KARTHIK SUNDARAM SARAVANAN

https://www.linkedin.com/in/karthik-sundaram-s
https://github.com/karthik-sundaram
Phone: +1-(979)-739-7809
Email: karthik.sun95@gmail.com

EDUCATION

Texas A&M University, College Station, TX — May 2019
Master of Engineering, Industrial Engineering (Data science/Statistics concentration) — **GPA:4/4**

National Institute of Technology (NIT), Trichy, India — May 2017
Bachelor of Technology, Production/Industrial Engineering — **GPA: 4/4**

TOOLS/SOFTWARE SKILL SET

- Python, SQL, Docker, Azure DevOps, AWS, Azure, Snowflake, Gen AI/LLMs, R, Tableau, Linux/Bash, PySpark, GitHub, NLP, Deep Learning, Airflow, Logz.io, MLFLow

WORK EXPERIENCE

Lightbeam Health Inc., Atlanta GA — Jul 2022 - Present
Senior Data Scientist

Product Development:

- Built a global plug-and-play ML model for our top 3 products, reducing development time by over 75% per client product
- Reduced patient disengagement by 20% through targeted interventions using a time series LSTM with Attention model, enhancing intervention effectiveness, patient health outcomes, and profitability
- Scoped intent routing NLP agents for dynamically directing customer queries to appropriate internal chatbots, leveraging Azure Foundry for scalable deployment.
- Developed a Multi-Agent natural language assistant to help customers navigate the product UI, utilizing langgraph and pydantic function calling for context-aware interactions.
- Built a text-to-SQL module using fine-tuning and agentic RAG to improve contextual schema and column retrieval for accurate query generation.
- Led cross-functional collaboration and technical contributions to achieve Microsoft ISV certification for GenAI solutions, aligning with AI Cloud Partner Program standards
- Redesigned the Patient Impact Machine Learning Model to predict engagement, achieving over 40% lift improvement in high-risk deciles, enabling more effective healthcare interventions
- Guided BI engineers to integrate data science metrics into live Tableau dashboards, transitioning from static reports to real-time performance monitoring

MLOps (Level 0 to Level 2) & Migration Initiatives:

- Led migration of code from R to Python and AWS to Azure, creating reusable libraries and managing version control
- Architected transition to Docker from manual VM-based processes, enabling scalable, version-controlled deployment
- Designed an Azure DevOps pipeline for ML deployments, with automated VM management with PowerShell scripting
- Architected Logz.io integration for real-time monitoring of ML scoring processes, replacing delayed & dependency-heavy log retrieval - enabling drift detection, alerting and logs analytics
- Implemented a bias detection framework to assess fairness across protected variables, using metrics like Equal Opportunity and Predictive Parity within the 80% rule
- Created a drift monitoring system leveraging statistical tests (Kolmogorov-Smirnov, Chi-squared, Z-tests) for ensuring consistent data quality and model reliability

Jvion Inc., Atlanta GA — Jul 2019 - Jul 2022
Senior Data Scientist

- Delivered end-to-end ML solutions providing clinical predictions and recommendations for 10+ top healthcare clients, addressing All-Cause Readmissions, Sepsis, Alzheimer's, and 30-day re-admissions, resulting in yearly savings of $500k-$3M+ per client product measured through statistical tests and Interrupted Time Series Analysis
- Developed a comprehensive data repository by integrating diverse open datasets, using spatial analysis leading to improved ML model performance

- Reduced annual costs by $85K through optimizing API usage and eliminating redundant data processes.
- Boosted predictive accuracy for acute/dynamic disease models by integrating dynamic features, leveraging research literature, and aggregating IoT healthcare data, improving model lift by 2.5x
- Led experimental ML projects for breast cancer and COVID-19 prediction models
- Enhanced analytics with dynamic Tableau visualizations for key metrics like ROC and PR-AUC
- Developed scripts for EC2 instance management to trigger production runs through Airflow integration
- Presented technical projects in weekly forums, contributed to codebase improvement, conducted root cause analyses, and assisted in interviewing and mentoring new data scientists

Glassbeam Inc., Santa Clara, CA Sep 2018 - Dec 2018
Machine Learning Intern

- Analyzed large-scale unstructured IoT and time series data from MRI & CT scanners of leading healthcare clients
- Enhanced the product platform by developing machine learning models for failure prediction
- Replaced threshold-based anomaly detection with DBSCAN and Multivariate Gaussian algorithm
- Parsed raw log data into structured formats, enabling correlation analysis of machine failures and improving company-wide data utilization

Stanley, Black & Decker , Houston, TX May 2018 - Aug 2018
Data Science Intern

- Worked in Oil & Gas division dealing with large scale datasets across 10M data points across 25+ tables
- Defined a clear ERD schema for SAP tables & developed an algorithm to create Return on Investment (ROI) per product
- Reduced inventory forecast error by 15% using Time Series Forecasting using ARIMA & 'prophet' package in python
- Generated sales & marketing insights by finding the pipeline parts rented together using Market Basket Analysis
- Developed a web app for time series correlation visualization using Bokeh package in python
- Implemented an ML model to reduce product rental forecast error by 55%
- Created automated ETL pipelines, saving finance team man-hours

Technical University of Munich, **Germany** May 2016 - Aug 2016
Operations Research Intern – DAAD German Scholarship

- Worked on a project with Arla Foods, Denmark, focused on optimizing water usage in cheese manufacturing
- Conducted data analysis and visualization using Gantt charts and Tableau to optimize process scheduling
- Developed a scheduling framework using MILP (Mixed Integer Linear Programming) to maximize water reuse and minimize fresh water intake, reducing tank capacity by 23% and boosting profitability 300%

PUBLICATIONS

- "Designing an omnichannel closed loop green supply chain network adapting preferences of rational customers" – Springer India - *Sadhana- Academy Proceedings in Engineering Sciences, 2018*
- "Application of fuzzy quality function deployment for sustainable design of consumer electronics products: a case study" - Springer Verlag - *Clean Technologies and Environmental Policy, 2016*

INDEPENDENT PROJECTS

- Developed an agentic multi-modal chatbot utilizing LangGraph and ReactJsonAgent for agentic orchestration, LangSmith for monitoring and feedback, and deployed via Gradio with an accessible live API https://github.com/karthik-sundaram/MultiModal_MultiAgent_Chatbot
- Developed a sensitive data masking solution as a screening step before sending queries to OpenAI API, using a fine-tuned DistilBERT model deployed with AWS Lambda and API Gateway https://github.com/karthik-sundaram/aws-bert-llm-data-masking
- Developed a Multimodal RAG system for extracting, summarizing, and retrieving text, tables, and images from PDFs, with source citation included in responses, using Redis for storage, Chroma for embeddings, and Docker for deployment https://github.com/karthik-sundaram/MultimodalRAG

- Fine-tuned GPT-2 with RLHF (Reinforcement Learning with Human Feedback)/PPO, improving response quality by 64%, enhancing helpfulness, truthfulness, and safety
  https://github.com/karthik-sundaram/RLHF_PPO_Tuned_GPT2
- Developed a fine-tuned agentic RAG model for SQL generation from natural language using PEFT, QLoRA, and FAISS, achieving 91% accuracy in functionally equivalent SQL queries evaluated by an LLM.
  https://github.com/karthik-sundaram/FineTuned_AgenticRAG_Text_to_SQL
- Developed a Python package 'circularAlgos' (available on PyPi) for circular distributions, featuring PDF/CDF calculations, random generation, MLE-based mixture fitting, and probabilistic clustering.
  https://github.com/karthik-sundaram/circularAlgos

CERTIFICATIONS

- **Natural Language Processing Specialization** (4-course specialization) – DeepLearning.AI
  Courses Included*:* Attention Models, Sequence Models, Classification & Vector Spaces, Probabilistic Models
- **Generative AI with Large Language Models** - DeepLearning.AI, Amazon Web Services
- **Improving Real World RAG Systems: Key Challenges & Practical Solutions** - Analytics Vidhya
- **Neural Networks and Deep Learning** – Coursera
- **Microsoft Certified: Azure Data Scientist Associate –** Microsoft (Jan 2020-2022)