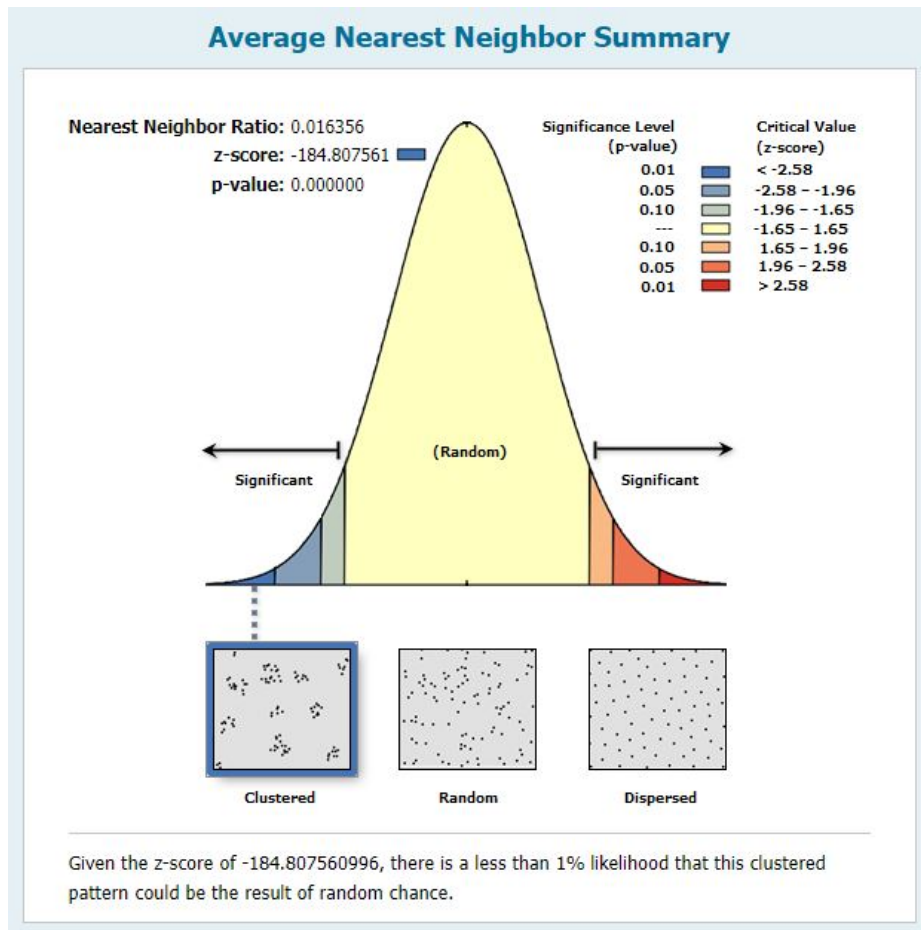

Figure 10 Spatial relationship of vaccine data using ANN

Hotspot Analysis

Hotspot shows the geographical areas where the vaccine sentiment polarity is high in rate while cold spots show the areas with less vaccine sentiment polarity.
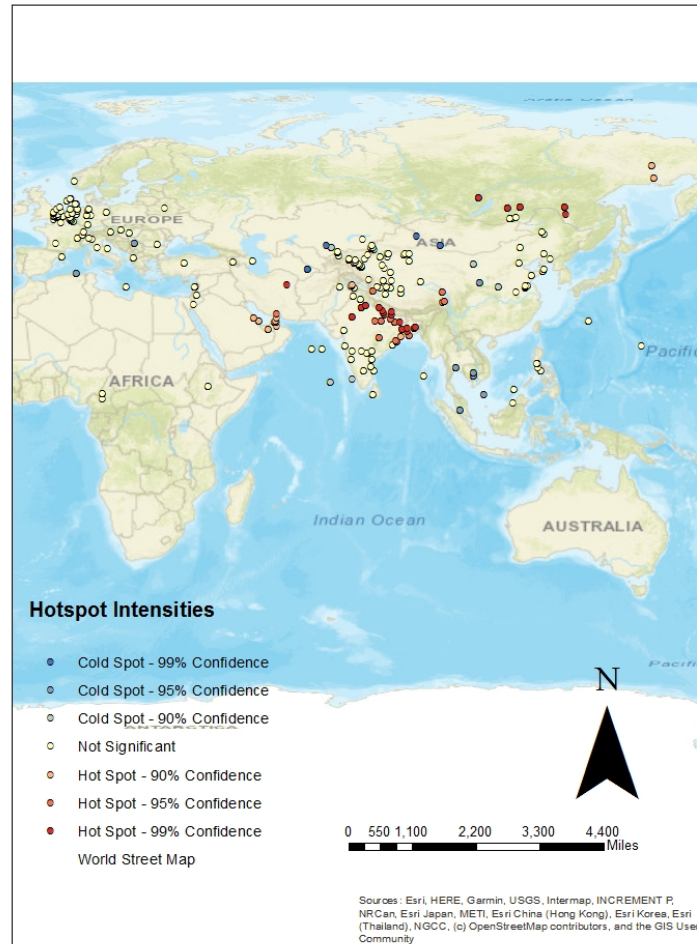
Figure 11 Hotspot analysis of COVID-Vaccine tweets

Analysis Using Kernel Density Estimation

Kernel density can be envisioned as putting the blob of ice-cream on the top of each feature point and then the density function is the measuring the height of the accumulated blobs
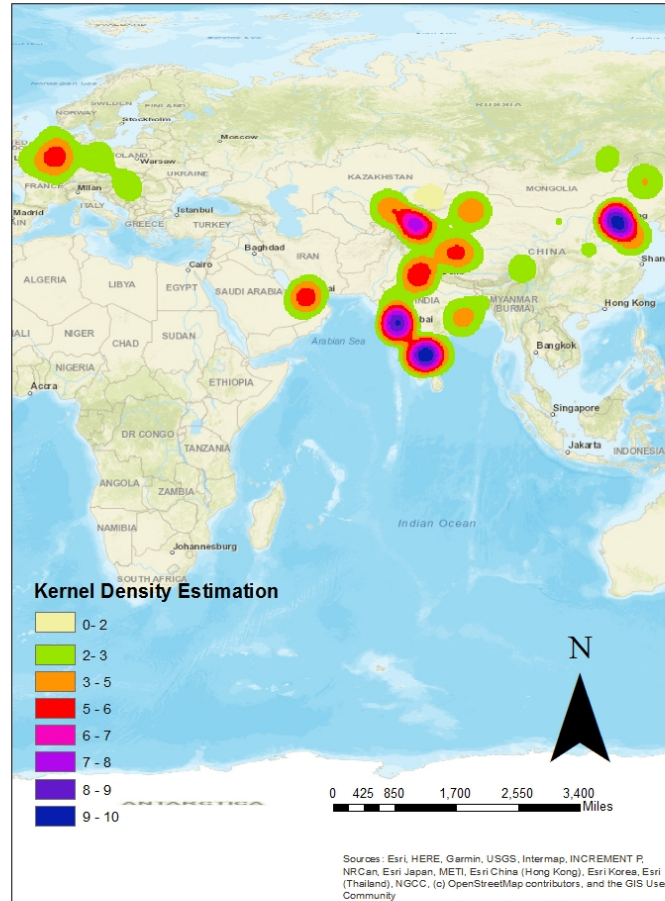
Figure 12 Kernel Density Estimation of COVID-Vaccine tweets

Buffering:

One hurdle for vaccination may be the mobility. People are unable to reach to the vaccination center or may be they are allocating vaccination center too far. Geo-spatial techniques like buffering can be efficiently used to address allocation problems.
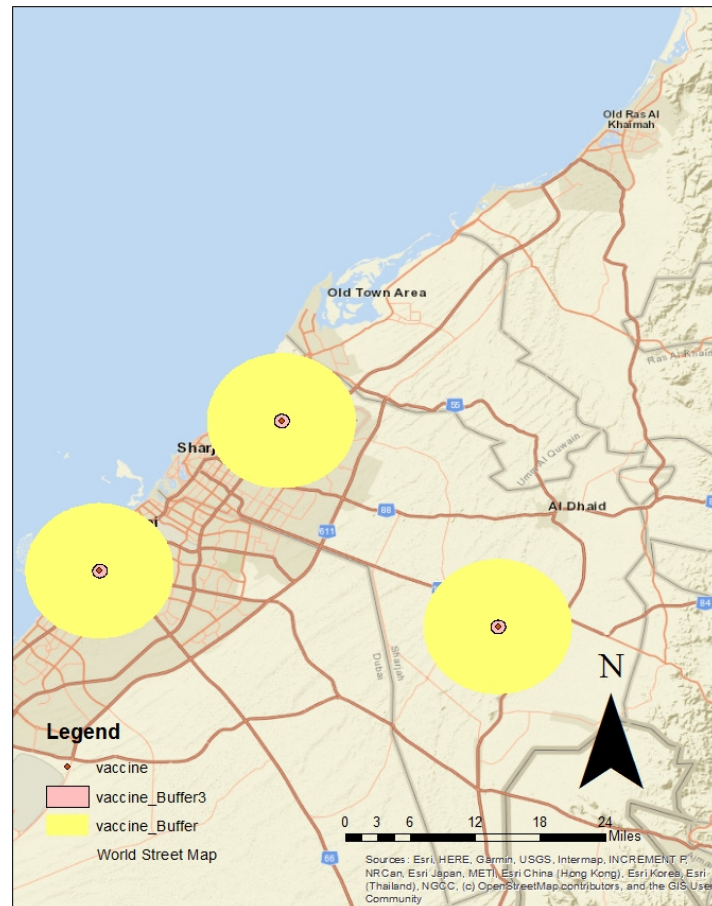
Figure 13 Buffering

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

Topic: 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. Computational Social Choice (COMSOC), as an emerged field, has tried to join various social fields (social choice theory) and technical fields (computer science, mathematics, economics and logic). 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. However, the results of voting based on aggregation of crowd opinion was not considered satisfied.

Methodology:

The majority judgement is a new method of election. It is the consequence of a new theory of social choice where voters judge candidates instead of ranking them. In this research, we used Eurovision song contest data of 2021 final round. Eurovision song contest is held annually, in which almost 40 countries participate. We applied majority judgement on the Eurovision song contest data and found that Italy got highest position, followed by Croatia and Australia acquiring second and third positions respectively in competition of 2021.
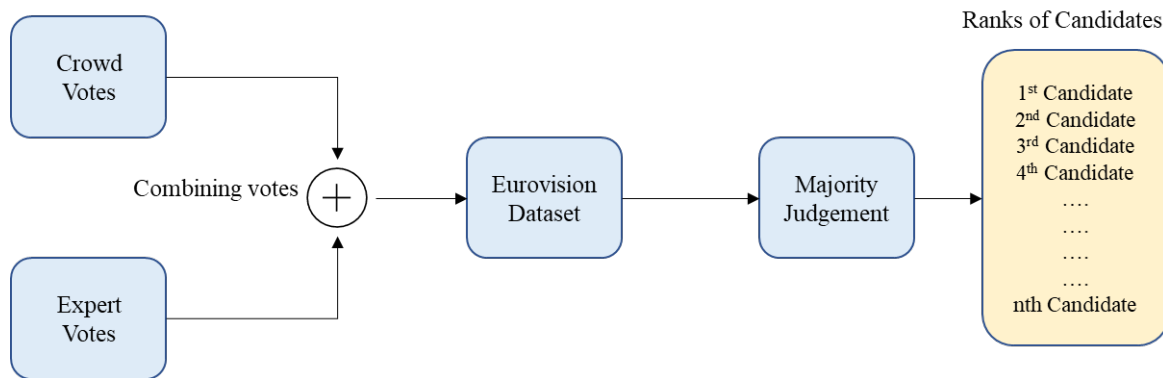

Figure 14 Majority Judgement on Eurovision song contest

Results:

Figure 2 shows the rank of each participating country. It is clear from the figure that the winner country stands at the highest point. Considering the results, and our dataset, we concluded that Italy is the country with highest rank, followed by Croatia and Australia with second and third position respectively.
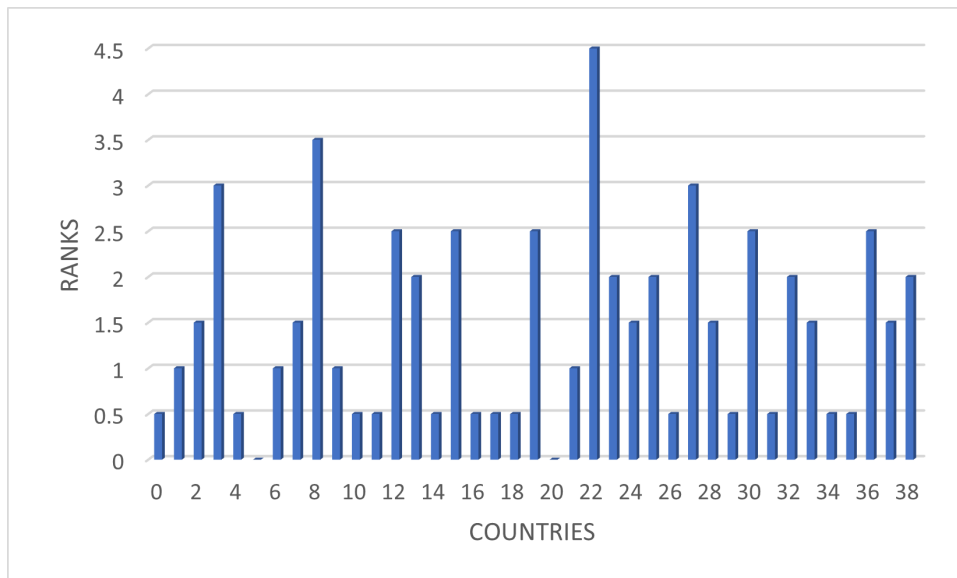

Figure 15 Results of Eurovision song contest

4. Research products:

1. 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)**.
2. 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)**.
3. 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)**.
4. Umair, Areeba, and Elio Masciari. "A Survey of Sentimental Analysis Methods on COVID-19 Research." SEBD (2022) **(Published) (SCOPUS indexed)**.
5. Umair, A., & Masciari, E. (2022). Human sentiments monitoring during COVID-19 using AI-based modeling. Procedia Computer Science, 203, 753-758 **(Published) (SCOPUS indexed)**.
6. 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)**.
7. Areeba Umair and Elio Masciari, Giusi Madeo and Muhammad Habib Ullah, Sentimental Analysis of COVID-19 Vaccine Tweets using BERT+NBSVM, NFMCP 2022. (**Accepted**)
8. 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**)
9. Umair Areeba et al., Sentimental Analysis of Arabic Tweets related to COVID-19 using AraBERT model, CICN 2022 **(Accepted**) (**SCOPUS indexed**).
10. Book chapter "Analyzing impact of COVID-19 on sentiments of people and its spatial trends using social media data"

5. Conferences and seminars attended

1. 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), December 9, 2021, USA (Presented my paper online)
2. 30th Euromicro International Conference on Parallel, Distributed and Network-Based Processing in Valladolid, Spain, March 9th - 11th, 2022 (Presented my paper Online)
3. 30th Symposium on Advanced Database System - Tirrenia (Pisa), Italy - 19-22 June 2022 (Presented my paper in-person)
4. ICIT 2022 International Conference on IT and Industrial Technologies, October 03-04, 2022

6. Periods abroad and/or in international research institutions

During my second year, I did not spend my time abroad.

7. Tutorship

During my second year, I did not perform tutorship.

8. Plan for year three

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.