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/citations?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

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

Familar OS: Ubuntu, Windows

KEY RELEVANT PUBLICATIONS

⋆ DAO Framework
⋆ ERC-721
⋆ Multi-Signature Contracts
⋆ Non-Transferable Tokens (NTTs)
⋆ IPFS
⋆ Hardhat

Umer Majeed et al., "DAO-FL: Enabling Decentralized Input and Output Verification in Federated Learning with Decentralized Autonomous Organizations," TechRxiv. Preprint, Dec. 2023. **www.github.com/umermajeedkhu/DAOFLcode/tree/main/contracts**

- Developed a decentralized framework for input and output verification in federated learning, enhancing data integrity and security using DAOs and ERC-721 tokens.
- Designed DAO Membership Tokens (DAOMTs) for governance, implementing mintable and soul-bound tokens to facilitate decentralized decision-making in machine learning systems.

⋆ Structured Transparency
⋆ Homomorphic Encryption
⋆ Input & Output Privacy
⋆ Output Verification

Umer Majeed et al., "ST-BFL: A Structured Transparency empowered cross-silo Federated Learning on the Blockchain framework," IEEE Access, Nov. 2021.

- Developed a framework enhancing data privacy in federated learning through structured transparency and homomorphic encryption, ensuring secure handling of sensitive data.
- Implemented output verification and smart contracts for accountability in collaborative machine learning, facilitating reliable model performance validation and effective data management.

⋆ Transfer Learning
⋆ Federated Learning
⋆ TensorFlow Federated
⋆ Secure Aggregation
⋆ Time-related Statistical Features

Umer Majeed et al., "Cross-Silo Model-Based Secure Federated Transfer Learning for Flow-Based Traffic Classification," ICOIN 2021. **https://doi.org/10.1109/ICOIN50884.2021.9333905**

- Developed a federated transfer learning scheme for traffic classification using deep learning 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.

⋆ Deep Learning
⋆ Traffic Classification
⋆ Horizontal FL
⋆ TensorFlow Federated
⋆ Feature Engineering
⋆ Time-related Statistical Features

Umer Majeed et al., "Cross-Silo Horizontal Federated Learning for Flow-based Time-related-Features Oriented Traffic Classification," APNOMS 2020. **https://doi.org/10.23919/APNOMS50412.2020.9236971**

- Developed a federated learning model for traffic classification using flow-based time-related features, enhancing privacy and data security.
- Demonstrated the effectiveness of deep learning techniques in traffic classification, showcasing advanced data analysis methodologies.

KEY RELEVANT CERTIFICATIONS AND MOOCS

⋆ Python Basics
⋆ Data Types
⋆ Data visualization
⋆ pandas
⋆ numpy
⋆ seaborn
⋆ matplotlib
⋆ EDA
⋆ SQL
⋆ Statistical Thinking
⋆ Statistical Analysis
⋆ Relational Database
⋆ SQL JOINs
⋆ SQL Aggregation
⋆ Git/ Github
⋆ Command line
⋆ CLI piping

Datacamp
https://www.datacamp.com/portfolio/umermajeed

- Introduction to Python - Completed - 2017
- Intermediate Python - Completed - 2017
- Data Types for Data Science in Python - Completed - 2017
- Python Data Science Toolbox (Part 1) - Completed - 2017
- Python Data Science Toolbox (Part 2) - Completed - 2017
- Statistical Thinking in Python (Part 1) - Completed - 2017
- Statistical Thinking in Python (Part 2) - Completed - 2017
- Introduction to Version Control with Git - Completed - 2018
- Intermediate SQL Queries - Completed - 2017
- Introduction to Shell - Completed - 2018
- Introduction to Data Visualization in Python - Completed - 2017
- Intermediate Data Visualization with Seaborn - Completed - 2018

⋆ Generative AI Tools
⋆ Data Augmentation
⋆ Querying Databases
⋆ Feature Engineering
⋆ Ethics in AI
⋆ Data Visualization
⋆ Matplotlib
⋆ Seaborn
⋆ Dash / Plotly
⋆ Geospatial Data
⋆ Dashboards
⋆ ML Models
⋆ Data Pipelines
⋆ DDL / DML
⋆ Advanced SQL
⋆ Python Integration
⋆ Cloud Databases
⋆ CRISP-DM
⋆ Financial Data
⋆ Data Management
⋆ Data Integration
⋆ Model Building
⋆ Model Deployment
⋆ Model Monitoring
⋆ Cloud-Based Tools
⋆ Data Literacy
⋆ ETL
⋆ Big Data
⋆ Data Pipelines

IBM Data Science Professional Certificate - Coursera - Audit Completed with Labs
https://www.coursera.org/professional-certificates/ibm-data-science

- **Generative AI: Elevate Your Data Science Career - July, 2024** - Learn generative AI tools for data preparation and querying, engage in hands-on labs, explore real-world use cases, and understand ethical considerations in data science.
- **Machine Learning with Python - June, 2024** - Study machine learning fundamentals, including regression, classification, and clustering methods. Gain practical experience with Python libraries and complete a final project to showcase your skills.
- **Data Visualization with Python - June, 2024** - Implement data visualization techniques with libraries such as Matplotlib and Plotly, build interactive dashboards, and apply skills through hands-on projects.
- **Data Analysis with Python - May, 2024** - Develop skills in data cleaning, exploratory data analysis, and visualization. Build and evaluate ML models, and create efficient data pipelines.
- **Databases and SQL for Data Science with Python - May, 2024** - Learn SQL from basics to advanced, integrate with Python, and work with real-world datasets. Explore relational and cloud databases.
- **Data Science Methodology - April, 2024** - Apply CRISP-DM methodology to structure projects, prepare data, build and evaluate models, and understand iterative improvements.
- **Python Project for Data Science - April, 2024** - Extract and analyze stock data using Python, build dashboards to visualize trends, and demonstrate proficiency in data analysis projects.
- **Python for Data Science, AI & Development - April, 2024** - Learn Python basics, data structures, Pandas, Numpy, web scraping, REST APIs, and data collection methods.
- **Tools for Data Science - April, 2024** - Explore tools for data management, integration, visualization, model building, and deployment. Learn about popular open-source and cloud-based tools.
- **What is Data Science? - April, 2024** - Understand data science fundamentals, career paths, big data processing, ETL, and data pipelines. Gain insights into data science applications and cloud computing.

⋆ Neural Networks
⋆ Back-propagation
⋆ Hyperparameters
⋆ Regularization
⋆ Optimization
⋆ TensorFlow
⋆ Dropout
⋆ CNN Basics
⋆ Advanced Architectures
⋆ Object Detection
⋆ RNNs
⋆ GRU
⋆ LSTM
⋆ Attention Models
⋆ NLP
⋆ Transformers

Deep Learning Specialization - Coursera
https://www.coursera.org/specializations/deep-learning

- **Neural Networks and Deep Learning - Completed - July, 2021 -** ✹ - Gain a deep understanding of neural networks, implement architectures, and optimize through hyperparameter tuning and regularization.
- **Improving Deep Neural Networks - Completed - August, 2021 -** ✹ - Explore advanced techniques like hyperparameter tuning, optimization algorithms (Adam, RMSprop), regularization methods (dropout, batch normalization), and implement models using TensorFlow.
- **Structuring Machine Learning Projects - Completed - Oct. 2021 -** ✹ - Diagnose errors in ML systems, implement strategies like end-to-end learning and transfer learning, and set human-level performance benchmarks for complex tasks.
- **Convolutional Neural Networks - Completed - Oct. 2021 -** ✹ - Explore CNN layers, advanced architectures like ResNet, apply object detection techniques (YOLO, U-Net), and create models for applications like face recognition and neural style transfer.
- **Sequence Models - In progress** - Implement RNNs, GRUs, LSTMs, and transformers for NLP tasks like machine translation and named entity recognition, and apply attention mechanisms for enhanced performance.

⋆ Privacy Techniques
⋆ Federated Learning
⋆ Secure Multi-Party Computation
⋆ Differential Privacy
⋆ Remote Execution

Private & Secure AI/Data Science Courses - OpenMined
https://courses.openmined.org/courses

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

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)