Pavithra Kandhasamy Selvaraj

📞 +1-945-393-3954 ✉ pavithrakselvaraj@gmail.com in Pavithra Selvaraj 🌐 Pavithra Portfolio

Education

University of California Santa Cruz	**Aug 2025 - Present**
Master's in Computer Science and Engineering	California, USA
Coimbatore Institute of Technology	**Jul 2016 - Sep 2020**
Bachelor's in Computer Science and Engineering (CGPA - 3.7/4.0)	Coimbatore, India

Technical Skills

Programming: Python, SQL, Shell Scripting, HTML, CSS.

AI & Data Science: Regression, Classification, Clustering, Time Series Forecasting, Statistical Modeling, Data Analysis and Visualization, Deep Learning, Gen AI, Pandas, Numpy, Sklearn, Matplotlib, Seaborn, PySpark, PyTorch, Keras, Tensorflow.

MLOps and Cloud: Google Cloud (BigQuery, VertexAI, PostgreSQL, Looker), Mlflow, Terraform, Tekton, Docker.

Other Tools: Git, Excel, Flask, Postman, Splunk.

Professional Experience

Tiger Analytics — **Dec 2024 - Aug 2025**
Data Scientist — Chennai, India

* Built ML attributes for a Credit Reporting Bureau using **Python** and **BigQuery** to process and transform over **16k+** census variables across **15+** large-scale tables; performing feature engineering including variable mapping, aggregation and normalization.
* Developed and deployed predictive models for attribute creation like **Logistic Regression, XGBoost and Decision Trees** using **PySpark** on **30GB+** records, with post-attribute analysis to ensure feature stability across historical and current periods.
* Supported the development of a nationwide **Customer Segmentation** model using **15 million** census records to cluster and classify the Australian population by neighborhood characteristics, enabling market insights aligned with evolving population.
* Streamlined large-scale model training and **feature processing pipelines** on GCP, enabling **distributed processing** of census data, ensuring automated data ingestion and transformation workflows to integrate new census data into downstream models.

Ford Motor Private Limited — **Jan 2022 - Nov 2024**
Machine Learning Engineer — Chennai, India

* Built **Random Forest** model to optimize Collections by identifying the effective communication channels and contact times for delinquent customers, achieving **73%** accuracy via preprocessing, feature selection and tuning to reduce contact frequency by **66%**.
* Implemented **Time Series Forecasting** models on historical repayment and delinquency data to predict customers at risk of default, enabling proactive collections strategies and reducing late-payment cases by improving early intervention accuracy.
* Migrated the application to **Google Cloud Platform**, leveraging **Cloud Run**, **Cloud Build** and **Cloud Scheduler**, resulting in a **45K USD** annual cost reduction through seamless database migration from On-Prem **MS SQL Server** to **PostgreSQL**.
* Ensured consistent cloud deployments across environments using **Terraform**, reducing cloud service provisioning time by **67%** and implemented a **Tekton**-based **CI/CD** pipeline to streamline the process from source code push to production deployment.

Ford Motor Private Limited — **July 2020 - Dec 2021**
Junior Machine Learning Engineer — Chennai, India

* Developed a scalable **Python** application to manage outreach of **70k** past-due customers, utilizing a rules-based system to determine optimal outreach strategy based on frequency limits, time restrictions and state-specific regulations reducing manual efforts.
* Stored and managed over **10 million** customer contact records by building **ETL pipelines** integrating data from **5+ systems** in an on-prem **MS SQL Server** ensuring data quality for downstream ML models that predicted repayment likelihood.
* Experimented with models like **Random Forest, SVC, XGB and Neural Networks** on customer repayment datasets to evaluate predictive performance, comparing precision–recall tradeoffs to identify approaches for enhancing the rules-based system.

Academic Projects

* Developed an intelligent **Symbolic Reasoning** system at the **AIEA Lab, UCSC**, integrating **Large Language Models** with **Prolog** to accurately translate complex natural language queries into symbolic representations for logical inference.
* Re-implemented the **Logical LLM framework** with modular components (classifier, solvers, refinement, interpreter), incorporating an iterative self-refinement loop to significantly improve reasoning accuracy on diverse and complex queries.
* Enhanced system scalability and modularity by integrating **LangChain** (RAG pipelines with Chromadb) and migrating to **LangGraph** for structured agent workflows, improving performance on direct, indirect, and ambiguous natural language problems.

Hackathon Projects

Question Your Dataset using LLM {Python, VertexAI, PaLM 2, Pandas, PySpark, Langchain, HTML, CSS, Flask} — **Aug 2024**
· Implemented an **LLM-powered system** that enables users to interact with datasets using natural language questions, identifying patterns, generating insights and answering complex queries to make data exploration more accessible for non-technical users.

Meeting to Article Content Generation {Python, VertexAI, Gemini-Pro, HTML, CSS, CloudBuild, GCS, Flask} — **Mar 2024**
· Designed an automated **meeting-to-article** generator that transcribes meetings using **Google Cloud Speech-to-Text**, processes and converts them into structured articles with speaker-aware summarization, streamlining internal communication workflows.

Multilanguage Video Summarization {VertexAI, Gemini-Pro, HTML, CSS, CloudBuild, Flask} — **Jan 2024**
· Built an LLM-powered video summarization tool using **Gemini Pro** and **Google Translate** that generates concise, multi-lingual video summaries and answers content-related questions, enhancing accessibility and user experience for lengthy video content

Accessing Knowledge Articles using LLM {Python, Langchain, RAG, VertexAI, PaLM 2, Gradio} — **Oct 2023**
· Engineered a multilingual, **LLM-powered** knowledge access system, enabling intuitive search and retrieval of relevant information from lengthy documents, improving accessibility, overcoming language barriers and delivering faster and relevant results.