

Rushyanth Nerellakunta

rushyanth55@gmail.com|[Linkedin](#)|[Github](#)|Indianapolis,IN

Summary

Data Scientist and **Machine Learning Engineer** with 3+ years of experience building predictive models, automating data pipelines, and deploying production-scale ML systems. Strong background in statistical modeling, experimentation, and optimization with hands-on expertise in **Python**, **SQL**, **PyTorch**, and **AWS** cloud services. Experienced in working with structured and unstructured data across domains such as **healthcare**, **finance**, and **consumer analytics**. Proven ability to design and operationalize end-to-end ML solutions, conduct A/B testing, develop recommendation systems, and apply causal inference for business impact. Adept at collaborating with cross-functional teams to translate data insights into strategic decisions.

Skills

- **Programming:** Python, SQL, R, Bash, C++, Git, JavaScript
 - **Data Science:** Data Ingestion, Feature Engineering, Model Evaluation, Data Structures, Data Modeling, Statistical Methods, Linear Algebra, Calculus
 - **Machine Learning & DL:** PyTorch, TensorFlow, W&B, GANs, CNN, RNN, ANN, NLP (spaCy, NLTK)
 - **GenAI & LLMs:** OpenAI, Gemini, Claude API's, Transformer, Bert, LangChain, LangGraph, MLflow, Prompt Engineering, RAG Pipelines, Embeddings, Fine-Tuning, ChromaDB, Pinecone
 - **Backend & Cloud:** FastAPI, Flask, Docker, Kubernetes, CI/CD, Airflow, Azure (exposure), AWS (SageMaker, S3, Lambda, CloudWatch), Spark, A/B Testing, Snowflake
 - **Visualization & BI:** Power BI, Tableau, Plotly, Streamlit, Excel, Power Automate
 - **GIS & PM:** ArcGIS Pro, Hazus, Project Management (Agile)
 - **ML Algorithms:** Linear/Logistic Regression, Decision Trees, Random Forest, XGBoost, LightGBM, MLR, K-Means, KNN, SVM
-

Professional Experience

AIML Engineer — Kahana(Aug 2025 – Present)

- Spearheading the development of an **AI-powered interaction layer** on top of Firefox, enabling users to control and navigate the browser through **natural voice commands**.
 - Integrating **real-time speech-to-text transcription** using Deepgram and Gemini APIs to deliver smooth and accurate conversational interactions between the user and browser.
 - Leveraging **AWS Lambda** to build scalable serverless functions and **AWS Cognito** for secure authentication and user session management.
 - Designing and implementing **modular voice command pipelines**, enabling dynamic handling of navigation, search, and browser automation tasks.
 - **Reduced LLM token usage by 35%** through intelligent **caching of frequently used commands** and lightweight prompt routing, optimizing latency and cost.
 - Collaborating with UI and backend teams to ensure **low-latency responses**, intuitive UX, and seamless integration with Firefox architecture.
-

Research Data Scientist — COMET Lab, Indiana University (Aug 2024 – May 2025)

- Engineered **ETL pipelines** and modeled structured/unstructured research data, reducing analytics turnaround time by **40%**.
 - Executed **60+ ML benchmarking experiments**, improving SQL translation model accuracy from **78% to 92%**.
 - Built prompt-engineering pipelines using **LLaMA** and OpenAI models, increasing automation efficiency.
 - Developed **Tableau dashboards** to monitor model KPIs and error distributions in real time.
 - Presented research findings to **technical and leadership stakeholders**, driving data-informed strategy.
-

Data Analyst — Virgenverse (Feb 2022 – Nov 2023)

- Developed **churn prediction and forecasting models** using Python and SQL, reducing customer churn by **8%**.
 - Automated ETL pipelines, improving refresh reliability and cutting manual effort by **40%**.
 - Designed and analyzed **A/B and multivariate tests** for marketing and pricing optimization.
 - Delivered **Tableau dashboards** visualizing product performance and retention metrics.
 - Partnered with cross-functional teams to integrate predictive insights into retention strategies.
-

Algorithmic Trader & Data Analyst (Independent) (Jan 2020 – Feb 2022)

- Built predictive **time series forecasting models** (ARIMA, SARIMA, LSTM) and signal-based engines for financial datasets.
 - Conducted **quantitative risk assessments** and “what-if” scenario modeling to optimize portfolio performance.
 - Automated ingestion and ETL processes, improving reporting speed and model reliability.
 - Presented actionable insights through **visual dashboards** for non-technical stakeholders.
-

Projects

- **Healthcare AI Agent (RAG System)** – Designed a multi-agent architecture using LangChain + ChromaDB to recommend personalized health insurance plans; improved factual accuracy by **27%**.
 - **Financial RAG Assistant** – Built a GenAI chatbot using LangChain and OpenAI APIs to automate financial reporting; reduced query response time by **50%**.
 - **Airline Price Prediction** - Developed an end-to-end flight fare prediction system by performing extensive feature engineering (date/time extraction, stops, airline/destination encoding) and training regression models including ARIMA/SARIMA for temporal trends and boosting models (XGBoost, LightGBM), then applied model-averaging ensembles to enhance predictive stability and accuracy.
 - **Apple Sentiment Analysis** – Built an end-to-end sentiment analysis pipeline (TF-IDF + Naive Bayes + VADER) and deployed a **Streamlit web app** for real-time text classification.
 - **Customer Churn Dashboard** – Developed predictive models using XGBoost & Logistic Regression; delivered interactive Tableau dashboards for retention strategy.
 - **GoDaddy Microbusiness Density Forecasting (Kaggle)** – Built a hybrid forecasting pipeline using ARIMA/SARIMA, XGBoost, and LightGBM, and implemented **model averaging ensembles** to boost accuracy and leaderboard performance.
 - **GAN Fashion MNIST** – Trained a Wasserstein GAN for apparel image generation, improving downstream model generalization.
 - **CNN Face Identification** – Built a TensorFlow/Keras CNN achieving **95%+ accuracy** with real-time inference.
 - **Chest X-ray Classification** – Developed ResNet50 + ViT multi-label classifier (ROC-AUC 0.86+) with Grad-CAM visual explanations.
 - **Flood Risk Assessment — Bartholomew County, IN** – Modeled flood hazard scenarios using ArcGIS Pro & Hazus; informed FEMA mitigation strategies.
-

Education & Certifications

- **M.S. in Applied Data Science**, Indiana University Indianapolis (2025)
 - **B. Tech in Electronics & Communication Engineering**, Sreenidhi Institute of Science and Technology (2022)
-

Professional Certification

- [**AWS Machine Learning Engineer - Associate**](#)