

KUSHWANTH KUMAR

☎ +1 (929) 361-8554 | ✉ kkushwanth77@gmail.com | 🌐 [github](#)

PROFESSIONAL SUMMARY

Dynamic, results-driven Data Scientist and **Software** developer with **over five years of experience** in building scalable ML pipelines, fine-tuning LLMs, and developing AI-driven solutions using MLOps techniques. Skilled in cloud platforms (AWS, GCP, Azure), containerization (Docker, Kubernetes), and AI frameworks (PyTorch, TensorFlow, LangChain). Adept at NLP, multi-agentic RAG, and Conversational AI with a strong focus on innovation, efficiency, and cross-functional collaboration. Successfully built and deployed a full-stack AI application from **scratch**, demonstrating end-to-end ownership and technical excellence.

SKILLS

Programming Languages: Python, R, SQL, Java, PySpark, Unix, Linux, Golang

Cloud Platforms & DevOps: AWS, Azure, GCP, Docker, Kubernetes (AKS, EKS), Git, GitHub, SSH, Harness, Jenkins, Azure AI Studio, A/B Testing, Apache Hadoop, Terraform

Big Data Technologies: Snowflake, Apache Airflow, Azure Data Factory (ADF), Databricks, Oracle SQL, MongoDB, Async PostgreSQL, SQL Server, Neo4j, Kafka, Vector Databases (ChromaDB, FAISS, Weaviate, Pinecone), Azure Cosmos DB, RDF Graphs (SPARQL, Gremlin, Cypher)

ML & AI: PyTorch, TensorFlow, Scikit-learn, Supervised learning models, Unsupervised learning models, Time Series Forecasting, SVM, Ensemble Techniques, Feature Engineering, Clustering, Model Deployment, Model Monitoring, Drift Detection, RLHF, AutoGen, Flair, BERT, Deep Learning

Gen AI & NLP: Fine-tuning LLMs, multi agentic RAG, CAG, GraphRAG, Re-ranking, Hugging Face, Prompt Engineering (CoT, ToT), LangChain, LangGraph, LlamaIndex, Haystack, Whisper AI, ASR, OpenAI GPT, Claude, Vertex AI, AI Governance and Ethics, Spacy, n8n, AI Workflow Automation, splunk, Autogen

Software Frameworks & Tools: Flask, FastAPI, SciPy, Redis, gRPC, NumPy, Pandas, MLOps/LLMOps (Airflow, MLflow)

Data Visualization & Business Analytics: Tableau, Power BI, ERP Systems, Supply Chain Analysis, Dashboard Generation

Development: Agile (Scrum), CI/CD (Harness, Jenkins), Pytest, Jira

WORK EXPERIENCE

Marsh Insurance

Software Engineer -ML/AI

Remote

Jan 2024 – Present

- Designed scalable **MLOps pipelines** using **Vertex AI Pipelines**, **Cloud Functions**, and **Cloud Storage** for ingestion, training, and deployment.
- Developed and deployed **multi-agentic RAG systems** using **LangChain**, **ChromaDB**, and **Vertex AI Search**.
- Built LLM-based applications with **FastAPI**, integrating with **Gemini**, **Claude**, and **OpenAI GPT** for enterprise use cases.
- Orchestrated model lifecycle with **Kubeflow**, **Cloud Build**, **Artifact Registry**, and CI/CD workflows via **Terraform**.
- Managed real-time model performance and drift with **Vertex AI Model Monitoring** and custom metrics using **Prometheus** and **Grafana**.
- Developed supervised fine-tuning workflows for LLMs, optimizing performance using ToT and self-consistency strategies.
- Utilized **BigQuery**, **Pub/Sub**, and **Dataflow** to engineer streaming analytics pipelines.
- Deployed AI apps in secure containers via **GKE**, applying best practices for distributed inference and scaling.
- Collaborated with cross-functional teams to define GenAI tooling architecture for high-throughput production tasks.
- Instrumented GCP IAM, Secret Manager, and audit logging for GenAI-specific governance and compliance.

Kimberly-Clark, Tredence

Data Engineer

Bangalore, India

June 2021 – Aug 2023

- Developed and scheduled robust **Azure Data Factory (ADF)** pipelines for ingesting and transforming large-scale retail data.
- Integrated **Azure Data Lake**, **Blob Storage**, and **SQL Pools** to optimize data storage and retrieval across analytics teams.
- Used **Databricks on Azure** for data cleansing, transformation, and batch scoring across distributed retail systems.
- Built and maintained **Power BI dashboards** to monitor pipeline performance and generate sales insights for stakeholders.
- Implemented real-time ETL frameworks with **Azure Event Hubs** and **Stream Analytics** to feed live telemetry to business systems.
- Collaborated with data modelers to standardize schema definitions using **Azure Purview** for data cataloging and governance.
- Automated pipeline deployments using **Terraform**, **Azure DevOps**, and **Git** for CI/CD across environments.
- Orchestrated GenAI reporting flows by embedding **FastAPI** endpoints into Azure Functions and AKS clusters.
- Built internal tools for schema discovery and config-driven pipeline generation to reduce manual intervention.
- Liaised with compliance teams to implement secure access controls via **RBAC**, **Managed Identity**, and **Key Vault**.

Union Bank of India

Data Scientist

Hyderabad, India

May 2020 - Jun 2021

- Designed and deployed **financial forecasting models** using **Python**, **scikit-learn**, and **Time Series** techniques.
- Built ML microservices for anomaly detection and risk analytics using **Flask**, **XGBoost**, and REST APIs.
- Constructed **time series pipelines** using **Airflow** and batch schedulers for model refresh and audit reporting.
- Developed high-accuracy fraud detection models leveraging **unsupervised clustering** and transaction-based heuristics.
- Used **BERT-based classifiers** to tag compliance documents and enhance audit readiness.
- Collaborated with analysts to create custom forecasting notebooks using **Jupyter** and model interpretability layers.
- Automated model deployment workflows using **Docker**, **Kubernetes (on-prem)**, and bash scripting.
- Enabled rule-based filters and parameter tuning overlays to support iterative experimentation.
- Worked with finance teams to build data marts and pre-aggregated features to accelerate analytics workflows.
- Implemented **LangChain**-driven assistants to summarize model outcomes and error analysis.

EDUCATION

Webster University <i>Masters in Information Technology; GPA: 3.93/4.0</i> National Institute of technology, Surat <i>Bachelors in Computer Science; GPA: 8.02/10.0</i>	San Antonio, TX <i>September 2023 - December 2024</i> Surat, Gujarat <i>August 2018 - May 2021</i>
--	---

PERSONAL PROJECTS

Leaf Disease Detection Using CNN <i>Python, Flask, Oracle SQL</i> • Developed a convolutional neural network model with a Flask-based web application for detecting leaf diseases, and integrated Oracle SQL for user data management.	<i>2021</i>
Autonomous Object Detection for Industrial Applications <i>OpenCV, YOLOv5</i> • Built a YOLOv5-based object detection system for real-time industrial tool identification, achieving 96% accuracy with multi-camera integration.	<i>2022</i>

CERTIFICATIONS

MLOps with VertexAI <i>VertexAI, GCP, MLOps</i> • Completed a hands-on certification covering the deployment of scalable machine learning pipelines using Vertex AI, including CI/CD workflows, model monitoring, and feature store integration.	<i>Apr 2025</i>
--	-----------------