Mutaz Younes

Albany, New York | +1 (641) 261-9124 | mutazyounes@gmail.com | linkedin.com/in/almutazyounes

Data Scientist and AI Lead with over 6 years of experience driving data-centric solutions and leading AI initiatives across multiple sectors, including healthcare, telecom, and finance. Expertise in designing and deploying scalable machine learning models and natural language processing (NLP) systems, with a focus on automation, data integrity, and privacy compliance. Proven track record in leading cross-functional teams, mentoring talent, and implementing cloud-based AI infrastructures to streamline decision-making and extract insights from complex data sources. Skilled in developing AI-driven applications powered by techniques such as retrieval-augmented generation (RAG) and large language models (LLMs) for tasks like clinical data extraction and decision support. Passionate about building ethical, privacy-conscious AI systems that enhance data accessibility and drive innovation.

Experience

Office of Mental Health | Data Scientist/ Programming Lead | Albany, USA | 2023 - Current

- Leading OMH's first AI development team by establishing a collaborative, innovation-driven culture focused on mentorship, professional growth, and cross-functional collaboration. Responsible for strategic planning, talent development, and ensuring alignment of AI initiatives with organizational objectives, while fostering a knowledge-sharing environment to empower junior data scientists and drive long-term success.
- Created and deployed a secure document interaction system using RAG, enabling staff to efficiently search sensitive healthcare documents through an intuitive chat interface.
- Developed an AWS-powered LLM solution to streamline EHR vendor selection by automating RFP requirement analysis, significantly reducing executive review time.
- Delivered multiple educational presentations to non-technical staff, simplifying complex AI concepts and demonstrating practical applications of AI tools, such as large language models (LLMs) and retrieval-augmented generation (RAG), to enhance organizational efficiency and decision-making. focused on promoting AI literacy and empowering staff to leverage AI technologies effectively in their roles.
- Conducted an internal Python training course to upskill the data analysis team, covering core programming concepts, data manipulation, and automation techniques to enhance team capabilities.
- Developing a database-aware synthetic data generation system that maintains referential integrity while creating artificial healthcare records for safe AI experimentation.
- Working on creating a local LLM-based meeting documentation tool that automatically generates structured minutes and action items from recorded meetings.
- Leading a comprehensive codebase modernization initiative to enhance maintainability and efficiency across all projects.

UNDP Jordan | Natural Language Processing Consultant | Amman, Jordan | 2022 - 2024

- Led a project to analyze 70,000+ tourist reviews, applying transformer-based models for topic extraction and sentiment analysis. Successfully identified up to 18 topics per review using zero-shot learning, providing the UNDP with nuanced insights into tourist sentiment.
- Developed and deployed a public-facing website that enabled citizens to report daily issues. Implemented machine learning techniques and LLMs to classify and filter reports, ensuring accurate categorization and prioritization before data storage. This improved the efficiency of issue management and reporting for the UNDP team.
- Utilized the Twitter API to collect and analyze over 1 million tweets. Applied NLP techniques to classify tweets into relevant topics, providing valuable social media insights. Created a Tableau dashboard to visualize and present the findings to UNDP stakeholders, aiding in data-driven decision-making.

Open Insights | Data Scientist | Amman, Jordan | 2020 - 2022

- Implemented Python-based AutoML framework to predict asset renewal, leveraging H2O platform, enhancing scalability and efficiency, ultimately saving the client significant time and resources when handling new datasets.
- Designed and implemented a scalable framework to analyze 200+ billion records of customer data for a large telecom company using PySpark, SQL, and Databricks. Used the feature store design pattern and applied machine learning models from MLlib in Spark to predict missing data, providing a more comprehensive understanding of the customer base.
- Produced compelling visual dashboards utilizing Tableau for two large telecom companies, displaying insights from data analysis in easy-to-understand format.
- Implemented time series prediction algorithms including ARIMA, Prophet, and XGBoost to analyze historical data and forecast future customer balance for a major bank in the US.
- Applied customers segmentation and data analysis techniques to smart home devices sales data, revealing valuable insights

and trends.
- Collaborated in authoring material for a major bank's Data Academy program in Jordan, developing 3+ modules, including the Deep Learning module, covering ANN, RNN, LSTM, and CNN.

JUST | Research Assistant / Teacher Assistant | Irbid, Jordan | 2019 - 2020
- Conducted research and published 5 research papers while pursuing my master's degree.
- Demonstrated expertise in NLP by participating in 4 online competitions and achieving top rankings in 2 of them.

Additional Projects | 2024 - 2025
- De-identification & Adverse Event Tool:
  - Collaborating with KHCC and Ahmed Younes, a pharmacist and researcher at Deggendorf Institute of Technology, on the development of Prophytech-AI. This AI-driven tool de-identifies unstructured medical notes, extracts adverse event information, and grades them based on CTCAE standards. The project employs a combination of rule-based algorithms, transformer-based NER models, and large language models to ensure PHI protection while maintaining data integrity for clinical research purposes.
- Urban Area Expansion Prediction:
  - Collaborated with a master's student from Jordan University of Science and Technology (JUST) on a research project using an Artificial Neural Network (ANN) to predict future urban growth based on historical land use patterns. The project involved analyzing geographic and demographic data to forecast expansion trends, providing insights to support data-driven urban planning and development decisions.
- Post-Pandemic Urban Planning and Regional Development:
  - Collaborated with a master's student from Jordan University of Science and Technology (JUST) on a research project exploring strategies for sustainable urban planning and regional development in the wake of the COVID-19 pandemic. Built predictive models using XGBoost and Artificial Neural Networks (ANN) to forecast urban expansion and population distribution trends based on historical land use and public health data. The project focused on identifying key urban growth patterns and provided data-driven insights to support more resilient infrastructure planning and policy recommendations for post-pandemic recovery.
- Data Science instructor:
  - Served as an instructor for a Data Science course at PHI Science Institute, focusing on core data science concepts, machine learning techniques, and practical applications in Python.

Education

- **Masters in data science**: Maharishi International University, Iowa, USA, 2022 – 2025
  - Relevant Courses: Cloud Computing, Big Data, Big Data Technology, Machine Learning, Big Data Analytics.

- **Masters in computer science (Thesis Track)**: Jordan University of Science and Technology, Jordan, 2019 – 2021
  - Relevant Courses: Advanced Artificial Intelligence, Machine Learning, Natural Language Processing, Advanced Methods in Data Mining, Big Data Management.

- **Bachelors in computer science**: Albalqa Applied University, Jordan, 2014 – 2018

Technical Knowledge

Large Language Models (retrieval-augmented generation, agentic frameworks, fine-tuning BERT, LangChain), Python, NumPy, Pandas, Scikit-learn, TensorFlow, Keras, PyTorch, Tableau, Power BI, Machine Learning (linear regression, logistic regression, decision trees, random forest, SVM, XGBoost), Deep Learning (ANN, RNN, LSTM, CNN), Big Data (Hadoop, Spark), Cloud Platforms (AWS, Databricks), Time Series Prediction, Data Analysis, Machine Learning Design Patterns, Model Validation, MLOps, Feature Store Design.