Big Data Analytics for Text Mining

Teacher(s)
Zafar Saeed

Course Website (optional)
https://classroom.google.com/
Code: *biuuwcc*

Course description (min 150, max 300 words)
The explosion of social media and the computerization of every aspect of social and economic activity resulted in creating large volumes of mostly unstructured data: weblogs, videos, speech recordings, photographs, e-mails, tweets, and similar. In parallel development, computers keep getting ever more powerful and storage ever cheaper. Today, we have the ability to reliably and cheaply store huge volumes of data, efficiently analyze them, and extract business and socially relevant information. The key objective of this course is to familiarize the students with the most important information technologies used in manipulating, storing, and analyzing big data. We will examine the basic tools for statistical analysis using Python, and several AI and Machine Learning algorithms.

The course will introduce the principles of Big Data Analytics and its role. It will develop a thorough understanding of the theoretical concepts of Big Data Analytics in combination with the practical implementation of real-world problems using tools and technologies essentially required to process big data. Students who complete this course successfully are expected to:

- understand the fundamentals of Big Data Analytics and the role it plays
- learn various APIs to process, estimate, and manage data in multiple formats
- demonstrate the ability to translate data into clear, actionable insights
- learn supervised, unsupervised, and graph learning methods in AI and ML
- gain the ability to initiate and design systems that can accept, store, and analyze large volumes of unstructured data
- demonstrate the ability to think critically in making decisions based on data and deep analytics

Course period
Nov-Dec 2024

SSD
INF/01

Course References (optional)
[1] R. Brachman and H. Levesque: Knowledge Representation and Reasoning. Morgan Kaufmann 2014
[2] J. Z. Pan, An Introduction to Knowledge Graphs, Tutorial at the BIAS2021 summer school, 2021.
[3] M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich: A Review of Relational Machine Learning for Knowledge Graphs. Proceedings of the IEEE, Volume 104

Credits and Hours

3 credits, 2 of lectures (15 hours) and one of practice (15 hour), for a total of 30 hours.

Exam Modality

Two alternatives are available to the student to pass this exam (Teacher(s) may choose other modalities):

1) Paper presentation. Students present the content of 2 papers suggested by the teacher. No groups are allowed.
2) Project. Students implement and experimentally validate an algorithm or its variation from a paper suggested/approved by the teacher. Projects can be done in groups of 1-3 students, depending on the algorithm.

Teacher(s) CV

Find below.

Teacher(s) Main Publications

1. *Exploring the impact of SEO-based ranking factors for voice queries through machine learning, Artificial Intelligenve Review.* (IF 12, SJR: Q1, In-print-2024**)**
2. Towards understanding the role of content-based and contextualized features in detecting abuse on Twitter. *Heliyon.* (IF 4, SJR: Q1, online-2024**)**
3. Dynamic Bipartite Network Embedding for incorporating structural and attribute information. World Wide Web. 2023 Sep;26(5):3463-81. (IF 3.7, SJR: Q1, online-2023**)**
4. Energy theft detection in smart grids with genetic algorithm-based feature selection, *Computers, Materials & Continua*, 74(3), 5431–5446. (IF 3.86, SJR: Q2, online-2023**)**
5. Investigating cognitive workload in concurrent speech-based information communication, *International Journal of Human-Computer Studies, Vol: 157-102728.* (IF: 4.866, SJR: Q1, online-2022**)**
6. Enhanced heartbeat graph for emerging event detection on twitter using time series networks. *Expert Systems with Applications*, *136*, 115-132. (IF: 8.665, SJR: Q1, online-2020**)**
7. Evesense: what can you sense from twitter?. In *European Conference on Information Retrieval* (pp. 491-495). Lisbon, Portugal, Springer, Cham. (Core Rank: A, online-2020**)**
8. Event detection in Twitter stream using weighted dynamic heartbeat graph approach [application notes]. *IEEE Computational Intelligence Magazine*, *14*(3), 29-38. (IF: 9.809, SJR: Q1, online-2019**)**
9. What's happening around the world? a survey and framework on event detection techniques on twitter. *Journal of Grid Computing*, *17*(2), 279-312. (IF 4.674, SJR: Q1, online-2019**)**
10. Text stream to temporal network-a dynamic heartbeat graph to detect emerging events on twitter. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 534-545). *Melbourne, Australia*, Springer, Cham. (Core Rank: A**,** online-2018**)**

Zafar Saeed

Education

2020	**Ph.D (Computer Science)**, Quaid-i-Azam University, Islamabad, Pakistan
	Thesis: Detecting Unspecified Events on Twitter
2018	**Ph.D (Visiting Scholar)**, University of Technology Sydney, Australia
	Thesis: Detecting Unspecified Events on Twitter
2010	**M.Phil (Computer Science)**, Quaid-i-Azam University, Islamabad, Pakistan.
	Thesis: Usable Layout for Scientific Documents to Support Online Reading
2007	**BS(Computer Science)**, BUITMS, Quetta, Pakistan.
2001-02	**Diploma in IT** and **Diploma in .Net Technologies**, BIIT, Quetta, Pakistan.

Academic Experience

Assistant Professor (RTD-A), University of Bari

Courses Taught: Combining Knowledge and Reasoning, Big Data Analytics for Text Mining.

Assistant Professor, University of Central Punjab, LHR, Pakistan   (09/2020 – 06/2022)

Courses Taught: Operating Systems, Emerging Trends in Information Retrieval, Big Data Analytics, Natural Language Processing, Artificial Intelligence.

Course Coordinator: Designing course outlines, monitoring teaching activities, managing, and leading the course instructors to uphold the quality of teaching environments. Coordinating three courses: Introduction to Data Science, Big Data Analytics, Data Analysis Techniques.

Research and Collaboration: Organizing faculty's research seminars, coordinating with the external institutes for research and collaboration.

Course Curriculum: Designing new course curriculum for bachelor's and master's programs in Data Science such as Social Network Analysis and Social Media Content Analysis.

Thesis Supervised:

Ushtar Ali, Sentiment analysis on political tweets using graph embedding (MSCS-2022)

Fozia Aslam, Impact of SEO-based ranking factors for voice queries through machine learning (MSCS-2021)

Lecturer, University of Central Punjab, LHR, Pakistan   (10/2012 - 02/2013)

Courses Taught: Data Structures, Discrete Structures

Lecturer (Visiting)

Quaid-i-Azam University, ISB, Pakistan   **(09/2009 - 02/2011) & (09/2018 - 06/2019)**

Courses Taught: Intro. to Computer Systems, Problem Solving and Programming, Object Oriented Programming

University of Technology Sydney, Australia   **(03/2018 - 06/2018)**

Course Taught: Data Structures and Algorithms

Air University, ISB, Pakistan   **(02/2015 - 06/2015)**

Courses Taught: Data Structures, Programming Fundamentals

Arid Agriculture University, RWP, Pakistan   **(02/2013 - 01/2015)**

Courses Taught: Modern Programming Languages, Software Engineering

Federal Urdu University of SA & Tech, ISB, Pakistan **(02/2012 - 01/2013)**
Courses Taught: Visual Programming, Distributed Databases

Professional Experience

Software Developer, Acoem (Ecotech Pty Ltd), Melbourne, Australia (06/2022 – 02/2023)
Responsibilities:
- Design, development, and integration of software features for Airodis
- Maintenance and enhancement of the software products which are used for weather data analytics
- Design and implementation of service-oriented AirodisX – a cloud-based system for AQMS

Tech Stack:
- C#, VB.Net, .Net Core, .Net Microservice Architecture, Entity Framework, WCF, SSH.Net, Fluent API, SOAP/RESTful APIs, AutoMapper, WinForms, JSON, XML, TOML, SQLite, SQLCe, SQL Server, VS.Net, Jira, Confluence, Jenkins (CI/CD), GitHub, GitKraken

Software Developer, Advanced Analytics Institute, UTS, Sydney, Australia (07/2017 - 06/2018)
Responsibilities:
- Development of novel graph-based methods to transform text stream.
- Design and development of FinQ - an AI-powered micro-finance application.

Tech Stack:
- Python, MongoDB, Scikit-learn, Pandas, numpy, NLTK, tweepy, networkX, Google Cloud Platform (GCP), PHP, CodeIgniter, RESTful APIs, JSON, Bootstrap, HTML5, MySQL, AWS(S3, EC2, RDS), GitHub

Software Engineer, Diyatech, Islamabad, Pakistan (02/2012 - 10/2012)
Responsibilities:
- Development, integration, and maintenance of software NCache
- analyzing newly developed features for evaluating system performance
- Writing source code for the integration of end-users' APIs with the existing software application

Tech Stack:
- C#, SQL Server, PostgreSQL, Sockets, Protocol Buffers, SOAP, VSS, Bugzilla

Software Engineer, M-Net Technologies, Islamabad, Pakistan (07/2010 - 02/2012)
Responsibilities:
- Project initialization, requirement gathering, design and development
- Writing efficient and scalable software source code
- Testing the source code, debugging for the error tracing, and ensuring the quality
- Time and cost estimation, resource allocation and client communication.

Tech Stack:
- ASP.Net, C#, VB.Net, Infragistics, Crystal Reports, Entity Framework, MVC, SQL Server, VS.Net, VSS

Project Developer, BUITEMS, Quetta, Pakistan (04/2006 - 04/2007)
Responsibilities:
- Design and development of security system for computer laboratories
- Reading data from ground sensors and laser sensors
- Controlled visual surveillance, and secure entrance & exit

Tech Stack:
- C#, VB.Net, Infragistics, Crystal Reports, SQL Server, parallel port programming

Research Interests and Publications

- Data science, machine learning, text mining, and Knowledge Graph
- Information extraction and retrieval systems
- Social computing (Social Network Analysis, Social Media Analytics)

Published citations?user=o99g-7oAAAAJ&hl=en

In-Progress
1. Social Network Analysis of Russia-Ukraine War based on Twitter Data - (Tentative Sub: Jan-2024)
2. Identifying Conspiracy Theories in Twitter based on Deep Learning - (Tentative Sub: Feb-2024)

Workshops and Events

- o Training workshop on "Enhancing Research Productivity", organized by University of Central Punjab, Lahore, Pakistan, held on 25th September 2021
- o Training workshop on "Teaching Excellence Program", organized by University of Central Punjab, Lahore, Pakistan, held on 30th August 2021 to 2nd September 2021
- o Conference Talk on "What can we sense from Twitter", ECIR'20, Lisbon, Portugal.
- o Workshop on "Use of Artificial Intelligence Against COVID-19", organized under China-Pakistan Science Workshop by PASTIC, QAU, Pakistan and BTBU China, held on 23rd July 2020.
- o Training workshop on "Introduction to teaching first year subjects ", organized by Institute for Interactive Media and Learning, University of Technology Sydney, held on 26th March 2018
- o Training workshop on "Active learning in UTSOnline", organized by Institute for Interactive Media and Learning, University of Technology Sydney, held on 27th March 2018
- o Training workshop for casual academics on "Tools for Online Feedback", organized by Institute for Interactive Media and Learning, University of Technology Sydney, held on 28th March 2018
- o Conference Talk on "Text Stream to Temporal Network", PAKDD'18, Malborne, Australia.
- o Poster Presentation in Faculty of Engineering & IT Research Showcase event, organized by FEIT, University of Technology Sydney, held on 14th June, 2018

Community Services

Invited/Sub Reviewer:
- International Joint Conference on Artificial Intelligence (IJCAI'24)
- International Conference on Computer Modeling and Big Data Analytics (CMBDA'22), China
- Pacific Rim International Conference on Artificial Intelligence (PRICAI'18), China
- Int. Conference on Advances in Social Networks Analysis and Mining (ASONAM'18), Spain
- IEEE International Conference on Big Data Science and Engineering, (BigDataSE'18), USA

Reviewing for Journals:
- Expert Systems with Applications, Elsevier
- Geo-Spatial Information Science, Taylor & Francis
- Social Network Analysis and Mining, Springer Link

Awards

- **Student Grant** by 42nd European Conference on Information Retrieval (**ECIR'2020**)
- **Research fellowship** by Advanced Analytics Institute, **University of Technology Sydney**, Australia. (Feb 2018 - July 2018)
- **Research funding** for the foreign research visit under **International Research Support Initiative Program** by the Higher Education Commission (**HEC**), Pakistan. (Aug 2017-Jan 2018)
- **Ph.D. Fellowship** by Higher Education Commission **(HEC)**, Pakistan (Feb 2013-Jul 2017)
- **Merit scholarship** in M.Phil, Department of Computer Science, **Quaid-i-Azam University**, Islamabad, Pakistan. (2007-2008)