SALMAN MOHAMMED

USA | Portfolio | 646-513-8347 | Salmanmohammed2018@gmail.com | GitHub

EXPERIENCE:

Smartrip Inc., Remote | **AI Intern, AI Travel Companion Platform** *Mar 2025 - Aug 2025*

- Architected voice-enabled AI travel assistant leveraging **Gemini LLM (Vertex AI)** and **ElevenLabs TTS**, generating personalized itineraries through natural language conversation—**reducing manual planning time by 70% for 500+ beta users**.
- Optimized **LLM prompt engineering** workflows for itinerary generation, **reducing token usage by 30%** through system prompt refinement and caching strategies while maintaining **4.2/5 user satisfaction scores**.
- Engineered **multi-agent recommendation pipeline** integrating **6+ travel APIs** (flights, hotels, activities) with **Supabase vector database** for preference embeddings, implementing fallback logic that **improved booking success rate by 40%**.
- Deployed **containerized FastAPI microservices** on **GCP (Cloud Run)** with **CI/CD pipelines via GitHub Actions**; instrumented **PostHog telemetry tracking 15+ user events**, enabling data-driven iteration with **<24hr deployment cycles**.

TechStack: *Python, FastAPI, Gemini LLM, Vertex AI, ElevenLabs, Supabase, Docker, GCP (Cloud Run), GitHub Actions, PostHog, Stripe.*

Dataevolve Solutions, Hyderabad, India | **Machine Learning Engineer, Digi Yatra** *Sep 2022 - Jun 2023*

- Deployed **facial recognition and boarding pass OCR system** for Digi Yatra (India's national airport program), processing **10K+ passengers daily across 8 airports** with **98% accuracy** and **<2s latency** per verification.
- Built **PyTorch OCR model** for boarding pass scanning (name/flight/seat extraction) using **transfer learning (ResNet-50 backbone)** and **ensemble methods**, achieving **94% field-level accuracy—18% improvement over baseline CNN**.
- Engineered **YOLOv7-based barcode detection microservice** handling **500 QPS**, containerized with **Docker** and deployed on **AWS EKS (3-node cluster)** via **KServe**, enabling **horizontal autoscaling to 15 replicas** during peak traffic.
- Automated **ML model deployment** via **AWS API Gateway + Lambda CI/CD pipeline (GitHub Actions)**, **reducing release cycle from 3 days to 4 hours** and achieving **99.7% API uptime across 12 production endpoints**.
- Scaled system to **8 airports** by deploying containerized services on **AWS ECS (Application Load Balancer + Target Groups)**, implementing **TLS encryption** and **IAM role-based policies**; established **blue-green deployment** environments using Lambda, achieving **zero-downtime releases**.

TechStack: *Python, PyTorch, YOLOv7, ResNet-50, Docker, Kubernetes, AWS (EKS, ECS, ECR, Lambda, API Gateway, CloudWatch), KServe, GitHub Actions, RabbitMQ, Amazon MQ, TLS.*

iNeuron.ai, Remote | **Data Science Intern** *Nov 2021 - Aug 2022*

- Developed **Random Forest regression model** for bike-sharing demand prediction (**MAE: 12.3, 80% improvement over baseline**) and **SVM classifier** for forest fire detection (**85% accuracy, F1: 0.82**); deployed to **Heroku via Docker** and built **Plotly dashboards** for stakeholder presentations.

TechStack: *Python, LangChain, ChromaDB, sentence-transformers, Streamlit, Docker.*

PROJECTS:

Health Risk Assessment Portal - Capstone Project

- Engineered **full-stack health risk prediction platform** (FastAPI backend, React frontend) with **3 Random Forest classifiers** achieving **87% average ROC-AUC** on synthetic EHR data (Synthea); deployed on **Vercel with Docker**, serving **1K+ predictions in pilot phase**.
- Implemented **feature engineering pipeline (30+ clinical features)** and **SHAP explainability module**, providing clinicians with interpretable predictions and identifying **top 5 risk factors per condition**.

TechStack: *Python, FastAPI, React, scikit-learn (Random Forest), SHAP, Pandas, Synthea, Docker, Vercel.*

Dobbs Agent - AI Customer Service Chatbot

- Built **production-ready AI customer service chatbot** for Dobbs Tire & Auto Centers with **React + TypeScript frontend** and **FastAPI backend**, implementing **keyword-based FAQ search** with **scheduling intent detection** across **15+ FAQ topics** covering hours, locations, tire brands, and services for **50+ store locations**.
- Integrated **ElevenLabs TTS API** for natural voice responses and **Web Speech API** for hands-free voice input; developed **intelligent appointment lead capture system** storing submissions in **JSON file storage** with **REST endpoints** for lead management and retrieval.

TechStack: *Python, FastAPI, React, TypeScript, TailwindCSS, Shadcn UI, TanStack Query, ElevenLabs API, Web Speech API, Uvicorn.*

TECHNICAL KNOWLEDGE:

Languages: Python, SQL, JavaScript, Bash, C, MongoDB, PostgreSQL, Redis, FAISS, Pinecone.
ML/AI & Frameworks: PyTorch, Deep Learning, TensorFlow, Hugging Face Transformers, LangChain, scikit-learn, Large Language Models (LLMs), RAG Systems, Computer Vision, NLP, Model Optimization, Prompt Engineering.
Cloud & DevOps: Jenkins, Kubernetes, Docker, AWS, ElasticSearch, Maven, Gradle, Docker, SSH, ELK, IaaS, SaaS, GCP.
Tools & Libraries: Pandas, NumPy, FastAPI, Flask, React, Plotly, Git, PostHog, Supabase, Stripe.

EDUCATION:

Saint Louis University, St. Louis, MO | Masters in Analytics *Aug 2023 - May 2025*

Jawaharlal Nehru Technological University, Hyderabad, India | Bachelor of Technology *Aug 2017 - May 2021*