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: 3.7
Master of Professional Studies, Data Science — *Graduated, May 2012*

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 (EC2, EMR, S3), Databricks
Data Integration & Processing :Apache Kafka, PySpark, ETL Pipelines, Managed File Transfer (MFT)
Monitoring & Debugging :Kibana, Grafana, Splunk, AWS CloudWatch
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*

Brief description:

CalAmp is a leading telematics and asset tracking company. This project focused on developing a ReAct-based AI chatbot that allows users to generate and execute SQL queries using natural language. The system uses Azure OpenAI, FAISS, and Streamlit to enable dynamic query generation, visualization, and retrieval-augmented generation (RAG) for improved business insights.

Key Responsibilities:

- Designed a ReAct-based AI chatbot using Azure OpenAI and FAISS for natural language to SQL conversion, supporting Retrieval-Augmented Generation (RAG) workflows.
- Developed a responsive Streamlit UI for real-time SQL query execution and interactive data visualization.
- Built a scalable ETL pipeline in Python to extract, transform, and load data from paginated APIs into SQL Server.
- Conducted EDA using Matplotlib and Seaborn to identify data patterns and support downstream modeling.
- Collaborated with data engineers to ensure pipeline integrity, schema consistency, and data accuracy.
- Created Power BI dashboards to visualize operational KPIs and automate reporting for internal stakeholders.
- Applied NLP and classification techniques using Scikit-learn, NLTK, and TF-IDF for structured output generation.
- Automated repetitive data handling tasks and model workflows using scripting and logging best practices.
- Evaluated model performance using precision, recall, and F1-score, tuning them for production readiness.
- Documented system logic, chatbot architecture, and deployment steps for internal handoff.

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

Brief description:

This project supported Universal Group's Corporate Finance team in generating data-driven insights for executive-level reporting, including retrospective analysis related to their 2016 and 2017 finance white papers. The focus was on building a scalable data analytics system to process, validate, and visualize large volumes of operational and financial data.

Key Responsibilities:

- Designed and implemented ETL workflows using Python, SQL, and AWS Glue to automate ingestion, transformation, and cleansing of multi-source financial data.
- Built Power BI dashboards for real-time performance monitoring, enabling executives to make data-informed decisions.
- Used AWS Lambda and Athena to process, query, and deliver analytics over structured data stored in S3.
- Automated log parsing and analysis workflows by integrating AWS Lambda with Athena to extract insights from S3-stored application logs in near real-time.
- Implemented version control for SQL scripts and pipeline logic, facilitating collaboration and rollback during updates.
- Developed SQL logic for generating ad-hoc reports, aligning technical solutions with stakeholder reporting needs.
- Ensured data quality through schema validation, anomaly detection, and pipeline-level monitoring.
- Created and maintained documentation of system design, data flows, and validation frameworks.
- Worked closely with analysts to gather requirements and deliver insights using Tableau and Power BI.
- Supported the integration of LLM-based summarization into white paper workflows, improving narrative generation.

- Conducted statistical analysis and developed simple ML models to evaluate risk and financial performance metrics.
- Automated repetitive data processing and dashboard refresh tasks to improve operational efficiency.
- Participated in Agile sprints and provided technical inputs during planning, testing, and stakeholder demos.

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

PARK PILOT (PP) for Driver Assistance Systems.:

Brief description:

Park pilot is one of the ECU in modern cars/Trucks with Automatic and Manual assist .It consists of different range of Ultrasonic sensors for detecting the parking slot and to provide multiple features like Side view Assist (SVA),Cross Parking(CP),Parallel Parking(PP), Home Zone parking, Park assist and Guidance.

Key Responsibilities:

- Conducted software validation for ECU behavior related to ultrasonic-based parking assistance features.
- Simulated sensor and signal behaviors using CANalyzer and CANoe to verify CAN bus communication.
- Supported execution of test cases built in Cantata and assisted in test data preparation.
- Worked on validation setups to verify system-level behavior from software components interacting with test benches.
- Verified signal routing, actuator triggering, and response logging within vehicle simulation environments.
- Assisted in identifying log-based issues and flagged anomalies during integration runs.
- Collaborated with developers to understand feature-level expectations and report test findings.
- Contributed to generation of traceability matrices and validation coverage documentation.
- Participated in team-level validation planning and test execution discussions.
- Supported partial automation for recurring software validation steps using in-house tools.

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

Near Range Camera Systems:

Four cameras are fitted in the car as one camera is used at front, and the other two are fitted at the side front mirrors on left and right respectively. At the time of parking the car, these cameras will take the snapshot of the object (or) obstacle and each snap will be captured in a chip called 'Salsa'. This salsa will send the image to 'Zynq' in 1Gbit/s speed with the help of Ethernet switch, then 'Zynq' will recognize the object, decode the image and store the image in its RAM memory. Then send the processed image to 'i.MX6' and then displays the image which is now converted into a video in the frame. It uses the latest CMOS(complementary metal oxide light semiconductor) technology for its cameras ensuring excellent image quality even when faced with great differences in brightness or poor conditions

Key Responsibilities:

- Performed test executions on simulation setups using pre-defined input signals and software test scripts.
- Monitored CAN-based data exchange between ECUs and camera modules using CANalyzer.
- Created and modified basic test scripts for functional checks and log analysis.
- Assisted in evaluating latency and image frame processing under different conditions.
- Supported issue reproduction and root-cause identification based on software behavior.
- Participated in test reviews and aligned testing outcomes with expected software specifications.
- Coordinated with teams to update calibration settings and verify their impact.
- Maintained and configured test benches for camera software validation.
- Documented anomalies, test coverage status, and regression results as part of engineering workflow.

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

Airbags:

Airbag system is a safety critical control system that resides inside a Car and is primarily responsible for passenger protection during a crash event.

Module# 1: **Remote sensors(TRW-FH)**

Description:

Airbag control units (ACUs) offer robust crash detection The airbag control unit detects and evaluates a crash, before triggering the appropriate restraint systems according to the type of collision and its severity. Information is supplied to the control unit by as many as six satellite acceleration sensors.

Module # 2: **Squibs(TRW-FH)**

Description:

Squibs are main output devices of our airbag system. Squibs are used in vehicles to deploy an airbag when crash occurs. When a crash occurs, the electronic control unit response to that crash condition causes a pre determinable quantum of electrical energy to be delivered to an ignitable airbag squib. The squib, which is in communication with the gas generator, causes the generator to inflate the airbag by releasing rapidly an inert, non-toxic gas, such as nitrogen.

Key Responsibilities:

- Supported module-level testing of embedded C code for airbag deployment logic.
- Conducted signal simulation and output validation for squib driver modules.
- Used CANoe and CANalyzer to monitor in-vehicle network data and responses.
- Analyzed ECU response to satellite sensor inputs during simulated crash scenarios.
- Performed manual and automated testing using Cantata and in-house tools.
- Participated in peer reviews for safety requirement traceability.
- Verified diagnostic routines using UDS protocol.
- Created boundary condition test cases and failure mode simulations.

- Logged defects and retested patches until closure.
- Maintained compliance documentation and participated in audit preparations.

ACADEMIC PROJECTS

UMBC Data Science Chatbot *Oct 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 *Apr 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 analysed 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.
- Pre-processed 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 *Apr 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 *Apr 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.

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

CERTIFICATIONS

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