Pavithra Kandhasamy Selvaraj

+1-945-393-3954 pavithrakselvaraj@gmail.com Pavithra Selvaraj Pavithra Portfolio

Education

University of California Santa Cruz **Aug 2025 - June 2027**
Master's in Computer Science and Engineering (CGPA - 3.85/4.0) California, USA

Coursework: Analysis of Algorithms, Neural Computation, Foundations of Data Science, Programming Language.

Coimbatore Institute of Technology **Jul 2016 - Sep 2020**
Bachelor's in Computer Science and Engineering (CGPA - 3.73/4.0) Coimbatore, India

Technical Skills

Programming: Python, SQL, Shell Scripting, HTML, CSS.
AI & Data Science: Regression, Classification, Clustering, Time Series Forecasting, Statistical Modeling, Predictive Modeling, Data Analysis and Visualization, Deep Learning, Gen AI, Pandas, Scikit-learn, Matplotlib, Seaborn, PySpark, PyTorch, Keras, Tensorflow.
MLOps and Cloud: Google Cloud (BigQuery, Looker, VertexAI, AutoML, PostgreSQL), Mlflow, Terraform, Tekton, Docker.
Other Tools: Splunk, Git, Excel, Flask, Postman.

Professional Experience

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

- **Engineered end-to-end machine learning attributes** for a Credit Reporting Bureau, leveraging **Python** and **BigQuery** to process and transform **16K+** census variables across **15+** large-scale tables, improving data readiness for predictive modeling.
- Developed large-scale predictive models (**Logistic Regression, XGBoost, Decision Trees**) using **PySpark** on **30GB+ data**, optimizing attribute creation and performing post-attribute stability analysis to ensure consistency across time periods.
- **Enabled data driven customer targeting** by developing **Customer Segmentation tool** using **Clustering techniques (K-Means and Hierarchical Clustering)** on **20M+** census records, enabling clients to refine campaign strategies.

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

- **Optimized collections efficiency by 66%** by building a **Random Forest** model to identify the most effective communication channels and times for delinquent customers, achieving **73%** accuracy through feature engineering and hyperparameter tuning.
- **Reduced late-payment cases by 24%** through proactive intervention by implementing **Time Series Forecasting** model on historical repayment and delinquency data, improving early-risk identification accuracy and enabling targeted collection strategies.
- **Lowered infrastructure costs by $45K** annually by migrating application to **Google Cloud Platform** using **Cloud Run, Cloud Build and Cloud Schedulers** and database migration from on-prem MS SQL Server to **PostgreSQL**
- **Accelerated cloud deployment** cycles by automating infrastructure provisioning with **Terraform** and **Tekton CI/CD pipeline** to ensure reliable multi-environment deployments and integrated **Splunk** monitoring to resolve application issues within 24 hours.

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

- **Reduced manual outreach efforts** by automating engagement for **70K+** past-due customers through a scalable **Python** application, implementing a rules-based system to determine optimal contact frequency and time with state-specific regulations.
- **Strengthened data-driven decision-making** by consolidating data from **5+ internal systems** into **MS SQL Server** and delivering insights through **Looker** and Python-based visualizations (**Seaborn & Matplotlib**) to optimize collections strategy.
- Enhanced outreach prediction framework by experimenting with Classic ML models and **Neural Networks** utilizing **PyTorch and Tensorflow**, analyzing **precision–recall** tradeoffs to identify effective algorithms for refining the rules-based outreach strategy.

Academic Project and Experience

- Serving as **Graduate Tutor** for course **Applied ML - Deep Learning** and **Grader** for Database Systems and Embedded OS.
- 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.
- 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.