Mohamed Yusuf

Kodungaiyur, Chennai - 600118
+91 6381810492 mohamedyusuf2528@gmail.com
Portfolio LinkedIn GitHub

Profile Summary

Data Scientist with 2 years of hands-on experience in building Machine Learning and Deep Learning solutions. Specialized in predictive modeling, transformer-based generative AI, and deploying scalable models in cloud environments (AWS Azure). Skilled in delivering end-to-end AI solutions that solve real-world problems and drive business value.

Skilled at converting complex data into insights, building scalable AI solutions including LLMs, RAG, and deploying them efficiently for real-world impact.

Skills

Strategic: Machine Learning, Deep Learning, Predictive Modelling, Forecasting, Pattern Recognition, ANN, CNN, VGG, RNN, LSTM, Optimization, YOLO, GAN, LLM, RAG, LangChain, LangGraph, MLOps, MLflow, OCR, Chatbot, OpenAI, MediaPipe

Languages: Python, SQL

Deployment: Flask, FastAPI, Streamlit

Databases: MySQL, PostgreSQL

Cloud: AWS, Azure

Visualization: Tableau, Matplotlib, Seaborn

Professional Experience

Data Scientist 2023 – Present
Colan Infotech Pvt. Ltd. *Chennai*

- Worked on ML/DL/LLM model development and deployment
- Developed advanced RAG systems including Multi-Modal RAG and Text-to-SQL bots
- Managed model pipelines on AWS and Azure platforms

Projects

Multi-Modal RAG Chatbot | *LangChain, CLIP, OpenAI, VectorDB, SQL, GitHub*

- Built a multi-modal RAG chatbot to process PDFs, including text, tables, and images, as well as handle structured quantitative data
- Extracted PDF contents and applied title-based chunking and summary generation for text, tables, and images
- Implemented a multi-vector retriever using Qdrant VectorDB with embeddings from OpenAI's large text embedding model; stored original documents in a docstore for retrieval during inference
- Parallelly built a RAG chatbot to handle quantitative queries with SQL generation
- Utilized GPT model with OutputParser agent to correct SQL syntax issues if happens
- Designed a custom schema creation method for the database
- Enabled accurate, context-aware responses for complex numerical data
- Integrated LangChain QA chain to provide document-aware answers

Custom AI Avatar Bot | *Transformers, HuggingFace, ML, SQL, Flask, XAMPP*

- Developed avatar chatbot interface for dynamic DB-driven Q&A
- Used Transformer model to generate SQL queries from user intent
- Routed user questions to relevant DB using ML-based query classifier
- Hosted data on XAMPP server with Flask API and custom UI with Avatar

Hospital Staff Salary Prediction | *XGBoost, MLflow, AWS*

- Achieved 89% regression accuracy using XGBoost on hospital salary data
- Performed data cleaning, transformation, and exploratory analysis
- Implemented CI/CD pipeline using SageMaker, S3, Lambda, Step Functions, and CodePipeline
- Tracked experiments with MLflow and deployed model as a SageMaker endpoint

AI Bill Review | *BERT, Azure, Streamlit, Flask, Threading, WebHook*

- Fine-tuned BERT to classify invoice descriptions into legal categories (GNCCode)
- Implemented rule-based workflow to trigger actions based on prediction outcomes
- Deployed the model via Azure VM with Flask and Blob integration for custom class training
- Configured a webhook to notify the team and user upon completion of custom class model training
- Resolved auto-restart issues using a systemd service file for continuous deployment

Home Automation | *MediaPipe, OpenCV, Deep Learning, Flask*

- Developed 3D hand and body pose detection using camera feed and MediaPipe
- Collected pose data using OpenCV and created labeled training datasets
- Trained a Deep Learning model achieving 94% accuracy in posture classification
- Deployed API via Flask integrating pose prediction and automation trigger

POC Projects

Text Clustering & Topic Generation | *SentenceTransformers, DBSCAN, T5*

- Generated sentence embeddings using all-mpnet-base-v2
- Used DBSCAN for clustering textual data
- Created topic summaries using T5
- Automated NLP pipeline for insight generation

Text Image Classification for OCR | *CNN, TensorFlow, Selenium*

- Built CNN model to classify images
- Automated dataset creation via Selenium
- Used early stopping and tuning for optimization
- Applied image augmentation for better accuracy

Education

Bachelor of Science in Mathematics 2018 – 2021
Sadakathullah Appa College, MS University *Tirunelveli, Tamil Nadu*

Certifications

MLOps with AWS Bootcamp - Zero to Hero Series – Udemy

Master Data Science Program – GUVI