

Sukumar Govindaraj

Plano, TX | (682) 405-4122 | govindaraj.sukumar@outlook.com | [GitHub](#) | [LinkedIn](#) | [Portfolio](#)

Professional Summary

Machine Learning Engineer with 4+ years' experience in data engineering and distributed systems. Skilled in building secure, scalable pipelines and APIs using Python, SQL, and Spark across GCP and Azure. Experienced in software engineering, data modeling, audit-compliant design, and end to end MLOps.

Skills

Technical Skills: Anomaly Detection System, Recommendation System, Predictive Modeling, NLP, Software Development Life cycle, Data Modeling
Programming Languages: Python, Java, JavaScript, SQL, GraphQL, C++, Shell scripting
Frameworks and Tools: PyTorch, PySpark, Scikit-learn, MLflow, Copilot, Mule, Talend, DVC, Node.js, React, Fast API, Tableau, Power BI
Cloud Platforms: **GCP** (BigQuery, Cloud Storage, IAM), **AWS** (EC2, Redshift, SNS, SQS), **Azure** (ML Studio, Synapse, Blob, Data Factory, Monitor)
Professional Skills: Problem-solving, Innovation, Collaboration, Leadership, Critical Thinking, Attention to Detail, Committed,

Professional Experience

UT Arlington – Shimadzu Research Center

10/ 2023 – Present

Machine Learning Engineer | Predictive Maintenance - Anomaly Detection System

- **AI Chatbot** - Developed an LLM-RAG based internal chatbot as part of MCP server by utilizing Vertex AI to assist researchers troubleshoot lab tools during experimentation, reducing engineering time and dependency on manual support.
- **Model development** - Collaborated with research team in building a system to monitor lab equipment's health in real time using machine learning to help the team to detect issues early and reduce unexpected tool failures.
- **ML Operations** - Automated model training and updating to ensure predictions stay accurate without manual checks.
- **Reporting and Data Visualization** - Created easy to understand dashboards and alert systems to support lab staff and shared weekly reports with the leadership, decision making and reduced unplanned equipment downtime by over 20% through early anomaly detection.
- **Skills** - Predictive modeling, Google Vertex AI, PyTorch, Neural Networks, LLM, RAG, NLP, Experimentation, MLflow, CI/CD, ML lifecycle, A/B Testing, Excel.

Victoria's Secret & Co

04/ 2022 – 07/ 2023

Data Engineer | Supply Chain Management - Inventory Forecasting System

- **Requirement Gathering** - Collaborated with a planning team to organize sales and inventory data to help predict inventory in stores and warehouses.
- **ETL Pipelining** - Created scheduled workflows that kept data up to date by fetching, cleaning and delivering accurate data to reporting dashboards, helping teams make faster and better decisions.
- **Collaboration** - Worked closely with business users to improve report quality and optimized data pipelines to reduce report delivery time by 70%.
- **Skills** - ETL tools, Mule Soft, Oracle ERP, Control-M, PostgreSQL, PySpark, AWS S3, Lambda, Git, Jira, Ad Hoc, Delta Batchting, SAP Master Data.

Web Wings Global Solutions

01/ 2021 – 03/ 2022

Data Engineer | Customer Analytics

- **Integration** – Built a GraphQL layer on Salesforce CRM to enable single-query access to customer data, simplifying API usage across 500+ accounts.
- **Data Modeling**- Designed structured data layers using medallion architecture to make dashboards more reliable for business users to interpret.
- **Quality Testing** – Orchestrated data pipelines using Control-M with Spark and SQL, implemented unit tests validation, and supported deployment.
- **Skills**: Salesforce CRM, GraphQL, AWS, Redshift, DMS, S3, Databricks, Oracle SQL, Version Control, Linux, Agile Scrum, KPIs, Relational Database.

Publications and Presentations

- "Transfer Matrix 3: A Web-based Simulation Tool for Optics and Plasmonic Structures", Presented at Nanofabrication Research Symposium, 2025
- "Survival Analysis Modeling for Predictive Maintenance for Lithography & Thin-Film Deposition Systems", Presented at UTA Research Symposium, 2024

Projects

- **Genomics Data Pipeline - Variant Detection, Expression Analysis in Genes**
Built a DNA/RNA data pipeline to identify gene variants and expression patterns using Bowtie2, SAMtools, and bcftools. Automated processing with Python and GCP services like BigQuery and Cloud Storage.
- **GenAI Coffee Shop Assistant - LLM, RAG based Agentic Chat Application**
Built a multi-agent conversational assistant using MCP, RAG, and LLMs to handle ordering, personalized recommendations and FAQs. Integrated the system into a React Native app with Python backend, Prompt Engineering, Pinecone, and Firebase.
- **Credit Risk Modeling – PD & LGD Prediction on Financial data**
Built ML models to estimate default probability and loss severity using financial and credit bureau data. Applied feature engineering, multi variable regression techniques, and model tuning to improve accuracy and reduce errors.

Education

- Master of Science in Applied Statistics and Data Science
- Bachelor of Engineering in Computer Science

University of Texas at Arlington, TX — GPA: 4.0/4.0
Sahyadri College of Engineering, India — GPA: 3.5/4.0