MALATHI

US Citizen - Bay Area CA · Malathi.kaliappan11@gmail.com · 925-699-5473 ·
https://www.linkedin.com/in/malathi-kaliappan014/

EXPERIENCE

Pryon
Senior Machine Learning Engineer
San Francisco, CA
January 2024 - Present

- Created AI assistants using Retrieval-Augmented Generation (RAG) with LangChain/LlamaIndex, utilizing vector databases, embedding models, sentence transformers, tokenization, and multiple knowledge sources for efficient information retrieval and processing.

- Implemented a multi-agent system using Agentic AI with AutoGen, enabling multiple models to interact and collaborate for complex task automation, improving decision-making and optimizing workflows.

- Implemented LangChain for prompt templating, memory, context handling, and task chaining, optimizing LLM-based workflows.

- Built interactive interfaces using Streamlit/React with FastAPI for seamless API integration and front-end interaction.

- Leveraged VectorDB for embeddings and integrated HuggingFace models to enhance the accuracy and speed of data retrieval.

- Conducted testing for retrieval and response evaluation using advanced metrics like BLEU, ROUGE, and retrieval precision to ensure the accuracy and effectiveness of AI-driven solutions.

- Applied advanced NLP techniques such as summarization, classification, and sentiment analysis using NLTK and spaCy, in conjunction with LLMs and LangChain for effective task automation and chaining.

- Established monitoring and logging systems for deployed models, ensuring real-time performance tracking and facilitating timely updates and troubleshooting.

- Deployed and optimized models on AWS Bedrock, utilizing its managed service to efficiently build and scale generative AI applications.

- Fine-tuned large language models (LLMs) with domain-specific datasets to optimize AI assistance, ensuring high accuracy and relevance in conversations.

- Optimized model deployment across cloud platforms, ensuring seamless integration with existing business processes.

- Implemented large-scale infrastructure solutions using EKS for container orchestration and Git for CI/CD pipeline management.

- Used Python libraries like Pydantic, FastAPI, and Pytest for developing applications, validating data, and automating testing.

Bank of the West
Machine Learning Engineer
San Ramon, CA
July 2021 - December 2023

- Developed and deployed real-time fraud detection models using Python and PyTorch, implementing algorithms for predictive analytics to identify suspicious transactions and enhance security.

- Designed neural networks in PyTorch, utilizing its dynamic computation capabilities to process real-time data efficiently.

- Built and maintained end-to-end machine learning pipelines, integrating tools like pandas and scikit-learn for data preprocessing, feature engineering, and model training.

- Implemented MLOps practices using Docker and AWS SageMaker to streamline model deployment, version control, and monitoring in production environments.

- Automated workflows for model training, tuning, and deployment on AWS SageMaker, ensuring scalability and reducing operational overhead.

- Designed and integrated CI/CD pipelines to automate testing, deployment, and updates for machine learning models, ensuring continuous improvement in production systems.

- Tested and validated models with scikit-learn, ensuring high accuracy and reliability by employing robust evaluation metrics.

- Leveraged monitoring tools on AWS to track performance metrics and ensure system stability with minimal downtime.

- Enhanced data transformation efficiency by 20% using pandas and NumPy for preprocessing and feature engineering in large datasets.

- Collaborated with cross-functional teams to integrate machine learning solutions into existing systems, improving data flow and prediction accuracy.

- Optimized model workflows by profiling processes and implementing cost-efficient solutions to improve performance and scalability in production environments.

Bank of the West   San Ramon, CA
Senior Data Analyst   June 2018 - June 2021

- Designed and developed interactive Tableau and Superset dashboards, incorporating KPIs, trends, and custom visualizations to support business decisions.

- Applied best practices in data visualization, creating user-friendly interfaces with dynamic filters and visual elements like heatmaps, bar charts, and trend lines.

- Analyzed complex datasets from multiple sources (SQL, APIs, Excel) to identify actionable insights, trends, and anomalies that drove business strategies.

- Leveraged SQL to write complex queries for data extraction, preprocessing, and aggregation, enabling accurate and timely reporting in Tableau and Superset.

- Integrated data from various sources (SQL Server, AWS Redshift, Google Sheets) into Tableau and Superset, ensuring seamless data connection and performance optimization.

- Established secure, efficient data source connections and performed complex SQL joins to merge data, improving reporting accuracy and efficiency.

- Translated business requirements into visual solutions, creating dashboards and reports that provided clarity and insights for both technical and non-technical stakeholders.

- Presented insights to stakeholders, effectively communicating complex findings through simplified visualizations and ensuring alignment with business objectives.

- Facilitated iterative feedback loops with cross-functional teams, continuously improving dashboards and reports based on stakeholder input.

- Created and managed automated data updates and scheduled refreshes in Tableau Server and Superset, ensuring timely data availability and consistency.

- Regularly audited dashboards to ensure data accuracy and integrity, troubleshooting performance issues and improving overall dashboard load times by 40%.

- Developed and delivered training sessions for internal teams on using Tableau and Superset for self-service data analysis, fostering a data-driven culture.

- Managed version control of dashboards, ensuring consistent updates and smooth deployment of new visualizations and features across teams.

Merck   Rahway, New Jersey
Data Analyst   January 2017 - May 2018

- Designed and developed interactive dashboards in Tableau and Power BI, enabling stakeholders to visualize trends and derive actionable insights.

- Performed statistical analysis using Python (NumPy, Pandas, SciPy) to uncover significant data patterns, supporting data-driven decision-making.

- Conducted data cleaning and validation processes to ensure data accuracy and reliability, reducing data discrepancies.

- Analyzed customer segmentation using k-means clustering and other techniques to optimize marketing strategies and improve customer engagement.

- Monitored real-time business metrics through dashboards, enabling proactive identification of trends and anomalies to support timely decision-making.

- Extracted, transformed, and analyzed large datasets in real-time to provide actionable insights for operations and strategic planning.

- Monitored and maintained data flow pipelines to guarantee timely updates for real-time dashboards, minimizing delays in data reporting.

- Established data governance practices to ensure data integrity, compliance with industry standards, and security protocols.

- Collaborated with cross-functional teams to gather data requirements, ensuring alignment with business objectives and delivering actionable insights.

- Integrated and transformed data from diverse sources, such as CRM systems and web analytics tools, to create unified datasets for analysis.

- Set up monitoring systems to track data pipeline performance, enabling quick resolution of issues and minimal disruptions.

- Provided data-driven recommendations that improved business operations and supported strategic decision-making processes.

TECHNICAL SKILLS

Programming Languages:	Python, SQL
ML Frameworks/Libraries:	Pytorch, Huggingface, MLFlow, Tensorflow, Keras
Cloud Platforms:	AWS Bedrock, AWS Sagemaker
Data Processing and Analysis:	Numpy, Pandas, Spark, Databricks
Natural Language Processing (NLP):	NLP, Pytorch, Huggingface
Web Development and Deployment:	Streamlit, FASTAPI, Docker, Kubernetes
Other Tools and Technologies:	LangChain/LlamaIndex, RAG, Gen AI, LLM, Vector DB