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: 600+

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 covers essential skills in **Python Basics**, including **Data Types**, **Data Visualization**, and libraries such as **Pandas**, **NumPy**, **Seaborn**, and **Matplotlib**. It also includes fundamental concepts of **Exploratory Data Analysis (EDA)**, **SQL**, **Statistical Thinking**, and **Statistical Analysis**. The curriculum emphasizes **Relational Databases**, **SQL JOINs**, **SQL Aggregation**, as well as tools like **Git/GitHub** and command line operations including **CLI piping**.

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

Deep Learning Specialization - Coursera -

This specialization covers key concepts and techniques in **Deep Learning**, including **Neural Networks**, **Back-propagation**, **Hyperparameters**, **Regularization**, **Optimization**, and frameworks like **TensorFlow**. Participants learn to implement various architectures, including **Convolutional Neural Networks (CNNs)** and **Recurrent Neural Networks (RNNs)**. Advanced topics such as **GRU**, **LSTM**, **Attention Models**, and **Transformers** for natural language processing (NLP) are also explored. The coursework emphasizes practical implementation and optimization strategies to achieve high performance in deep learning tasks.

1. Neural Networks and Deep Learning - Completed - Jul. 2021 -
2. Improving Deep Neural Networks - Completed - Aug. 2021 -
3. Structuring Machine Learning Projects - Completed - Oct. 2021 -
4. Convolutional Neural Networks - Completed - Oct. 2021 -
5. Sequence Models - In Progress

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

This comprehensive certification program encompasses a wide range of essential skills in **Data Science**, focusing on **Data Visualization**, **Data Management**, **Machine Learning**, and **Data Analysis**. Participants gain hands-on experience with various tools and techniques, including **Python**, **SQL**, and **CRISP-DM methodology**. The coursework covers topics like **Data Pipelines**, **Feature Engineering**, **Data Augmentation**, **Big Data**, and **Model Deployment**. Participants also engage in practical projects, such as the **Applied Data Science Capstone**, where they perform **data collection**, **wrangling**, and exploratory analysis using real-world datasets, including predicting Falcon 9 rocket landings. This certification is ideal for developing a robust foundation in data science and preparing for a successful career in the field.

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 comprehensive certification equips participants with job-ready skills in **Data Analytics**, emphasizing practical experience in **data cleaning**, **data visualization**, and **dashboards**. The program covers essential tools such as **Python**, **Excel**, and **SQL**, with advanced training in **Python libraries** (e.g., **Pandas**, **NumPy**, and **scikit-learn**), **Jupyter Notebooks**, **Google Looker** and **Cognos Analytics**. Participants develop proficiency in **exploratory data analysis**, **predictive modeling**, **generative AI**, and **machine learning**, and complete hands-on projects, including building interactive dashboards and analyzing real-world datasets. The program also offers **interview preparation** and career support to ensure a smooth transition into the field of data analytics.

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 -

SpaceX Falcon 9 ML Project - Kaggle Notebook - , Dash App -

This project focuses on **SpaceX Falcon 9 launches**, covering 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.

Tesla and GameStop Stock/Revenue Data and Dashboard - Kaggle Notebook -

This notebook provides an **exploratory data analysis (EDA)** of **Tesla and GameStop Stock/Revenue Data**, including **data fetching** via 'yfinance', analysis of key metrics and trends, and a summary of **market behavior** and **financial performance**.

Socioeconomic Indicators in Chicago (2008-2012) - Kaggle Notebook -

This notebook presents an EDA of **socioeconomic indicators in Chicago (2008-2012)** through **pairplots**, **heatmaps**, **correlation matrix**, and **descriptive statistics**, identifying trends and relationships between key features.

Sales and Service Analysis Report for SwiftAuto Traders - Looker Report -

This Looker dashboard provides a comprehensive analysis of **car sales and service performance** at SwiftAuto Traders. The **Sales Dashboard** highlights KPIs such as **total profit**, **quantity sold**, and visualizations of **sales by model** and **profit by dealer**. The **Service Dashboard** focuses on **recalls per model**, **customer sentiment**, and trends in **monthly sales** and **profit**.

Products and Sales Analysis Report for Customer Loyalty Program - Looker Report -

This Looker report presents a detailed analysis of **product sales** and **customer loyalty** metrics, including **total revenue**, **quantity sold**, and various visualizations such as **line charts**, **bar charts**, and **treemaps** for product lines, **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)