

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

**DOTTORATO DI RICERCA / PhD PROGRAM IN
INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING**

Activities and Publications Report

PhD Student: **Areeba Umair**

Student ID: DR995146

PhD Cycle: XXXVI

PhD Cycle Chairman: Prof. Stefano Russo

PhD program student's start date: 01/11/2020

PhD program student's end date: 31/10/2023

Supervisor: Prof. Elio Masciari

e-mail: elio.masciari@unina.it



PhD scholarship funding entity:

Università Federico II.

General information

Areeba Umair received in year 2020 the Master Science degree in Computer Science from the National University of Computer and Emerging Sciences, Islamabad, Pakistan. She attended a curriculum in Information Technology and Electrical Engineering within the PhD program in Information Technology and Electrical Engineering. She received a grant from Università Federico II.

Study activities

Attended Courses

Year	Course Title	Type	Credits	Lecturer	Organization
1	Digital Forensics; methods, practices and tools	Ad hoc course	3	Dr. Giovanni Cozzolino	ITEE
1	Statistical data analysis for science and engineering research	Ad hoc course	4	Prof. Roberto Pietrantuono	ITEE
1	Hardware and Software Architectures for Big Data – Mod. B	MSc course	6	Prof. Vincenzo Moscato	ITEE
1	Big Data Analytics and Business Intelligence	MSc course	6	Prof. Giancarlo Sperli'	ITEE
1	Safety Training Course	course	0	Dott.ssa Liliana Lista	UNINA
2	Software Defined Radio Applications for Radar and Localization Systems	Ad hoc course	3	Proff. Antonio De Maio, Augusto Aubry, Dr. Vincenzo Carotenuto	UNINA
2	Ultra High Field Magnetic Resonance Imaging	Ad hoc course	3	Dr. Arturo Brunetti, Dr. Christopher Collins, Dr. Riccardo Lattanzi, Dr. Rita Massa, Dr. Giuseppe Ruello	UNINA
2	Data Visualization	MSc course	6	Dr. ROBERTO PIETRANTUONO	UNINA
3	Corso di Italiano livello A1	course	40 hours	Prof. Ana Baldan	CLA, UNINA

Attended PhD Schools

Year	School title	Location	Credits	Dates	Organization
1 st	2021 Spring School on	Napoli,	2	4-5 May	Department of Pharmacy

Activities and Publications – Final Report

UNINA PhD in Information Technology and Electrical Engineering – XXXVI Cycle

PhD candidate: Areeba Umair

	Transferable Skills	Italy		2021	University of Naples Federico II
1 st	AIRO PhD School 2021 and 5th AIRO Young Workshop	Online	3.6	8-12 February 2021	UNINA
2 nd	Scuola Nazionale per Dottorandi “F. Gasparini”. XXIV Stage, Napoli	Online	4	24-28 January 2022	University of Napoli Federico II

Attended Seminars

Year	Seminar Title	Credits	Lecturer	Lecturer affiliation	Organization
1 st	Robot Manipulation and Control	0.5	Prof. Paolo Dario, Scuola Sant’Anna Pisa	Scuola Sant’Anna Pisa	ITEE
1 st	Patent Searching Best Practices with IEEE Xplore	0.2	Rachel Berrington	IEEE Client Services Manager Europe	IEEE
1 st	How to get published with the IEEE	0.3	Dr.ssa Eszter Lukacs	IEEE Client Services Manager Europe	IEEE
1 st	Advances in Machine Learning for Modelling and Understanding in Earth Sciences	0.3	Prof. Antonio Iodice	DIETI	ITEE
1 st	NLP ED AI NEL MONDO ENTERPRISE	0.4	Prof. Marco PASSAROTTI	Universita cattolica del sacro cuore	ITEE
1 st	AI: Artificial Intelligence for notary's sector - a case study	0.2	Salvatore Palange		ITEE
1 st	Machine Learning: Causality loss in translation	0.3	Edwin A. Valentijn	University of Groningen	ITEE
1 st	Approaches to graph machine learning	0.3	Miroslav Cepek	Czech technical university	ITEE
1 st	IEEE Authorship and Open Access Symposium: Best Practices to Get Published to Increase the Exposure and	0.3	Rachel Berrington	IEEE	IEEE

Activities and Publications – Final Report

UNINA PhD in Information Technology and Electrical Engineering – XXXVI Cycle

PhD candidate: Areeba Umair

	Impact of Your Research				
1 st	Wireless Intelligence: From Reconfigurable Surfaces to Edge/Cloud Communications	0.8	Prof. M. Martalò, Prof. R. Raheli	University of Cagliari	ITEE
1 st	Logic-based Learning of Answer Set Programs	0.2	Prof. Mark Law	University of Florida	ITEE
1 st	Why Do We Cooperate? Understanding and Modelling Societies using Reinforcement Learning	0.3	Prof. Mirco Musolesi	University college london	ITEE
1 st	Visual Interaction And Communication in Data Science	0.4	Prof. Marco Quartulli		ITEE
1 st	Big Data and Computational Linguistics	0.4	Prof. Francesco Cutugno	UNINA	ITEE
1 st	SENSORIA Health	0.2	Prof. Stefano Rossotti		ITEE
1 st	The Science of Science	0.3	Prof. Giuseppe Longo	UNINA	ITEE
1 st	Distributional Semantics Methods: How Linguistic features can improve the semantic representation	0.3	Prof. Alessandro Maisto	University of salerno	ITEE
1 st	Risk assessment in real life: experiences from the railway domain	0.3	Prof. Valeria Vittorini	DIETI	ITEE
1 st	Synchronization in complex networks, hypergraphs and simplicial complexes	0.2	Dr. Marco Coraggio, Dr. Micol Benetti	UNINA	ITEE
1 st	Optimized Graph Representations for Right-Wing Reddit Community Detection Using Graph Neural Networks	0.2	Prof.ssa Silvia Rossi.	DIETI	ITEE
1 st	Introduction To legged Robots and Examples of lit's Dynamic Legged Systems la	0.4	Dr. Fabio Ruggiero	DIETI	ITEE
1 st	Introduction To Underwater Robotics	0.4	Dr. Fabio Ruggiero	DIETI	ITEE

Activities and Publications – Final Report

UNINA PhD in Information Technology and Electrical Engineering – XXXVI Cycle

PhD candidate: Areeba Umair

1 st	Ethics of Quantification	0.4	Andrea Saltelli		ITEE
1 st	Sadas Engine, an innovative DBMS for the DATA WAREHOUSE, great performance in the VLDB environment.	0.3	Luca De Rosa	UNINA	ITEE
1 st	5G: Esposizione ai Campi Elettromagnetici e Metodologie di Misura	0.8	Prof. Nicola Pasquino	DIETI	ITEE
2 nd	Cyber security in Akka Technologies	0.4	Dr. Luigi Villa-Sara Belluccini - Matteo Pracchia	DIETI	ITEE
2 nd	Single cell omics leverage Machine Learning to dissect tumor microenvironment and cancer immune editing	0.5	Dr. Raoul J.P. Bonnal	UNINA	ITEE
2 nd	Possible Quantum Machine Learning Approaches in HEP	0.4	Dr. Michele Grossi	University of Pavia	ITEE
2 nd	Connecting the dots: Investigating an APT campaign using Splunk	0.4	Dr. Antonio Forzieri	Politecnico di Milano	ITEE
2 nd	Complexity and the City: transitioning towards the smart cities of the future	0.3	Dr. Mary Avery, Luigi Amodio and Luis Bettencourt	UNINA	ITEE
2 nd	Threat Hunting Essentials	0.4	Group-IB		ITEE
2 nd	The learning landscape in deep neural networks and its exploitation by learning algorithms	0.2	Prof. Riccardo Zecchina	Bocconi University	ITEE
2 nd	Torsional instability in suspension bridges: a new mathematical explanation of the Tacoma Narrows Bridge collapse	0.2	Prof. Filippo Gazzola	Politecnico di Milano	ITEE
2 nd	The Spatial structure of Bi-photon States	0.2	Dr. Alessio D'Errico	University of Ottawa	ITEE
2 nd	Computational analysis of cancer genomes	0.3	Dr. Nuria Lopez-Bigas	IRB Barcelona	ITEE
2 nd	Artificial intelligence and	0.4	Fosca Giannotti,	Scuola di Normale	ITEE

Activities and Publications – Final Report

UNINA PhD in Information Technology and Electrical Engineering – XXXVI Cycle

PhD candidate: Areeba Umair

	autonomous weapon systems		Guglielmo Tamburrini	Superiore di Pisa	
2 nd	RAILS MID-TERM WORKSHOP	1	Valeria Vittorini, Lorenzo De Donato, Francesco Flammini, Denis Miglianico, Autilia Vitiello, Ruifan Tang	DIETI	ITEE
2 nd	Explainable Natural Language Inference	0.3	Dr. Marco Valentino		ITEE
2 nd	Bench to Bytes to Bedside: Converting genomic data into Healthcare tools	0.2	Serena Nik-Zainal	University of Cambridge	ITEE
2 nd	Ethics and Philosophy of Artificial Intelligence	0.2	Mark Coekelberg		ITEE
2 nd	Can a Text-to-Speech Engine Generate Human Sentiments?	0.2	prof. Vijay K. Gurbani	Illinois institute of Technology	ITEE
2 nd	Da profughi a rifugiati: i corridoi umanitari in Europa	0.4	Francesco Dondolo, Amedeo Arena, Flora Di Donato	UNINA	ITEE
2 nd	Using delays for control	0.2	Prof. Emilia Fridman	TEL AVIV university	ITEE
2 nd	Ciberconflitti e minacce per la pace e la stabilità internazionale	0.4	Simon Pietro Romano, Guglielmo Tamburrini	DIETI	ITEE
2 nd	On using simple optimization techniques for tuning of UAVs	0.4	Ph.D., D.Sc. (eng.) Dariusz Horla	Poznan University	ITEE
2 nd	Probing and infusing biomedical knowledge for pre-trained language models	0.4	Dr. Zaiqiao Meng	University of Glasgow	ITEE
2 nd	Towards AI-Driven Cancer Precision Medicine	0.2	Olivier Elemento	Cornell University	ITEE
3 rd	From Cyber Situational Awareness to Adaptive Cyber Defense: Leveling the Cyber Playing Field	0.4	Prof. Massimiliano Albanese	George Mason University	ITEE

Research activities

Areeba Umair participated in the research related to “Devising artificial intelligence tools for complex data”. During the three years, the focus of research was on analysing sentiments of people regarding COVID-19.

1. The initial phase of this project consisted of conducting an extensive examination of the scientific literature within the relevant field. This thorough review aimed to augment the existing knowledge base and shed light on the current state of affairs. Through this preliminary study, we identified potential deficiencies or areas where further investigation is warranted. These insights were instrumental in formulating research inquiries, which, in turn, served as the foundation for crafting a comprehensive research proposal.
2. The initial phase of the experimental activity focuses on the design and development of an autoencoder tailored for COVID-19 sentiment analysis using complex social media data. In this context, a novel hybrid algorithm has been proposed for this purpose. The results obtained from this research effort demonstrate that these novel algorithms are capable of achieving superior performance compared to existing state-of-the-art methods, despite their simplicity. This suggests that the hybrid algorithm, which combines elements from different approaches, offers an effective and efficient solution for COVID-19 sentiment analysis in the context of social media data. By surpassing more complex algorithms in terms of performance, it not only showcases its efficacy but also potentially reduces computational complexity and resource requirements, making it a valuable contribution to the field. Further details and insights from the experiment can be discussed to provide a comprehensive understanding of the research findings.
3. The second phase of the experimental activity encompasses the creation of a recommendation system that incorporates sentiment analysis. Within this phase, significant enhancements were made to the BERT model, including the addition of a convolutional layer. These improvements resulted in notably improved results, particularly in the context of seven distinct sentiment classes. The augmentation of the BERT model with a convolutional layer likely contributed to its ability to capture more nuanced sentiment patterns within the text data. Convolutional layers are proficient at detecting local features and patterns, which can be invaluable for sentiment analysis, as it involves understanding subtle textual cues. Furthermore, the reference to "review categorization" suggests that after sentiment analysis, the reviews were further organized or categorized based on their sentiment. This categorization step could be beneficial for generating more targeted and personalized recommendations.

Overall, this phase of the experiment demonstrates the effectiveness of combining sentiment analysis with improved BERT models, convolutional layers, and review categorization to

develop a robust recommendation system capable of delivering more relevant and tailored recommendations to users across various sentiment categories.

Tutoring and supplementary teaching activities

Credits summary

PhD Year	Courses	Seminars	Research	Tutoring / Supplementary Teaching
1 st	20	5	35	0
2 nd	16	7	45	0
3 rd	02	0.4	60	0

Motivare eventuali difformità significative nel percorso di formazione rispetto a quanto l'organizzazione del dottorato prevede come standard e che è riassunto dalla tabella seguente – Cancellare se non ci sono difformità significative.

Iter formativo	corsi / scuole	seminari	attività ricerca	tutorato / did. int.
1 anno	min 20 - max 40	min 5 - max 10	min 10 - max 35	min 0 – max 1.6
2 anno	min 10 - max 20	min 5 - max 10	min 30 - max 45	min 0 – max 1.6
3 anno	min 0 - max 10	min 0 - max 10	min 40 - max 60	min 0 – max 1.6
TOTALE	min 30 – max 70	min 10 – max 30	min 80 – max 140	min 0 – max 4.8

Research periods in institutions abroad and/or in companies

PhD Year	Institution / Company	Hosting tutor	Period	Activities
1 st	NULL			
2	NULL			
3	NULL			

PhD Thesis

In the PhD Thesis, Areeba Umair worked on devising artificial intelligence tools for complex data. Complex data in data science refers to information with intricate structures, such as high-dimensionality, mixed data types, temporal or spatial dependencies, graph/network formats, unstructured content, missing values, nonlinear relationships, hierarchical organization, or dynamic changes over time. Social media data is considered complex due to its diverse formats, high volume, unstructured nature, temporal dynamics, network structure, noise, missing data, sentiment, and contextual challenges. "Covid-19 tweets data" refers to a collection of tweets posted on Twitter, related to the COVID-19 pandemic. This dataset encompasses a range of information, including public reactions, opinions, news, and discussions about various aspects of

the pandemic. Analyzing this data provides insights into public sentiment, trends, misinformation, and also help to recommend the appropriate vaccine to people.

Various techniques have been employed for COVID-19 sentiment analysis including Lexicon-based methods, supervised machine learning models, like SVM and Naive Bayes, Deep learning approaches, such as RNNs and CNNs, transformer-based models, like BERT, Ensemble methods, and aspect-based analysis. In this thesis, we used X app (former Twitter) complex data and proposed BERT+NBSVM for classifying negative and positive tweets regarding COVID-19 vaccines, after applying necessary pre-processing steps. We also proposed sentiment analysis based recommender system for COVID-19 vaccines. For this purpose, proposed an ensemble of random forest with CT-BERT_CONVLayerFusion model, for classifying the tweets into seven different categories of sentiments. Sigma high performance architecture is used for better and efficient computing processing.

Our work also focuses on the usage of geo-spatial approaches to identify the geo-spatial patterns in the vaccination data. We performed buffering to suggest proper vaccination centers based on sentiment analysis. Thus, policymakers can benefit from our methods by analyzing people's concerns and understanding their mindset to improve proper planning to inform people about vaccines, identify misinformation or rumors spreading across the country, and launch ad-hoc campaigns suited to avoid confusion on this important topic.

Publications

Research results appear in 2 (+1 to be submitted) papers published in international journals and 4 contributions to international conferences.

List of scientific publications

International journal papers

Areeba Umair, and Elio Masciari.

"Sentimental and spatial analysis of covid-19 vaccines tweets."

Journal of Intelligent Information Systems

60.1 (2023): 1-21.

Areeba Umair, Elio Masciari, and Muhammad Habib Ullah.

"Vaccine sentiment analysis using BERT+ NBSVM and geo-spatial approaches."

The Journal of Supercomputing

(2023): 1-31.

Umair, Areeba, and Elio Masciari

"Sentiment Analysis using Improved CT-BERT_CONVLayer Fusion Model for COVID-19 Vaccine Recommendation"

Journal, 2023 (to be submitted)

International conference papers

Areeba Umair , Elio Masciari, and Muhammad Habib Habib Ullah.

"Sentimental analysis applications and approaches during covid-19: a survey."

Proceedings of the 25th International Database Engineering and Applications Symposium.

2021.

Areeba Umair, Elio Masciari.

"Sentimental Analysis of COVID-19 Vaccine Tweets Using BERT+ NBSVM."

Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham:

Springer Nature Switzerland,

2022.

Areeba Umair, 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.

Areeba Umair, and Elio Masciari.

"Human sentiments monitoring during COVID-19 using AI-based modeling."

Procedia Computer Science

203 (2022): 753-758

Patents and/or spin offs

Awards and Prizes

Date 20/10/2023

PhD student signature



Supervisor signature

