UMER MAJEED

Data Scientist

umermjd11.github.io

Islamabad, Pakistan

+92 311 1577 484

umermjd11@gmail.com

/in/umermjd11

Kaggle umermjd11

CV umermjd11.github.io/cv

github.com/umermjd11

github.com/umermajeedkhu

scholar.google.com

user=LrsLEJgAAAAJ

Citations: 650+

SUMMARY -

Experienced Data Scientist and Ph.D. candidate in Computer Science & Engineering, Umer Majeed excels in utilizing Python, R, and SQL for data analysis and machine learning projects. Skilled in data visualization tools like Plotly and Dash, with a strong foundation in Pandas and NumPy for data manipulation. Demonstrated success in developing predictive models and conducting in-depth exploratory data analysis. Published researcher with a focus on leveraging data science and AI for innovative solutions. Eager to apply expertise in statistical analysis and machine learning to drive impactful insights in a collaborative data science environment.

SKILLS -

PLs & Frameworks: Python, R, SQL, C++, Julia, Dash, TensorFlow, PyTorch, Keras.

Libraries & Technologies: NumPy, pandas, Matplotlib, Plotly, Seaborn, scikit-learn, NLTK, ggplot2, Microsoft Excel, IBM Cognos Analytics, Google Looker Studio

Familar IDEs: JupyterLab/ Jupyter Notebook, PyCharm, RStudio, VS Code, Google Colab

Familar OS: Ubuntu, Windows

KEY RELEVANT PUBLICATIONS -

Umer Majeed et al., "Cross-Silo Model-Based Secure Federated Transfer Learning for Flow-Based Traffic Classification," ICOIN 2021.
Developed a **federated transfer learning** scheme for traffic classification on **time-related statistical features** using DL and **TensorFlow Federated** on multi-organizational datasets, enhancing accuracy and efficiency through knowledge transfer in a cross-silo setting. Ensured data privacy in federated learning by implementing a **secure aggregation** protocol.

Umer Majeed et al., "Cross-Silo Horizontal Federated Learning for Flow-based Time-related-Features Oriented Traffic Classification," APNOMS 2020.
Developed a **horizontal federated learning** model for **traffic classification** on **TensorFlow Federated**, utilizing flow-based **time-related statistical features** to enhance **data privacy** and **security**. Demonstrated the effectiveness of deep learning techniques in traffic classification in **cross-silo** settings.

Umer Majeed et al., "Vanilla Split Learning for Transportation Mode Detection using Diverse Smartphone Sensors," KCC 2021.
Implemented a **split learning** framework for **transportation mode detection** leveraging **smartphone sensors** to enhance **data privacy** and reduce **client-side computation**. Showed that the **split neural network** achieves comparable performance to traditional deep learning models while being more robust against **inference attacks**.

Umer Majeed et al., "Blockchain-assisted Ensemble Federated Learning for Automatic Modulation Classification in Wireless Networks," KCC 2020.
Proposed an **ensemble federated learning** scheme for **automatic modulation classification** (AMC) using **deep learning** techniques on **decentralized data**. Leveraged a **blockchain network** to enhance model training and demonstrated improved performance of the ensemble model over base federated models in **wireless communication** systems.

RELEVANT CERTIFICATIONS AND MOOCS -

DataCamp - - https://www.datacamp.com/portfolio/umermajeed 2017-2018

These certifications cover **Python Basics**, including **Data Types**, **Data Visualization**, and libraries like **Pandas**, **NumPy**, **Seaborn**, and **Matplotlib**. They also include **EDA**, **Statistical Thinking**, **Statistical Analysis**, **Relational Databases**, **SQL**, **SQL JOINs**, **SQL Aggregation**, and tools like **Git/GitHub** and **CLI piping**.

1. Introduction to Python - - 2017
2. Intermediate Python - - 2017
3. Intermediate SQL - - 2017
4. Introduction to Shell - - 2018
5. Functions in Python - - 2017
6. Python Toolbox - - 2017
7. Statistical Thinking (1) - - 2017
8. Statistical Thinking (2) - - 2017
9. Version Control - Git - - 2018
10. Data Types in Python - - 2017
11. Data Visualization - - 2017
12. Data Visualization - Seaborn - - 2018

Deep Learning Specialization - Coursera -

This specialization covers key **Deep Learning** concepts like **Neural Networks**, **Back-propagation**, **Regularization**, and **Optimization**, using frameworks such as **TensorFlow**. It includes architectures like **CNNs** and **RNNs**, advanced topics like **GRU**, **LSTM**, **Attention Models**, and **Transformers** for NLP, with a focus on practical implementation and **optimization strategies**.

1. Neural Networks and Deep Learning - Jul. 2021 -
2. Improving Deep Neural Networks - Aug. 2021 -
3. Structuring Machine Learning Projects - Oct. 2021 -
4. Convolutional Neural Networks - Oct. 2021 -
5. Sequence Models - Audit Completed with Labs - Nov, 2024.

IBM Data Science Professional Certificate - Coursera - Audit Completed with Labs -

This certification covers essential **Data Science** skills, including **Data Visualization**, **Data Management**, **Machine Learning**, and **Data Analysis**. It emphasizes hands-on experience with **Python**, **SQL**, and **CRISP-DM**, exploring **Data Pipelines**, **Feature Engineering**, **Big Data**, and **Model Deployment**. Practical projects involve **data collection**, **wrangling**, and **exploratory analysis**, building a solid foundation for a career in data science.

1. What is Data Science? - April 2024
2. Tools for Data Science - April 2024
3. Data Science Methodology - April 2024
4. Python for Data Science, AI & Development - April 2024
5. Python Project for Data Science - April 2024
6. Databases and SQL for Data Science with Python - May 2024
7. Data Analysis with Python - May 2024
8. Data Visualization with Python - June 2024
9. Machine Learning with Python - June 2024
10. Applied Data Science Capstone - August 2024
11. Generative AI: Elevate Your Data Science Career - July 2024
12. Career Guide and Interview Preparation - August 2024

IBM Data Analyst Professional Certificate - Coursera - Audit Completed with Labs -

This certification provides job-ready **Data Analytics** skills, focusing on **data cleaning**, **visualization**, and **dashboards**. It covers tools like **Python**, **Excel**, **SQL**, and libraries such as **Pandas**, **NumPy**, and **scikit-learn**, along with **Jupyter Notebooks**, **Google Looker**, and **Cognos Analytics**. Skills in **EDA**, **predictive modeling**, **generative AI**, and **machine learning** are applied in projects involving dashboard creation and real-world data analysis.

1. Introduction to Data Analytics - Sep. 2024
2. Excel Basics for Data Analysis - Sep. 2024
3. Data Visualization & Dashboards - Excel & Cognos - Sep. 2024
4. Generative AI: Enhance your Data Analytics Career - Sep. 2024
5. Career Guide & Interview Preparation - Oct. 2024

Private & Secure AI/Data Science Courses - OpenMined -

1. ***Our Privacy Opportunity*** - Completed - Mar. 2021 - Explore structured transparency, **privacy techniques**, and the **privacy-transparency trade-off**.
2. ***Foundations of Private Computation*** - Ongoing - Progress 80% - Implement **federated learning**, **secure multi-party computation**, **homomorphic encryption**, and **differential privacy**.
3. ***Introduction to Remote Data Science*** - Completed - Feb. 2022 - Use **remote execution tools**, deploy **Domain Nodes**, and apply **privacy-preserving techniques** for distributed data science.

PROJECTS & PORTFOLIO -

ML Project - - SpaceX Falcon 9 launches - Kaggle Notebook - , Dash App - - This project covers key aspects of **machine learning** such as **data collection** (via API and web scraping), **data wrangling**, **exploratory data analysis (EDA)**, and the creation of **visualizations** and **interactive dashboards** using **Plotly Dash** and **Folium**. The project also applies **predictive analysis** through classification techniques to forecast launch success rates.

DL projects - - using TensorFlow, keras, PIL, transformers

1. Simple CNNs - Happyface & Digit hand Signs - - Github
2. ResNet - Digit hand Signs - - Kaggle NB -
3. Transfer Learning - MobileNet - - Kaggle NB -
4. Object Detection using yolov2 - - Github NB
5. Image segmentation using Unet - - Kaggle NB -
6. Face recognition using facenet - - Github NB
7. DL Art - Neural Style Transfer - - Kaggle NB -
8. RNN from Scratch - Dinosaur Island - - Kaggle NB -
9. Text generation - LSTM based RNN - - Kaggle NB -
10. Music Generation - LSTM based RNN - - Kaggle NB -
11. Word Embeddings - Similarity & Debiasing - - Github NB -
12. Emojifier: Expressiveness with Emoji - - Github NB -
13. Neural Machine Translation with Attention - - Github NB -
14. Trigger word detection - from voice - - Kaggle NB -
15. Transformer from Scratch - - Github NB -
16. Explore Positional Encodings - Transformer - - Github NB -
17. Named-Entity Recognition - Transformer - - Kaggle NB -
18. Extractive QA - Transformer - - Kaggle NB -

Exploratory Data Analysis (EDA) Projects - - using matplotlib, plotly, pandas

1. **Tesla and GameStop Stock/Revenue Data** - Kaggle Notebook - : involves data fetching via `yfinance`, analysis of key metrics, trends, and a summary of **market behavior** and **financial performance**.
2. **Socioeconomic Indicators in Chicago (2008-2012)** - Kaggle Notebook - : using **pairplots**, **heatmaps**, **correlation matrix**, and **descriptive statistics**.

Dashboard & Visualization Projects - - using Google Looker

1. **Sales and Service Analysis Report for SwiftAuto Traders** - Looker Report - : A comprehensive dashboard analyzing **car sales and service performance**, featuring KPIs like **total profit**, **quantity sold**, and visualizations of **sales by model**, **profit by dealer**, **recalls per model**, **customer sentiment**, and trends in **monthly sales** and **profit**.
2. **Products and Sales Analysis Report for Customer Loyalty Program** - Looker Report - : Detailed insights into **product sales** and **customer loyalty**, with data on **total revenue**, **quantity sold**, and visualizations including **line charts**, **bar charts**, **treemaps**, **gender slicers**, and **revenue by geography** through maps and word clouds.

EDUCATION

2017 - Present — **Master & Ph.D. (Combined) in Computer Science & Engineering** — **CGPA 4.11/4.3**
Department of Computer Science & Engineering, Kyung Hee University, Yongin, South Korea

2011 - 2015 — **BS Electrical (Telecommunication) Engineering** — **CGPA 3.83/4.00**
National University of Sciences & Technology (NUST), Islamabad, Pakistan

EXPERIENCE

2015 – 2016 — **PHP developer** — **Artologics, Islamabad, Pakistan**

- Developed robust back-end applications using Core PHP and CodeIgniter framework.
- Implemented jQuery and JavaScript to facilitate smooth communication between the user interface and server-side components via AJAX requests, enhancing the interactivity of web application.
- Employed SQL queries to interface with MySQL databases, ensuring data integrity and reliability while developing robust solutions for efficient data management.

`PHP / SQL / CodeIgniter / jQuery / AJAX / JavaScript / APIs`

LANGUAGES

English - Proficient (written and verbal), **Urdu** - Native, **Korean** -Beginner (TOPIK Level 2)