

AMITHA AKEPATI

Email: amithaake29@gmail.com | Mobile No: +1(469) 318-8841 | [GitHub](#) | [LinkedIn](#) | [Kaggle](#)

PROFESSIONAL SUMMARY

Results-driven Data Scientist with 4+ years of experience designing and deploying end-to-end AI/ML solutions focused on data analysis, predictive modeling, and Generative AI, including Large Language Models (LLMs), Vision Models (LVMs), Multimodal Models (LMMs), NLP, and Deep Learning. Skilled in building and fine-tuning scalable models for enterprise applications in predictive analytics and automation. Proven ability to extract, transform, analyze, and visualize data to deliver actionable insights and impactful dashboards that shape business strategy across multiple domains. Strong knowledge on driving innovation through AI/ML including LLM development and Agentic AI frameworks.

TECHNICAL SKILLS

Programming & Scripting: Python, R, SQL, PySpark, HTML, CSS

Machine Learning & AI: Supervised & Unsupervised Learning (Regression, Classification, Clustering), Deep Learning, Neural Networks (CNNs, RNNs, LSTMs, Transformers), Reinforcement Learning, Computer Vision, Natural Language Processing (NLP), Generative AI (LLMs, LVMs, LMMs), Agentic AI, Prompt Engineering (Zero-shot, Few-shot, Chain-of-Thought, ReAct), Retrieval-Augmented Generation (RAG), A/B Testing, Statistical Analysis (Hypothesis Testing, ANOVA, Correlation)

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

Data Engineering & Pipelines: ETL/ELT Pipelines, Databricks, Data Wrangling, Data Warehousing

Cloud Platforms & Operating Systems: AWS (S3, EC2, Lambda, SNS, SQS, STS), Microsoft Azure (Data Factory, Blob Storage, ML Studio), Google Cloud Platform (BigQuery, Vertex AI)

Business Intelligence & Visualization: Power BI (DAX, Power Query), Excel (Advanced, PivotTables), Jupyter Notebook & Lab, Google Colab

Tools & Methodologies: Git, GitHub, Anaconda, Spyder, Visual Studio Code, SDLC, Agile, Scrum, Jira

PROFESSIONAL EXPERIENCE

Data Scientist	Thermo Fisher Scientific	August 2024 – Present
<ul style="list-style-type: none">Developed Python scripts to integrate continuously incoming laboratory data, ensuring smooth data flow and reducing manual intervention.Built complex SQL queries with advanced analytical functions to extract actionable insights from large, multi-source datasets, supporting research planning and operational decisions.Automated ETL workflows using Databricks, reducing data processing time by up to 40% and improving overall operational efficiency.Designed and deployed machine learning models achieving 96% forecasting accuracy on lab research data, improving predictability and planning for experiments.Managed data integrity by handling missing values, encoding categorical variables, and performing feature engineering, enhancing overall data quality by 33%.Conducted experiments comparing classification and regression algorithms to identify the most effective predictive approaches, uncovering key factors influencing lab outcomes.Integrated AI-driven anomaly detection pipelines and explored Generative AI and LLM techniques to generate synthetic research data and simulate potential lab scenarios, improving model robustness and operational insights.Collaborated with cross-functional teams to gather business requirements, define project scopes, and ensure alignment with organizational objectives, enabling effective teamwork and project success.Ensured ethical data handling and transparency while delivering scalable, data-driven solutions aligned with organizational values.Documented workflows, ML experiments, and AI explorations for reproducibility, knowledge sharing, and continuous improvement of research processes.Designed scalable AI/ML pipelines for lab automation using Azure services and Databricks, reducing sample processing time by 27% and manual interventions by 40%.		

Tech Stack: Python, SQL, Databricks, PySpark, Scikit-learn, Pandas, NumPy, Matplotlib, ETL Automation, Machine Learning, Generative AI (LLMs, Agentic AI frameworks), Git, Jupyter Notebooks, Azure

Data Scientist Intern	Tomorrow's Leaders Today	January 2024 – May 2024
<ul style="list-style-type: none">Conducted end-to-end analysis on 10,000+ grant records, uncovering funding trends and improving outreach strategies for underrepresented communities.Developed ML classification models to identify high-potential grant opportunities, increasing targeting accuracy by 40% and boosting funding success rates.Applied AI/ML techniques to automate grant summarization and matching, reducing manual effort by 45% and streamlining proposal workflows.Designed and deployed interactive dashboards using Power BI and Tableau to present insights to leadership and non-technical stakeholders, enhancing strategic decision-making by 32%.		

- Automated data ingestion, preprocessing, and reporting pipelines using Python and SQL, improving data processing efficiency and retrieval speed by 20%.
- Ensured data quality and consistency through validation frameworks, increasing overall reliability across analytics systems.
- Created 50+ dynamic Power BI dashboards for key business stakeholders, enabling actionable insights from large datasets.

Tech Stack: Python, SQL, Power BI, Tableau, Scikit-learn, Pandas, NumPy, Machine Learning, Data Cleaning & Automation, Dashboarding, Data Validation, Jupyter Notebooks

Research Assistant – AI/ML

The University of Texas at Arlington

October 2023 - May 2024

- Collaborated with Prof. Aera on Generative AI applications in agriculture, focusing on crop yield prediction and recommendation systems.
- Developed ML and deep learning models (CNNs, LSTMs) on multi-source agricultural datasets including satellite imagery, sensor, weather, and soil data achieving >85% accuracy in crop yield prediction and soil health assessment.
- Designed and trained Generative Adversarial Networks (GANs) to generate synthetic agricultural data, enhancing model robustness and improving performance by up to 25% in data-scarce regions.
- Integrated Generative AI techniques with traditional agronomic models to create data-driven crop recommendations tailored to environmental factors, improving recommendation accuracy for farmers.
- Conducted statistical analyses (regression, correlation, hypothesis testing) to identify key variables influencing crop outcomes, guiding model optimization.
- Performed data preprocessing, feature engineering, and validation to ensure high-quality inputs for model training and evaluation.
- Documented research methodologies, results, and visualizations for peer-reviewed publications and real-world pilot projects.

Tech Stack: Python, TensorFlow, Keras, PyTorch, Scikit-learn, Pandas, NumPy, Matplotlib, CNNs, LSTMs, GANs, Satellite & Sensor Data, Statistical Analysis, Feature Engineering

Software Developer

DXC Technology

June 2021- August 2022

- Worked on Hewlett Packard Service Bus (HPSB), a middleware tool that transfers files between systems, ensuring smooth and reliable data movement between publishers and subscribers.
- Built Python scripts to monitor and automate file transfers, improving system reliability and reducing manual checks.
- Designed data validation and transformation workflows to clean and standardize files before delivery, improving accuracy and consistency across systems.
- Developed SQL queries and Power BI dashboards to track key metrics like transfer success rates and error trends, giving teams real-time visibility into data operations.
- Automated error detection and alert notifications using Python, AWS, and HPSB, helping reduce incident response time and improving issue resolution.
- Created a reusable data model and self-service tracking tools that made it easier for teams to monitor message flows and troubleshoot faster.
- Integrated Databricks for scalable data processing, making it easier to analyze transfer logs and operational metrics efficiently.
- Migrated legacy scripts to Python-based ETL frameworks, improving maintainability and speeding up processing by 25%.
- Helped streamline deployments by setting up a change management process in JIRA, cutting release timelines and reducing deployment issues.
- Collaborated with support teams to handle incidents and change requests in ServiceNow, ensuring smooth communication and timely resolutions.
- Documented workflows and created onboarding guides to help new team members understand systems faster.

Tech Stack: Python, AWS (EC2, S3, CloudWatch), SQL, Power BI, Databricks, HPSB (Hewlett Packard Service Bus), ServiceNow, JIRA, Git, ETL, Microsoft Visual Studio, Agile

EDUCATION

Master's in Data Science

The University of Texas at Arlington

January 2023 – May 2024

Arlington, Texas, US

Bachelor's in Computer Science

Annamacharya Institute of Technology and Sciences

June 2017 – July 2021

Rajampet, AP, India