

Amitha A
(469) 581-9587 | amitha8841@gmail.com

SUMMARY

Highly driven Data Scientist and AI/ML Engineer with 6+ years of experience delivering end-to-end analytics, machine learning, and GenAI solutions across supply chain, healthcare, and BFSI domains. Skilled in designing scalable data pipelines, developing production-grade ML systems, and building advanced Agentic AI and RAG architectures using LangChain and LangGraph. Adept at transforming business requirements into automated, data-powered solutions using Python, SQL, PySpark, Databricks, and cloud technologies across AWS, Azure, and GCP. Recognized for improving operational efficiency, driving measurable business impact, and leading cross-functional teams to accelerate AI adoption.

- Specialized in AI/ML, Generative AI, Agentic AI, and enterprise-grade analytics across Manufacturing, Supply Chain, Healthcare, and BFSI.
- Expert in building end-to-end ML pipelines including ingestion, feature engineering, model training, deployment, and monitoring.
- Advanced expertise in RAG systems, LangChain, LangGraph, vector databases (Chroma/FAISS/Pinecone), and multimodal LLM workflows.
- Strong experience implementing distributed ETL pipelines using PySpark, Databricks, AWS Glue, and Apache Airflow.
- Proficient in developing ML models using TensorFlow, PyTorch, Scikit-learn, XGBoost, and Transformer-based architectures.
- Proven ability to design cloud-native AI platforms leveraging AWS (S3, Lambda, Glue), Azure (ADF, Synapse), and GCP (BigQuery, Vertex AI).
- Deep knowledge of data modeling, data quality frameworks, data governance, and scalable data architecture.
- Skilled in MLOps including CI/CD, MLflow, Docker/Kubernetes, feature stores, model registry, and model drift detection.
- Experienced in building enterprise dashboards using Power BI, Tableau, and Looker to support data-driven decisions.
- Excellent communicator known for translating complex ML concepts into actionable business insights, reducing costs, and enabling advanced analytics adoption.

WORK EXPERIENCE

- Senior Data Scientist, Americold Logistics, Dallas, TX** Aug 2024 - Present
- Designed and deployed ML models for demand forecasting, inventory optimization, and cold-storage anomaly detection, improving prediction accuracy by 18% and reducing spoilage events.
 - Built scalable RAG pipelines using enterprise documents, embedding models, and vector databases to support operational decision-making.
 - Developed GenAI solutions using LangChain, Azure OpenAI, and LangGraph, including summarization, agent workflows, and warehouse Q&A systems.
 - Engineered automated ETL frameworks in Databricks + PySpark, reducing data processing times from 4 hours to under 40 minutes.
 - Created real-time Power BI dashboards for inventory KPIs, warehouse performance, SLA adherence, and temperature variance alerts.
 - Implemented feature stores and model registries using MLflow, improving consistency and auditability of production model updates.
 - Optimized picking/packing and route-planning workflows using ML, reducing operational delays by 12% across multiple facilities.
 - Built and containerized APIs for ML serving using FastAPI, Docker, and Kubernetes, ensuring low-latency inference.
 - Implemented data quality validation using Great Expectations, ensuring 98% accuracy and preventing downstream model failures.
 - Partnered with supply chain, operations, quality, and IT teams to drive enterprise adoption of predictive analytics.

- Data Analyst, Tomorrow's Leaders Today, Frisco, TX** Jan 2024 - May 2024
- Delivered advanced analytical insights for healthcare operations, improving reporting turnaround and insight clarity by 30%.
 - Created interactive Power BI dashboards integrating clinical workflows, appointment data, and compliance metrics.
 - Built secure SQL ETL pipelines to consolidate EHR, scheduling, billing, and claims data into unified analytical datasets.
 - Developed complex DAX calculations, RLS security models, and semantic layers for scalable enterprise reporting.
 - Automated monthly/weekly reporting workflows using Python + Power Automate, reducing manual effort by 40%.
 - Performed root-cause analysis on patient wait times, appointment cancellations, and throughput delays.
 - Built predictive indicators and anomaly alerts to identify deviations in patient engagement and care delivery.
 - Implemented star-schema data models supporting KPI dashboards and clinical performance evaluations.
 - Ensured strict HIPAA compliance in all analytics workflows by securing PHI data and access controls.
 - Collaborated directly with clinicians and directors to convert raw data into actionable operational improvements.

- Software Engineer, DXC Technology, Bangalore, IN** Jul 2019 - Dec 2022
- Developed ML models for customer churn prediction, fraud detection, and segmentation using Python and Scikit-learn.
 - Built enterprise-scale ETL pipelines using PySpark, Hive, and Hadoop, supporting high-volume data processing.
 - Designed Tableau and Power BI dashboards tracking customer KPIs, operational metrics, and risk indicators.
 - Monitored model performance using MLflow and built drift detection mechanisms to maintain model stability.
 - Engineered robust feature pipelines with automated missing-value handling, encoding, and outlier detection.
 - Wrote optimized SQL queries across Teradata, Oracle, and SQL Server, improving analytics execution time by 60%.
 - Automated weekly reporting processes using Python, enabling a 25% increase in team productivity.
 - Conducted data audits and implemented data quality checks across critical BFSI datasets.
 - Documented business logic, data dictionaries, lineage, and ETL workflows for cross-team clarity.

- Worked with cross-functional stakeholders including PMs, data engineers, SMEs, and client teams to deliver analytics solutions.

KEY SKILLS

Programming & Scripting: Python, SQL, Pyspark, HTML

Machine Learning & AI: Supervised & UnSupervised Learning, Deep Learning (CNN, RNN, LSTM, Transformers, GANs), Reinforcement Learning, NLP, Computer Vision, Generative AI, Agentic AI, RAG, Prompt Engineering (Zero-shot, Few-Shot, Chain-of-Thought, ReAct, Tree-of-Thoughts), LLM Fine-Tuning (LoRA, QLoRA), A/B Testing, Statistical analysis (ANOVA, Correlation)

Frameworks & Libraries: PyTorch, TensorFlow, Keras, Scikit-Learn, XGBoost, Hugging Face, LangChain, LlamaIndex, CrewAI, LangGraph, Pandas, NumPy, Matplotlib, Seaborn, OpenCV, NLTK

Data Engineering & Cloud: ETL / ELT Pipelines, Databricks, AWS (S3, EC2, CloudWatch, SNS, SQS), Azure (ADF, Blob Storage, ML Studio), GCP (BigQuery, Vertex AI), Vector Databases (FAISS, Chroma)

Tools, Visualization, BI & Collaboration: PowerBI, Tableau, Excel, Jupyter notebook, Google Colab, Git, GitHub, Cursor AI, VS Code, Copilot, FastAPI, Jira, ServiceNow, Agile/Scrum, SDLC

EDUCATION

UNIVERSITY OF TEXAS AT ARLINGTON

Arlington, TX

MASTER OF SCIENCE IN DATA SCIENCE

Jan 2023 - May 2024

Concepts: Machine Learning, Neural Networks, Deep Learning, Probability, Statistics, Data Visualization, Artificial Intelligence

ANNAMACHARYA UNIVERSITY

Andhra Pradesh, IN

BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING

Jun 2016 - Jul 2019

Concepts: Python, Database Management, Big Data, Cloud Computing, Machine Learning, Web development, Data Science, Data Structures and Algorithms, Thesis: "Object Detection with Deep Learning, Computer Vision and dual stage Convolutional Neural Network"