Rakul Angamuthu Venkatesan

Edmond, USA | 405-230-2654 | rakulav01@gmail.com | LinkedIn | GitHub

Summary

Aspiring Data Scientist with 3.6 years of experience building scalable, production-grade ML systems and deploying cutting-edge generative AI and agent-based models. Specialized in large language models (LLMs) and cloud-based inference workflows. Proven record in improving model efficiency, reducing processing latency, and enabling real-time decision systems for high-stakes environments.

Education

University of Central Oklahoma — Edmond, United States
Master of Science - Data Science — Sep 2023 - Aug 2025

Skills

- **Languages & Databases:** Python, SQL, R, Java, PL/SQL, PostgreSQL, MySQL
- **ML & AI:** Scikit-learn, TensorFlow, PyTorch, Keras, Hugging Face, GPT (3.5/4), LoRA, OpenCV, NLTK
- **LLM & NLP:** RAG, LangChain, spaCy, Vector Embeddings, Pinecone, Prompt Engineering
- **Data & Analysis:** Pandas, NumPy, Tableau, Power BI, Matplotlib, Seaborn, Plotly
- **Cloud & MLOps:** AWS (SageMaker, EMR, Glue), Kubernetes, Git, Docker, Jupyter
- **Soft Skills:** Cross-functional collaboration, Problem Solving, Communication, Time Management, Agile Leadership

Professional Experience

Covalense Digital — Bengaluru, India
Senior Data Engineer — Aug 2020 - Jul 2023

- Designed and deployed **anomaly detection** models (Isolation Forest, Autoencoders) on financial datasets, reducing reconciliation false positives by 30% and accelerating fraud identification workflows by 2x.
- Architected an ML pipeline on AWS (SageMaker, Glue, EMR) to enable **daily retraining** and **real-time fraud detection**, improving model responsiveness under live transaction streams.
- Boosted transaction accuracy to 99.5% and cut latency by 40% through intelligent **feature engineering** and predictive validation models integrated into automated BRM data pipelines.

Petrasys Global Pvt Ltd — Mumbai, India
Machine Learning Intern — Jan 2020 - Jun 2020

- Developed and deployed a real-time **CNN-based acoustic classification** model using TensorFlow, achieving 96%+ accuracy in species detection and enabling wildlife surveillance at scale.
- Engineered a **multi-threaded monitoring system** for forest rangers, leveraging edge ML and concurrent audio stream processing to reduce manual tracking time by 60% and expand real-time coverage.
- Streamlined data handling with **SQL** for structured data storage and Firebase for real-time updates. Designed an intuitive app using Java, improving usability by 35%.

Projects

KnowBot: Semantic-Aware Chatbot with GPT & Vector Search — Jan 2025 - March 2025
Tech Stack: Python, RAG, TensorFlow, Hugging Face Transformers, NumPy, spaCy

Built a production-ready chatbot using RAG with GPT-4 and LoRA-tuned LLaMA, powered by LangChain agents and Pinecone vector search. Enabled multi-step reasoning, semantic retrieval, and dynamic tool use across 50k+ embeddings. Achieved 90%+ accuracy on 10K+ daily queries with sub-300ms latency. Deployed on AWS Lambda and Kubernetes for scalable, real-time performance in a modular GenAI pipeline.

FinAgent: Autonomous AI for Real-Time Market Forecasting — Sep 2024 - Nov 2024
Tech Stack: Python, TensorFlow, Scikit-learn, Plotly,VADER, Streamlit, AWS

Designed a forecasting system integrating LSTM-GRU models with LLM-driven sentiment analysis using GPT-4 for market news interpretation.Combined technical signals and real-time headlines using agentic workflows via LangChain, enabling autonomous reasoning over financial indicators. Achieved 97% accuracy on a decade of data. Deployed on AWS EC2 with a Streamlit dashboard for interactive, real-time market trend analysis.

Certifications

- Data Scientist Career Path - Codecademy
- Machine Learning by Andrew Ng - Coursera
- Data Science & Data Analytics with Python And R - Microsoft