

PRAVEEN MANIMARAN

San Jose, CA | 408-913-5566 | praveenmanimaran61@gmail.com | [linkedin.com/in/praveen-manimaran](https://www.linkedin.com/in/praveen-manimaran) | praveenmanimaran.github.io

Professional Summary

Data scientist and applied ML developer with experience building data pipelines, developing AI/ML solutions, and working with LLMs in production settings. Skilled in Python, Spark, SQL, and cloud platforms such as AWS. Hands-on experience with prompt engineering, vector search, and deploying AI applications using modern frameworks. Passionate about applying data and AI to real-world challenges in fast-paced environments.

Education

University of California, San Diego

Sep 2023 - Mar 2025

M.S. Data Science (GPA: 4.0/4.0)

University of California, Santa Barbara

Sep 2021 - Sep 2023

B.S. Statistics and Data Science (GPA: 3.4/4.0)

Skills

Data Visualization: Tableau, Microsoft PowerBI, Seaborn, Matplotlib, Panel, Bokeh

Machine Learning & AI: TensorFlow, PyTorch, LangChain, Scikit-learn

Data Engineering & Cloud: AWS (Bedrock, SageMaker, Lambda, S3), Docker, Git, Azure (Synapse Analytics, Data Factory)

Big Data & Databases: SQL, PySpark, ChromaDB, Snowflake, PostgreSQL, MongoDB, DuckDB, Hadoop, Hive

Programming Python, R, C++, Java, SAS

Awards and Certifications

AWS Certified AI Practitioner

June 2025

Microsoft Certified: Azure Data Fundamentals

May 2025

Academic Excellence Award, UC San Diego – Halicioğlu Data Science Institute

March 2025

Experience

Lawrence Livermore National Laboratory

Sep 2024 - Present

Data Scientist

Livermore, CA

- Presented analytical findings to engineers and product stakeholders, informing **design improvements** and **anomaly mitigation strategies**
- Engineered scalable data workflows using **Python** and **SQL**, automating recurring tasks and reducing operational overhead across large sensor datasets
- Migrated analytical workloads from **PostgreSQL** to **DuckDB**, improving **query performance** and enabling faster experimentation and anomaly detection

Data Science Intern

June 2024 - Sep 2024

- Built an **interactive dashboard** using **Bokeh** and **Panel**, enabling test engineers to analyze anomalies in real-time
- Built and deployed **anomaly detection pipelines** over **3TB** of sensor data, applying **ML techniques** to identify irregular system behaviors

Ascend Leadership

Jan 2024 - Apr 2024

Data Management Intern

Remote

- Processed and analyzed **500+ database entries** to generate reports and improve **categorization accuracy**
- Automated **ETL pipelines** using **SQL** and **Python**, reducing manual data entry time by **20%**
- Created **Tableau dashboards** to uncover **demographic insights** in corporate leadership, driving strategic DEI conversations

AIByte

Oct 2023 - Dec 2023

Data Science Intern

Remote

- Developed a **real-time sign language recognition system** using **TensorFlow**, **MediaPipe**, and **OpenCV**, achieving **87% gesture classification accuracy**
- Designed a **Python-based interface** for live inference, improving accessibility and enabling real-time demos for **500+ users**

Amazon

June 2022 - Aug 2022

Area Manager Intern

Indianapolis, IN

- Analyzed and interpreted **150+ employees' performance and retention data** to optimize operational efficiency
- Built **AWS dashboards** to visualize **5 KPIs**, improving scheduling efficiency by **15%**

Projects

Fine-Tuning an Open Source LLM for SQL Query Generation

Mar 2024

- Fine-tuned **Llama** on a dataset of English-to-SQL query pairs using techniques like **LoRA** and parameter-efficient fine-tuning, reducing inference errors by **25%**

- Developed a preprocessing pipeline to clean, tokenize, and balance the dataset, improving model training efficiency and SQL query accuracy
- Deployed the fine-tuned model into a **RESTful API**, enabling seamless integration with applications

Cold Email Generator

Jan 2025

- Built an **AI-powered cold email generator** for consultancy firms, leveraging **Llama 3.1**, **ChromaDB**, and **LangChain** to automate personalized outreach and enhance client engagement
- Applied **advanced prompt engineering** strategies, including few-shot examples and role-based framing, to generate highly targeted, **natural-sounding emails** based on client profiles

Semantic Book Recommender

Oct 2024

- Developed a **content-based book recommendation system** using **LLMs**, **vector search**, and **zero-shot classification** to match users with books based on **semantic similarity** and **sentiment analysis**
- Built an **interactive Gradio dashboard** to allow users to input book descriptions and receive personalized recommendations, enhancing discoverability of relevant books

ECG Irregular Heartbeat Diagnosis

June 2024

- **Achieved 94% accuracy** in detecting irregular heartbeats using various machine learning models, including **Random Forests**, **1D CNNs**, and experimented with **LSTMs** to explore temporal sequence modeling
- Performed **multi-class classification** to diagnose heartbeat types, leveraging time-series preprocessing and **feature extraction techniques**