Sai Jahnavi Gamalapati

Jersey City, NJ • jahnavi.sai1@gmail.com • + 1 (667) 403-0136 • LinkedIn • Github • website

PROFESSIONAL EXPERIENCE

Senior Data Scientist, Health Data Analytics Institute Inc | Remote May 2023 - Present

HDAI delivers predictive analytics and GenAI-powered insights through its HealthVision™ platform, helping health systems improve patient outcomes, reduce costs, and ease clinician burden.

- Developed a Logit-based RAG solution for clinical note chatbot, building filtering models with BM25 retrieval and MedSpaCy features to remove ~80% of irrelevant notes, reducing LLM inference costs and cutting latency from 3 minutes to 20 seconds.
- Developed a scalable feature engineering pipeline powering ~200 hospital risk models using regularized logistic regression (glmnet), boosting AUC by 1.5× and improving discharge planning accuracy.
- Collaborated with MLOps teams (JIRA/Confluence) to validate, monitor, and deploy predictive models in production, performing parity testing, partitioned validation, and feature distribution matching on client datasets to ensure fairness and robustness.
- Developed utility-driven evaluation framework to rank models by client use cases, prioritizing cost savings, risk reduction, and clinical relevance, which improved end-user utility by 10% on average compared to accuracy-only benchmarks.
- Engineered large-scale datasets using SQL (SAS, Python, R) and AWS ETL workflows; streamlined data pipelines that supported production deployment of patient risk stratification and outcomes forecasting models

Graduate consultant, Inside Airbnb | Stamford CT Sep 2022 - Dec 2022

- Built ETL pipelines in Python/SQL for ingestion of U.S. rental data into Tableau dashboards; reduced processing time by 1 hour and enabled advanced clustering and regression analysis to evaluate housing market impacts across 53 states.
- Investigated Airbnb's effect on U.S. rental markets by designing drill-down dashboards with licensing, frequency, and revenue metrics, supporting commercial evaluations at the city and state levels.

Data Science Intern, The Lego Group | Stamford CT Feb 2022 - Apr 2022

- Queried and analyzed 1B+ global sales records using SQL (sub-queries, window functions, hypothesis testing) to identify growth opportunities for girls first themes.
- Created Tableau dashboards revealing correlations (~0.54) between online and girls sales, informing DMA-level expansion strategies with projected sales growth up to 12.4%.

Senior Engineer, Qualcomm | Bengaluru Jan 2018 - Jan 2022

- Applied statistical analysis and optimization techniques across 20+ CPU design projects, automating workflows in Python/TCL to cut turnaround time by 30%+ and consistently meet aggressive low-power design targets

SKILLS

Languages & Databases: R, Python SQL, Git | MySQL, DuckDB, AWS S3

Machine Learning & Statistics: Regression (linear, logistic, regularized), Classification (Random Forest, XGBoost), Clustering (K-Means), Deep Learning (CNNs, RNNs, Transformers), Survival Models, Statistical Inference, A/B Testing

NLP & Generative AI: Embeddings, RAG pipelines, LLM fine-tuning, Prompt Engineering, AI Agents, Hugging Face, GPT APIs, LangChain, CLIP models) **Analytics & BI Tools**: Tableau, Power BI, SAS, MATLAB

EDUCATION

UCONN School of Business, **Master of Science in Business Analytics and Project Management**, Stamford May 2023

IIIT Hyderabad, **B.Tech and MS by Research in Electronics and Communication Engineering**, Hyderabad Dec 2017

PERSONAL PROJECTS

Table-Augmented Generation (TAG) System

- Designed a natural-language-to-SQL pipeline with DuckDB, LangChain, and OpenAI API, enabling non-technical stakeholders to self-query datasets; improved decision-making speed by delivering instant context-enriched product insights.

Basket-to-Basket Similarity (Customer Segmentation)

- Built segmentation models on retail basket data using association rule mining, cosine similarity, and clustering in Python, uncovering cross-purchase patterns (e.g., snack–beverage bundles) enabling targeted promotions and cross-sell campaigns.

Fine-Tuning LLMs for Human Preference Alignment (RLHF)

- Trained a preference classifier using DeBERTa-v3 + TRL on the Kaggle Chatbot Arena dataset, experimenting with cross-encoders, Bradley-Terry loss, and bias-mitigation features. Achieved ~0.64 validation accuracy in predicting user-preferred responses.

Retail Product Recommendation Engine

- Developed a hybrid recommendation system using the Amazon Product Reviews dataset as a proxy applying collaborative filtering (SVD) and LLM embeddings of product text to deliver personalized recommendations (~20% lift in top-5 recall).

Rewear AI (DreamAI Hackathon finalist for sustainable fashion idea)

- Prototyped an agentic fashion assistant with CLIP embeddings, OpenAI LLMs, and FastAPI; digitized wardrobes and integrated weather/calendar context to generate personalized outfit recommendations.

Study to Diminish Housing Insecurity

- Led a team of three and won 2nd position - $20K in Humana Mays HealthCare Analytics Case Competition for developing a predictive model with 900 features (AUC: 0.75, Disparity Score: 0.99) to characterize members with housing insecurity.