Sai Jahnavi Gamalapati

Jersey City, NJ, 07302 • jahnavi.sai1@gmail.com • + 1 (667) 403-0136 • LinkedIn • Github • GoogleScholar

Data Scientist with 2+ years of experience in analytics, machine learning, and Gen-AI, delivering data-driven solutions. Skilled in translating complex data into actionable insights for healthcare, retail and other data driven industries.

PROFESSIONAL EXPERIENCE

Health Data Analytics Institute Inc, **Senior Data Scientist**, Dedham, MA May 2023 - Present

- Reduced AI summaries bot inference time from 3 minutes to 20 seconds by creating a **lightweight, BM25-based document filtering framework** using bag-of-words and custom greedy algorithm, eliminating 85% of irrelevant notes and saving millions in processing costs.
- Developed and deployed an internal **Table Augmented Generation (TAG)** system using **DuckDB, AWS, Nova Lite**, to parse and query CSV and Parquet datasets. Enabled natural language access to performance metrics for 300+ models across multiple settings, reducing cross-team dependencies and fostering smoother collaboration.
- Designed **feature engineering framework** for in-hospital models (~200 models), allowing continuous updates to patient risk profiles in real time; improved model performance by **1.5× (AUC)** and enhanced discharge planning.
- Led **validation of model suites** across multiple client datasets, improving accuracy through targeted feature re-engineering and conducting parity testing to ensure consistent production deployment.
- Developed an R-based model selection utility aligned to **client-specific use cases**, moving beyond standard AUC/sensitivity metrics to **prioritize costlier events**, reducing unnecessary operational spend.
- Built decision tree models to identify cost-saving opportunities, achieving a **5% reduction in patient costs** and providing hospitals with targeted feature recommendations for close monitoring.

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

- Performed analysis of **1 billion Lego sales** in 2020/21 by applying advanced SQL queries such as sub-queries, joins and hypothesis testing methods such as **t-tests and ANOVA to sample** and explore growth opportunities for "girls first" theme sales.
- Created Tableau dashboards to gain insights into girl sales and online sales correlation (~0.54), leading to the identification of potential regions at the DMA level to expand girls' sales (with projections up-to 12.4%) through strategic recommendations.

Inside Airbnb, **Data Science Intern**, Stamford CT Sep 2022 - Dec 2022

- Investigated impact of Airbnb's on US housing rental markets. Built dashboards with a drill-down option to city level to view commercial metrics (% licensing, frequency of renting etc) and **monitor revenue to make commercial evaluations** over 53 states.
- Streamlined the **data processing pipeline** by creating a scalable Python script that automates pre-processing of country-level data before ingestion into Tableau. Reduced **processing time by up to 1 hour,** resulting in faster data delivery and analysis.

Qualcomm, **Senior Engineer**, Bengaluru IN Jan 2018 – Jan 2022

- Developed SQL queries to extract data from **GCP** environments to support ad-hoc and recurring analyses, collaborated with cross-functional teams to reduce cycle-time of chip using ML algorithms resulting in 15% reduction in PPA goals
- **Spearheaded AI initiative** by leading a team of 5, managed over 10+ projects by developing predictive models using machine learning algorithms to optimize performance and power efficiency of VLSI circuits, resulting in 15% reduction in PPA goals.

SKILLS

Languages: R, SQL, Python, MATLAB **ML:** TensorFlow, PyTorch, Numpy, Scipy, Pandas, Scikit-learn

Tools: Tableau Power BI, SAS EG/SAS studio **Others:** AWS, Azure, GitHub

PROJECTS

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. Recommended data-driven solutions to mitigate problem by performing sensitivity analysis for reducing claim costs

House insurance Customer Retention - Kangaroo Insurance

- Awarded a **bronze medal in Kaggle competition for developing a predictive model** (AUC: 0.69) to identify customers at risk of auto insurance policy cancellations. Identified credit score, sales channel, and Age as factors leading to policy cancellations. Further, offered solutions to increase retention rate in Virginia with a sales forecast of up to $56.6M.

Rewear AI (Dream AI Hackathon MA)

- Finalist at DreamAI Hackathon (LIKELION US, Cambridge, MA) for developing *Rewear AI*, an **AI-powered closet assistant** that digitizes wardrobes using **CLIP embeddings**, integrates weather/calendar data for outfit recommendations via OpenAI LLMs, and personalizes suggestions through lightweight memory. Built end-to-end prototype with **FastAPI, CLIP, and OpenAI**, encouraging sustainable fashion through outfit rotation and reuse.

PUBLICATIONS

- Detection of neovascularization in retinal images using **semi-supervised learning** (2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)
- **Multimodal Registration** of Retinal Images, NCVPRIPG 2017
- Radial peripapillary capillary **density measurement** using optical coherence tomography angiography in early glaucoma, Journal of Glaucoma 2017

ACHIEVEMENTS

- Humana Mays Healthcare Analytics Competition - **Silver**
- 35th New England Statistics Symposium – **Bronze**
- Traveler's BI&A LDP Analytics Case Competition – **Gold**
- JMP Discovery Summit – **Best Poster Award**
- UNICC Data Thinkathon – **Audience Choice Winner**
- Well4Tech Health Hackathon – **Gold**

EDUCATION

University of Connecticut, Stamford **May 2023**
Master of Science in Business Analytics and Project Management Stamford, CT
Relevant Coursework: Data Mining, Time series forecasting, Business Decision Modeling, Statistics, Data Science (NLP)

International Institute of Information Technology Hyderabad **July 2018**
B.Tech and MS by Research in Electronics and Communication Engineering **(Thesis)** Hyderabad
Relevant Coursework: Statistical Methods in AI, Computer Vision, Deep Learning