Almas Fathimah

667-334-1464 | almasfathima4@gmail.com

https://www.linkedin.com/in/almas-fathimah-90/

https://github.com/AlmasFathimah124/

EDUCATION

University of Maryland, Baltimore County, Baltimore, MD — GPA: 3.77
Master of Professional Studies, Data Science — *Expected, Dec 2024*

Nimra College of Engineering and Technology, Ibrahimpatnam,India — GPA: 7.1
Bachelor of Engineering and Technology,Electronics and Communication Engineering — *Graduated, May 2012*

SKILLS

Programming Languages	**:**	Python, SQL, C,C++, HTML/CSS, PL/SQL, Java
Databases	**:**	MySQL, Oracle, MongoDB, SQL Server
Tools	**:**	IBM Rational DOORS,ALM
Analytics Tools	**:**	SSRS, SSIS, Microsoft Excel, Tableau, SAS, Power BI, SSAS
Big Data Technologies	**:**	Hadoop, Hive, Spark, Kafka, YARN
AWS Cloud Tools	**:**	EC2, EMR, S3
Python Packages	**:**	Pandas, PySpark, NumPy, Matplotlib, SciPy, OpenCV, CNN models, Dash, sklearn, keras, Rest API's
IDE's	**:**	PyCharm, Jupyter Notebook, Eclipse, Visual Studio
OS/Platforms	**:**	Windows, Mac OS, Linux
Embedded Debugging Tools	**:**	UART, SPI, I2C
Real-Time Systems	**:**	(Learning RTOS - FreeRTOS)

Version Control & Development Tools : Git, SVN, DOORS, ALM

PROFESSIONAL WORK EXPERIENCE

Data Science Intern, *CalAmp , CA, United States of America* — *Sep 2024 – Dec 2024*

- Designed a ReAct-based AI system for natural language SQL generation, integrating Azure OpenAI embeddings and FAISS with 50+ query-SQL examples to enable RAG reducing data retrieval time by 70%, and improving query accuracy by 60%
- Developed a Streamlit UI for SQL execution and result visualization, providing instant insights through interactive charts.
- Demonstrated multistep reasoning of ReAct in Streamlit, improving query execution clarity and increasing user satisfaction by 40%.
- Built an ETL pipeline in Python to extract and load data into SQL Server, incorporating API pagination handling for efficient and seamless data ingestion from source application.

Senior System Engineer, *Amadili Labs Private Limited , Bangalore, India* — *Jun 2020 – Aug 2022*

- Developed and optimized software solutions in Python and SQL, leveraging Apache Spark, Hadoop for data processing.
- Implemented big data pipelines and BI tools like Power BI, Tableau to generate actionable insights.
- Automated data validation, ETL processes, and reporting using SQL, Pandas to streamline decision-making.
- Secured API integrations & data flows, ensuring data integrity and encryption for seamless data exchange.

Embedded Software Developer, *UST GLOBAL TECHNOLOGIES , Bangalore, India* — *April 2017 – Jan 2018*

- Developed unit test procedures for embedded software written in c,c++
- Developed automated test cases using Cantata for the ParkPilot system at Bosch, enhancing software validation efficiency.
- Implemented unit testing and integration testing for embedded automotive software, adhering to industry standards.
- Followed MISRA C coding guidelines to maintain high-quality, safety-compliant software development.
- Conducted boundary value checks to validate variable ranges and ensure software reliability.
- Collaborated with cross-functional teams to analyze and resolve software defects, improving system reliability.

Software Engineer, *Artech Infosystems, Bangalore, India* — *Aug 2016 – Feb 2017*

- Developed test cases in C/C++ using Cantata for an automotive Near-Range Camera System at Bosch.
- Uploaded Cantata test results to DOORS and managed test scripts/code in MKS Integrity.
- Ensured requirements compliance and test coverage by linking Cantata test cases with system specifications.
- Worked on Embedded Systems with UART, SPI, I2C, UDS protocols.
- Automated system validation and testing using Python scripting

Contract Assignee, *Tata Consultancy Services, Hyderabad, India* — *Apr 2015 – Jan 2016*

- Designed and implemented test cases using Cantata, based on customer requirements for airbag systems.
- Developed and tested embedded software in C/C++ and Assembly across multiple platforms.
- Ensured compliance with safety standards and validated software for airbag control systems.

- Implemented low-level device drivers and performed debugging for microcontroller-based embedded systems.
- Gained exposure to real-time embedded systems principles, including task prioritization and interrupt handling

ACADEMIC PROJECTS

UMBC Data Science Chatbot Development *October 2024*

- Scraped website data using Selenium and scheduled automated updates with Jupyter Scheduler for real-time content retrieval.
- Generated embeddings from scraped data using OpenAI's Ada model and stored them in ChromaDB for efficient semantic search and fast information retrieval.
- Developed an NLP-powered chatbot leveraging LLM models to assist with queries related to UMBC's Data Science program.
- Implemented Retrieval-Augmented Generation (RAG) using Dialogflow ES, LangChain, and FastAPI, enabling dynamic and context-aware responses.

Technologies Used: NLP, LLM models (OpenAI Ada), Selenium, ChromaDB, FastAPI, LangChain, Dialogflow ES, Python, Jupyter Notebook.

Multi-Label Clinical Text Eligibility Classification and Summarization System *April 2024*

- Processed and embedded clinical text data using Word2Vec and TF-IDF, enabling multi-label classification.
- Developed classification models using Support Vector Machine (SVM) and Random Forest, enhancing eligibility prediction accuracy and decision-making efficiency.
- Generated text summaries with TextRank, Luhn summarization, and GPT-3, ensuring concise and relevant content extraction.
- Evaluated summarization quality using ROUGE scores, ensuring high text coherence and accuracy.

Technologies Used: NLP, LLM models, word embedding (word2vec), TF-IDF, random forest, SVM, TextRank, Luhn summarization, GPT-3, ROUGE Score,Transformers.

Python Powered Stock Market Analytics with MongoDB, PySpark, and Tableau *Dec 2023*

- Extracted stock data from APIs and stored it in MongoDB, utilizing PyMongo for seamless database interactions.
- Processed and analyzed market trends using PySpark, applying transformations and aggregations for deep insights.
- Visualized market insights in Tableau, enabling data-driven decision-making.

Technology and Tools used: Apache Spark, MongoDB, PySpark, Tableau, PyMongo, Twelve Data API, Yahoo Finance API

Predicting Patient Outcomes using ASA Classifications *Dec 2023*

- Built a machine learning model to classify ASA physical status and predict discharge times for perioperative risk assessment.
- Preprocessed data by handling missing values, encoding categorical variables, and removing outliers using IQR filtering.
- Used Random Forest, Gradient Boosting, and Ada Boost classifiers for ASA classification and Linear Regression, Random Forest Regressor for discharge time prediction.
- Applied hyperparameter tuning and class balancing techniques to improve model performance and mitigate overfitting.
- Evaluated models using accuracy, F1-score, R^2 score, MAE, and RMSE, with Gradient Boosting performing best for ASA classification.
- Developed a predictive system for clinical decision-making and preoperative risk assessment, leveraging Vital Db Clinical Data for enhanced patient outcome analysis.

Technology used: Machine Learning: Python, scikit-learn (sklearn); Data Analysis: Pandas, NumPy; Data Visualization: Matplotlib, seaborn; Web Development: Flask (for building web applications to visualize results or interact with models)

Financial Data Analysis: *April 2023*

- Built an ETL pipeline using Visual Studio, integrating SQL Server Integration Services (SSIS) for automated data extraction, transformation, and loading.
- Connected and managed SSIS workflows to streamline data processing and ensure efficient data flow.
- Performed data cleansing and transformation using SSMS, optimizing sales data for accurate analysis.
- Developed automated data pipelines, enabling seamless sales performance tracking and reporting.
- Visualized key sales insights using Power BI, supporting data-driven decision-making.

Technology used - ETL, SQL,Power BI,Visual Studio

Project Management System: *April 2023*

- Built a project management system using Python, leveraging Object-Oriented Programming (OOP) principles for modular and reusable code.
- Implemented search and sorting algorithms to optimize system performance and data retrieval.
- Developed unit and integration tests in Python, ensuring code reliability and maintainability.
- Applied Python's data structures and algorithms to enhance efficiency and scalability.

Technology used - Python, Object-Oriented Programming, Search Algorithms, Sorting Algorithms, Testing

CERTIFICATIONS

- **CJET in Embedded Systems** *April 2015*
- **Foundations of Project Management Offered by Google issued by Course era** *August 2021*
- **HTML, CSS, and Javascript for Web Developers Offered by Johns Hopkins University** *July 2021*
- **Complete Java Certification Program Udemy** *Sep 2020*