Raj Kumar Nelluri

(201) 554-8009 | rajkumarn2002@gmail.com | LinkedIn | GitHub | Portfolio

Python • SQL • AWS SageMaker • Kinesis • Lambda • XGBoost • TensorFlow • ETL Pipelines • MLOps • REST APIs

Professional Summary

AWS Certified Cloud Practitioner and MS Computer Science graduate specializing in **machine learning systems** and **cloud-native data pipelines** on AWS. Architected end-to-end MLOps workflows – real-time distributed stream processing (Kinesis, Lambda), scalable data lake ingestion (S3), automated model training (SageMaker), and **model monitoring** with CloudWatch-driven retraining across datasets exceeding 500,000 records. Deployed **real-time inference endpoints** and REST APIs for live ML prediction serving in fraud detection, customer churn prediction, and demand forecasting systems.

Education

Pace University — New York, NY
Master of Science in Computer Science — *Sep 2023 – April 2025*
Amrita Vishwa Vidyapeetham — Bengaluru, India
Bachelor of Technology in Artificial Intelligence — *July 2019 – April 2023*

Professional Experience

Insurance Fraud Detection | *Python, Scikit-learn, Lambda, Kinesis, RDS, SageMaker* — *Capstone – Pace University* | 2024

- Architected a 4-stage cloud-native fraud detection pipeline – S3 data lake ingestion → Lambda ETL trigger → Kinesis distributed stream processing → SageMaker inference – automating end-to-end claim scoring without manual intervention
- Engineered ETL transformation layer using Python to validate schema, enforce data quality checks, impute missing values, and normalize 15,000+ insurance claim records across 30+ features before model ingestion
- Trained Random Forest and Gradient Boosting classifiers with 5-fold cross-validation and hyperparameter tuning; optimized classification threshold achieving approximately 90% accuracy and lower false positive rate on held-out test data
- Deployed SageMaker batch inference job and persisted fraud scores to AWS RDS data warehouse, enabling analyst teams to query ML predictions via SQL without re-executing the training pipeline

Customer Churn Predictor | *Python, XGBoost, Lambda, SageMaker, Comprehend, CloudWatch, EventBridge* | GitHub — 2025

- Designed end-to-end automated ML workflow on AWS – S3 data lake ingestion → Lambda ETL orchestration → SageMaker managed training job – reducing manual pipeline execution to zero after initial deployment
- Engineered NLP feature extraction layer using AWS Comprehend to transform 7,000+ unstructured customer support tickets into structured sentiment scores and key-phrase signals, expanding model feature dimensionality
- Deployed XGBoost classifier as a SageMaker real-time inference endpoint; performed hyperparameter tuning and decision threshold optimization achieving 0.67 recall for early high-risk churn detection
- Implemented automated model monitoring using CloudWatch metric alarms and EventBridge rules, triggering retraining pipelines upon data drift detection in production feature distributions

Retail Sales Forecasting | *Python, XGBoost, Pandas, NumPy, AWS S3, SageMaker* | GitHub — 2025

- Built scalable batch ETL data pipeline using Python and AWS S3 data lake to ingest, aggregate, and transform 500,000+ retail transactions into structured daily time-series datasets for downstream ML forecasting
- Engineered 10+ temporal features – lag variables, 7/30-day rolling averages, and holiday indicators – across a 2-year sales history to capture weekly seasonality and demand patterns for XGBoost forecasting
- Trained XGBoost regression model via SageMaker managed training job with hyperparameter tuning; deployed as a persistent real-time inference endpoint and evaluated against a held-out validation set using MAE and RMSE

Crypto Price Forecasting | *Python, XGBoost, TensorFlow, LSTM, CNN, Flask* — *Capstone – Amrita University* | 2023

- Built automated REST API ingestion pipeline to collect Bitcoin OHLCV price history and on-chain blockchain metrics daily, producing a versioned structured dataset used to train and benchmark 3 ML architectures
- Benchmarked 3 ML architectures (LSTM, CNN, XGBoost) on an 80/20 chronological train-test split; selected XGBoost as the production model based on lowest MAE and RMSE performance on held-out test data
- Deployed XGBoost model as a Flask REST API serving live 7-day Bitcoin price forecasts; exposed a JSON prediction endpoint powering a web dashboard with sub-second response latency

3D Face Generation (NeRF) | *Python, JAX, TensorFlow, OpenCV, COLMAP* — *Capstone – Amrita University* | 2023

- Implemented Deformable Neural Radiance Fields in JAX/TensorFlow for photorealistic 3D face reconstruction from monocular video; built multi-stage preprocessing pipeline using OpenCV and COLMAP for structure-from-motion camera pose estimation
- Evaluated novel-view synthesis quality using PSNR metrics and reduced GPU memory consumption by over 30% through iterative batch size optimization, enabling training on standard academic GPU hardware

Certifications

AWS Certified Cloud Practitioner – Amazon Web Services — 2025
Adobe Journey Optimizer Foundation – Adobe — 2025
Complete Python Developer – Udemy — 2023

Technical Skills

Programming: Python, SQL
ML & MLOps: Scikit-learn, XGBoost, TensorFlow, Keras, Feature Engineering, Hyperparameter Tuning, Model Deployment, Model Monitoring
Data Engineering: ETL/ELT Pipelines, Batch Processing, Distributed Stream Processing, Data Validation, Data Quality, Schema Enforcement, Scalable Data Architecture
Cloud Infrastructure (AWS): S3 (Data Lake), SageMaker, Lambda, Kinesis, RDS (Data Warehouse), Glue, EC2, CloudWatch, EventBridge
APIs & Deployment: REST APIs, Flask, Real-Time Inference, Batch Inference
Data Processing: Pandas, NumPy, JSON, CSV
Visualization: Tableau, Power BI, Matplotlib
Tools: Git, Jupyter Notebook, VS Code