Kartheek Yakkala

+1 9136369562

kartheekyakkala.se@gmail.com ◇ Portfolio ◇ LinkedIn ◇ Blogs ◇ Github

Education

Master's in Computer Science
University of Central Missouri — Aug 2023 - Dec 2024

Bachelor of Technology, Electronics and Communications Engineering
Jawaharlal Nehru Technological University Hyderabad — Aug 2017 - Jul 2021

Skills

Programming Languages : Python, Java, SQL, R, Bash, C, C#, Html, CSS, JavaScript.
Technologies : Machine Learning, Natural Language Processing, Large Language Models(LLMs), Generative AI, Retrieval-Augmented Generation (RAG), Prompt Engineering, Fine tuning, Deep Learing, Data analysis, LoRA, QLoRA.
DevOps and Cloud : Continuous Integration and Deployment, Docker, Kubernetes, Helm charts, Jenkins, Github Actions, Azure Functions, Google Cloud run, Azure File share, Azure Log Analytics, Azure Kubernetes Services, Azure Container Registry, Docker Compose, Git, GithHub, Azure Machine Learning, AWS Sage Maker, Google Vertex AI, Azure Data Factory, Azure OpenAI, Anthropic, Meta Llama, Google Gemini.
Frameworks and Databases : PyTorch, TensorFlow, sci-kit-learn, LangChain, React, MongoDb, MySql, ChromaDB, FAISS, Pinecone, BigQuery, AlloyDB, Pandas, Sendgrid, HuggingFace, Redis, VertexAI vector store, Redis Vector Store, LangServe, LangSmith, Mlflow, Kubeflow.

Open-source contributions

LangChain: Contributed to the development and documentation of the largest framework for developing LLM powered applications. Repository, Commits
LLamaIndex: Contributed to the documentation of Gen AI framework. Repository, Commits
PandasAI: Contributed to the development of LLM project. Repository, Commits
Medium: Wrote blogs on Large Language Models and Generative AI applications. (Blogs)

Work Experience

Tata Consultancy Services(TCS)
Machine Learning Operations (MLOps) Engineer — Jul 2021- Aug 2023
Client: Albert Heijn

- Developed an ML model to resolve 30% of incidents in ServiceNow, reducing manual resolution time by 95%.
- Achieved 95% accuracy in incident prediction using Random Forest and SVM algorithms.
- Performed data analysis with Pandas and Numpy to understand distribution and patterns.
- Integrated ML models with MongoDB to manage and store over 10,000 tickets.
- Migrated on-prem Python app to Azure cloud using microservices, reducing infrastructure costs by 20%. **Technologies**: Docker, Kubernetes, Azure DevOps, Docker Compose.
- Deployed Python REST APIs using Flask and FastAPI for seamless integration between containers.
- Implemented CI/CD processes, increasing deployment speed and efficiency by 30%.
- Monitored app performance using Azure Log Analytics and optimized system reliability.
- Automated Azure DevOps release pipelines using SDKs, reducing downtime and manual effort.
- Collaborated with cross-functional teams via Jira and Confluence in an Agile environment.
- Reduced project budget by decommissioning underutilized servers and migrating dependencies to Azure.
- Developed a chatbot using Rasa, Spacy, and BERT for ServiceNow ticket management.
- Used Azure Functions to host and integrate chatbot components with external services.
- Documented code and processes for team knowledge sharing and future collaboration.

Cognizant Technology Solutions
Python Developer — Feb 2021- Jun 2021

- Implemented data preprocessing pipelines using Python libraries like pandas and NumPy to cleanse and transform raw user interaction data.
- Built data preprocessing pipelines using Pandas and NumPy to cleanse and transform raw user data.
- Engineered features from user data, focusing on product attributes and user preferences.
- Integrated collaborative filtering algorithms in the recommendation system using scikit-learn.
- Developed Python scripts to train recommendation models using TensorFlow and PyTorch.

Projects

Proximity-Based User Recommendation System

- Developed a recommendation agent to to identify nearby users with shared interests using Large Language Models (LLMs), Tools, Retrieval-Augmented Generation (RAG), and advanced prompt engineering.
- Integrated custom tools to LLM for database interaction and geospatial analysis using VertexAI agent builder to improve the accuracy and relevance of user recommendations by querying databases and performing distance calculations. Used Cloud run functions to deploy custom tools developed.
  Technologies used: VertexAI, VertexAI Agent builder, LangChain, Generative AI, Large Language Models (LLMs), MongoDB, geopy, google maps api, Uber H3, VertexAI Vector Store, OpenAI GPT-4. Project Link

Chat with your docs

- Developed a context-aware application powered by Large Language Models (LLMs) and LangChain
- Applied the Retrieval-Augmented Generation (RAG) approach to utilize a provided knowledge base as context, enabling the application to engage in informed and coherent conversations. To make it handle more data, I have used redis vector store to cache all documents and utilise them for RAG.
  Technologies used: HuggingFace, Transformers, ChromaDB, LangChain, Vector DBs, Vertex AI vector search, Generative AI, Large Language Models (LLMs), Gemini Flash, LLMOps. Project Link

Achievements and activities

- Received **On The Spot** award from TCS for making architect changes to decommission few servers and migrate them to azure cloud, resulted in a 40% reduction in project budget.

Certifications

- Microsoft Certified: DevOps Engineer Expert (AZ-400). Certificate
- Microsoft Certified: Azure Administrator Associate (AZ-104). Certificate
- Microsoft Certified: Azure Fundamentals (AZ-900). Certificate