JACOB KURIAKOSE

jkuriak3@asu.edu |+1 (602) 802-6591 | Tempe, AZ | www.linkedin.com/in/jacob-kuriakose

PROFESSIONAL SUMMARY

Data Scientist specializing in Machine Learning with hands-on experience in NLP and Time Series Forecasting. Led the migration of a production-grade AI chatbot to AWS, improving deployment efficiency by ~85%. Proficient in Python, TensorFlow, an cloud deployment, seeking Spring 2026 opportunities in data science, NLP, or machine learning engineering..

PROFESSIONAL EXPERIENCE

Arizona State University, *Research Assistant* ***May 2025 – Present***

- Led the migration of a production-grade AI chatbot infrastructure ("Waterbot") from CIC to CISA AWS environment.
- Provisioned AWS CDK (ECS, CloudFront, ECR, ALB, S3, RDS PostgreSQL, DynamoDB, Lambda), cutting manual setup ~40%.
- Designed RDS integration within CDK for persistent data storage, enabling Lambda and ECS tasks to securely access PostgreSQL via Secrets Manager and IAM roles.
- Built & pushed image to ECR; used ECS CLI redeploys instead of full stack deploys, reducing release time ~85%.
- Configured EC2 with IAM role-based access to securely authenticate with AWS services, eliminating the need for AWS configure.
- Resolved IAM, CDK bootstrap, and CloudShell bottlenecks to enhance migration stability and reduce deployment downtime by ~30%.
- Removed unintended CloudFront basic-auth function, restoring chatbot access for 2,000+ campus users.

Red Hat, *Software Quality Engineer Intern* ***January 2023 - July 2023***

- Performed manual and automated testing with GoLang, reducing average test execution time by 30%.
- Crafted detailed test cases using Polarion for thorough test coverage.
- Troubleshot software issues and leveraged Docker for scalable deployments.

BraynixAI, *Natural Language Processing Intern* ***June 2022 - August 2022***

- Improved legal document classification accuracy by 10% using Transformer-based NLP techniques.
- Built FastAPI-based ML model APIs to process legal and non-legal documents, streamlining workflows.

PROJECTS

Medical Chatbot *(Source Code)* ***Personal Project, Generative AI Model***

- Transitioned the medical chatbot from a static RAG pipeline to a ReAct-based agentic architecture with hybrid LLM orchestration (MedGemma + LLaMA-3), improving reasoning transparency, factual grounding, and response coherence.
- Integrated Pinecone vector retrieval (7K medical encyclopedia embeddings) with a cross-encoder reranker and ReAct tool-calling nodes (database + web), cutting retrieval latency by 35% while balancing verified medical data with real-time updates.
- Designed explicit Thought → Action → Observation reasoning chains and structured prompting to generate cited, user-friendly answers with medical disclaimers, improving interpretability and user trust.
- Built evaluation guardrails (grounded, safety, reasoning-trace scores) to ensure factual and medically safe outputs.
- Evaluated the ReAct-based medical chatbot on a 10-question QA dataset using 16 hybrid metrics, achieving BERTScore F1 ≈ 0.84, Overall System Quality ≈ 0.65, and Safety = 1.0, confirming factually grounded, safe, and semantically accurate reasoning behavior.
- Added conversation summarization & memory compression to preserve context while reducing token use and latency in long chats.
- *Tech Stack: Python, LangChain, LangGraph, Pinecone, HuggingFace, Cross-Encoder, MedGemma, LLaMA3 (Ollama), Flask*

Data-Driven Walmart Sales Predictions (*Source Code*) ***Personal Project, Time Series Forecasting***

- Analyzed SARIMA, Exponential Smoothing, LSTM, Random Forest, and LightGBM models to forecast store-level sales.
- Achieved 38.92% RMSE improvement with Exponential Smoothing over LSTM, enhancing inventory planning.
- Improved Gross Margin and Average Order Value through accurate sales predictions.
- *Technologies Used*: Python, Pandas, Matplotlib, NumPy, Scikit-Learn, Statsmodels, TensorFlow, Keras

SKILLS

Competencies: Data Cleaning, ML Algorithms (Linear Regression, Logistic Regression, Random Forest, KNN, SVM, etc.), Data Visualization, Exploratory Data Analysis, Data Mining, Generative AI, A/B Testing.
Programming Languages: Python (NumPy, Pandas, Matplotlib, Scikit-learn, Seaborn, SpaCy, NLTK etc.), C++, SQL.
Tools/Frameworks: TensorFlow, Keras, PyTorch, Docker, Git, PySpark, Jupyter Notebook, AWS Services, Tableau, MS Excel.

EDUCATION

Master of Science in Data Science, GPA: 4.00 ***August 2024 – May 2026***
Arizona State University *Tempe, Arizona*
Courses: Software Security, Data Mining, Statistical Machine Learning

Bachelor of Technology in Information Technology, CGPA: 9.0 ***August 2019 – July 2023***
Dr. Akhilesh Das Gupta Institute of Technology and Management *Delhi, India*
Courses: Operating System, Database Management System, Object-Oriented Programming, Computer Networks
Published Research Paper: *Monument Tracker: Deep Learning Approach for Indian Heritage*