

# SHANMUKHA JWALITH KRISTAM

📍 New York 📞 (934) 451-9567 ✉️ jwalithkristam@gmail.com 🌐 github.com/jwalith 📄 in/jwalith-kristam

## PROFESSIONAL SUMMARY

Data Scientist with 3+ years of experience leveraging ML and analytics to optimize client engagement and operational performance. Skilled in causal experimentation, predictive modeling, and translating complex data into business impact.

## EDUCATION

**Masters of Science in Data Science**, Stony Brook University, New York (GPA: 3.61/4) **Aug 2023 – May 2025**

**Relevant Courses:** Large Language Models, Statistical Learning, Statistical Computing, Big Data Analytics, Probability

**Bachelor of Technology in Computer Science**, GITAM University, India (GPA: 8.1/10) **Jun 2016 – Apr 2020**

**Relevant Courses:** Data Structures & Algorithms, Operating Systems, Machine Learning, Data Mining, Calculus

## TECHNICAL SKILLS

- **Programming Languages:** R, Python(NumPy, pandas, scikit-learn, PyTorch, PySpark), SQL, JavaScript
- **Data Science & Machine Learning:** Hypothesis Testing, Statistical Analysis, Time Series Forecasting, Causal Inference, Supervised & Unsupervised Learning, Feature Engineering, Model Evaluation, Interpretability, Dimensionality Reduction
- **Cloud & MLOps:** AWS (EC2, S3, SageMaker, Lambda, Redshift), Docker, Airflow, Flask APIs, Weights & Biases
- **Generative AI:** RAG, Fine Tuning, DeepSpeed, Langchain, CLIP, Stable Diffusion, GANs, Hugging Face, FAISS

## WORK EXPERIENCE

### AI Engineer Intern

**Jul 2025 – Sep 2025**

Schizophrenia & Psychosis Action Alliance

Alexandria, Virginia

- Scraped public housing data from multiple sources using Playwright and Javascript, generating semi-structured JSON.
- Built a **distributed scraping pipeline** with Celery and Redis, enabling parallel data extraction across all 50 U.S. states.
- Developed an agentic AI chatbot using **RAG** with FAISS for retrieval, powered by a fine tuned LLaMA 70B model using **LoRA**.
- Orchestrated the system with **LangChain** and LangSmith for evaluation, tracing, and performance optimization.

### Data Science Intern

**May 2024 – Aug 2024**

Welspot

Miami, Florida

- Conducted offline **experimentation** on AWS SageMaker, developing and comparing XGBoost and ARIMA for **financial risk forecasting**, leveraged S3 for data storage and achieved 88% validation accuracy with XGB as the final model.
- Applied **Bayesian** Optimization (Optuna) to tune XGBoost and used SHAP to interpret and regularize model predictions.
- Improved the F1 score from 0.78 to 0.89 by conducting **causal analysis** and statistical hypothesis testing (**t-tests, ANOVA**) on 20+ financial features and implementing time series feature engineering.

### Software Engineer | Analyst

**Oct 2020 – Jun 2023**

Tata Consultancy Services

Bengaluru, India

- Built **client similarity** and hybrid **recommendation models** using LightFM on transactional and operational data (**1.5 TB**) in PySpark and SQL improving Precision@5 by 15% and boosting client engagement.
- Extracted, processed, and analyzed **1M+** records from AWS Redshift using Python and SQL to generate CRM insights on client behaviors and lifecycle patterns, **informing engagement strategies** that improved retention and cross-sell rates.
- Designed and trained **autoencoder** model in TensorFlow on EC2 to generate latent embeddings from 1,100+ customer attributes, improving risk classification accuracy by 9% and enhancing downstream credit-risk analysis.
- **Optimized** and productionized ML **inference** pipeline on AWS Lambda via Flask REST APIs, serving 50K+ predictions daily with automated retraining and monitoring to sustain >99% reliability and stable model performance.
- Led **cross-functional** collaboration with application teams and data architects to design, train, and deploy scalable ML systems enabling **real-time inference** for fraud detection and business analytics across key platforms.

## PROJECTS

### Multimodal Sentiment Analysis System

**Mar 2025 – Mar 2025**

- Built a multimodal sentiment analysis system using the MVSA-Single dataset, combining **CLIP image embeddings with BERT text embeddings**, implementing fusion strategies and serving predictions via API, achieving 87% accuracy.

### Student Academic Performance Analysis

**Nov 2024 – Nov 2024**

- Analyzed student performance using **logistic regression in R with GLM** with **chi-square tests** and EDA, visualized insights via **ggplot2** and **RShiny** to support targeted educational interventions.

### Music Recommendation System

**Jun 2025 – Jun 2025**

- Built a music recommender (32K+ songs, 92K+ interactions) using KNN, RandomForest, **SASRec**, and **Two-Tower NCF**, achieving 88% HitRate@5, 83% NDCG@5 with CUDA, FAISS, cold-start handling, and **W&B** tracking.