Rekha Srinivas Chimpiri

College Park, MD | rekhasrinivaschimpiri@gmail.com | 240-604-1490 | LinkedIn | GitHub

PROFESSIONAL SUMMARY

Data Scientist and Machine Learning Engineer with hands-on experience building and evaluating machine learning and generative AI systems in Python. Skilled in RAG and agentic LLM systems, predictive modeling, causal inference, and forecasting, along with the evaluation harnesses and data pipelines that make results reliable and reproducible. Experienced in taking models from problem framing through validation and deployment using Python, SQL, PyTorch, scikit-learn, LangGraph, FastAPI, and AWS.

WORK EXPERIENCE

Data Scientist, DreamStudio, Atlanta, GA **Sep 2025 - Present**

- Built a Retrieval-Augmented Generation (RAG) pipeline in Python with LlamaIndex and a Chroma vector store for natural-language Q&A and summarization across multi-document enterprise PDF collections.
- Implemented model evaluation workflows for LLM outputs using faithfulness, answer relevance, retrieval quality, and regression-style checks to keep responses reliable as the system changed.
- Benchmarked a multi-model ML pipeline (Logistic Regression, SVM, MLP, Random Forest, XGBoost) to uncover interpretable links between industry activity, biodiversity, and land use across 3,000+ U.S. counties.
- Partnered with stakeholders to translate ambiguous business questions into structured analyses and production-ready outputs, owning delivery end to end.

Graduate Teaching Assistant, Machine Learning Systems, University of Maryland, MD **Aug 2024 - May 2025**

- Debugged model-training, containerization, deployment, and data-pipeline failures across Docker, Kubernetes, MLflow, and Git in 160+ students' ML projects, resolving issues in experiment tracking and reproducibility.
- Reviewed and evaluated 160+ students' ML implementations, identifying gaps in model design, code quality, experimentation methodology, and deployment workflows.

AI Intern, Coincent, Remote, India **Jun 2022 - Dec 2022**

- Developed end-to-end CNN image-classification pipelines in Python and PyTorch, applying structured experimentation, benchmarking, and systematic architecture comparison under consistent training conditions, documenting experiments for reproducibility, and achieving over 90 percent test accuracy.
- Optimized the training pipeline through hyperparameter tuning, data augmentation, and architecture refinement, validating each change with precision, recall, F1-score, and error analysis, improving test accuracy 12 percent and reducing training time 20 percent versus baseline.

AI Intern, Internship Studio, Remote, India **Jan 2022 - May 2022**

- Fine-tuned transformer-based machine translation models through an iterative evaluation cycle, pairing targeted preprocessing with sentence alignment to prepare training data, and improved test BLEU score roughly 20 percent over the pretrained baseline.
- Automated a multilingual data pipeline spanning cleaning, tokenization, sentence alignment, preprocessing, and evaluation steps into one repeatable end-to-end workflow, cutting overall preprocessing time 40 to 50 percent for model training cycles.

EDUCATION

M.S. in Data Science, University of Maryland, College Park **Aug 2023 - May 2025**

Coursework: Probability and Statistics, Big Data Systems, Deep Learning, Cloud Computing, Natural Language Processing

TECHNICAL SKILLS

Machine Learning and AI: classification, regression, clustering, tree ensembles (XGBoost, LightGBM), deep learning, computer vision, NLP, LLMs, RAG, agentic workflows, model evaluation

Statistics and Experimentation: A/B testing, hypothesis testing, ANOVA, experimental design, causal inference, uplift modeling, forecasting, time-series, model validation

Programming and Data: Python, SQL, R, PyTorch, scikit-learn, Pandas, NumPy, PySpark, DuckDB, ETL, data pipelines, feature engineering

MLOps and Cloud: FastAPI, Docker, Kubernetes, MLflow, CI/CD, GitHub Actions, ONNX, AWS (Bedrock, Lambda, SageMaker), Azure

Certifications: Microsoft Certified Azure Data Scientist Associate, Azure AI Fundamentals

PROJECTS

Agentic RAG System with Tool Calls and Evaluation Harness **Apr 2026 - May 2026**

Python, LangGraph, FastMCP, BM25, sentence-transformers, cross-encoder reranking, DeepEval, FastAPI, CI/CD

- Designed an agentic RAG system with a LangGraph ReAct loop and FastMCP tool calls, using hybrid retrieval (BM25 plus embeddings fused with reciprocal rank fusion) and cross-encoder reranking over real FDA drug-label data.
- Measured quality with a DeepEval harness (faithfulness, answer relevancy, contextual recall and precision) gated in CI, lifting contextual recall from 0.67 to 0.83 and contextual precision from 0.82 to 1.00 over the baseline.

Causal Inference and Uplift Modeling on Ad-Incrementality Data **Apr 2026 - May 2026**

Python, FastAPI, scikit-learn, Propensity-Score Matching, T-Learner Uplift, Qini, DuckDB

- Estimated the causal effect of an advertising treatment on the 14M-row Criteo incrementality dataset using propensity-score matching and a T-Learner uplift model, measuring a PSM-adjusted ATE of +0.82 percentage points across 11.9M matched pairs.
- Quantified uplift with the Qini metric at 0.088 out-of-sample and surfaced heterogeneous treatment effects across deciles, serving the analysis and an automated briefing through a FastAPI dashboard.