KARTHIKEYAN K

Data Scientist / ML Engineer

|(+91) 8248663423 |karthickinfo45@gmail.com|

| Github/kkarthik3 | LinkedIn/kkarthick-k |Portfolio| kaggle/kk3103 |

EXPERIENCE

Machine Learning Engineer[Kadit Innovations, Tenkasi, India]   **Mar 2025 – Present**

- **Protein Vector Search :** Designed and implemented an AI-powered protein search engine using Neo4j vector databases to perform similarity-based retrieval, integrating metadata from multiple protein databases to map sequences and identify relevant matches, including detailed annotation of post-translational modifications (PTMs).
- **CCTV Attendance Monitoring System**: Built a real-time attendance and security monitoring solution using RTSP camera feeds with on-device motion detection. Integrated WebSocket triggers to capture events, applied YOLO for person detection and InsightFace facial embeddings for identity recognition, and logged verified entries into PostgreSQL for attendance tracking and auditing.
- Deployed the applications on AWS, leveraging services such as EC2, S3, EKS and Lambda to ensure scalable performance and reliability for processing large biological datasets in a cloud-native environment.

Associate Machine Learning Engineer [ISPG Technologies Pvt.Ltd, Kochin, Kerala]   **Dec 2024 – Mar 2025**
(Intern: 7 Months → Associate ML Engineer: 3 Months)

- **RAG-Based Product User-Manual Chatbot (Development)**: Created a RAG-based product manual chatbot API by crawling and scraping over 5000 websites using Selenium, stored embeddings in Mongo-VectorSearch, and structured the workflow using LangGraph with tools and agents for particular sections.
- **AI Chatbot Workflow with LangGraph Agents & AWS (Pre-Production):** Developing a product with a structured workflow for an AI chatbot using LangGraph tools and agents, enabling efficient, task-specific interactions by automating responses and decision-making processes, improving user experience and achieving accuracy levels of over 70% compared to manual flow approaches. The solution features a centralized chatbot with an admin dashboard for controlling and customizing functionalities.
- **RAG Technology Integration (Development):** Developed and integrated Database Retrieval-Augmented Generation (RAG) into the CRM database, enabling seamless translation of user NLP queries into Cypher queries (Graph DB) and converting Cypher responses into human-readable formats. Implemented restricted delete, create, and update operations. This integration enhanced client interaction and operational efficiency by 40%.
- **AI-Based E-Commerce Query System (Production):** Designed and developed two AI-based e-commerce query systems and an admin dashboard for cost and query analysis using MongoDB VectorSearch for automobile spare parts and jewelry retailers. Implemented AI-driven search systems that enhanced customer experience with fast, accurate queries and natural language processing by 200ms. Integrated advanced search and AI recommendations, improving user experience by 50% and driving sales.

Data & Operations Associate [MK Labels, Chennai]   **May 2021 – Jul 2022**

- Collected, organized, and maintained large datasets of financial transactions (credit, debit, accounts payable/receivable) using Excel and Pandas, ensuring data accuracy and integrity.
- Managed inventory Data of 500,000+ stock units/Day using MySQL, ensuring 99% on-time delivery and accurate stock records.
- Streamlined workflows between clients and production teams using MySQL, Cron jobs, and automated email alerts, increasing operational efficiency by 20%.

EDUCATION

Bharathiar University, Coimbatore   **Aug 2021 - Apr 2024**
Masters in Data Analytics, Cumulative CGPA: 7.91/10

Ramakrishna Mission Vivekananda College, Chennai   **Jun 2018 - Jun 2021**
Bachelors in Physics, Cumulative CGPA: 8.02/10

SKILLS

Libraries	Python (Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Tensorflow, Keras, NLTK, Streamlit), Fast-API, Flask, Langchain, Langgraph, Neo4j
Languages	Python (Intermediate), R (Fundamentals), SQL , Cypher (Neo4J)
Data Wrangling	Data Preparation, Data Extraction, Data Cleaning, Exploratory Data Analysis(EDA), Data Visualization, Feature Engineering,Feature selection, Model Building, IoT .
Machine Learning	Data Modeling, Clustering & Classification, Quantitative Analysis, Regression, Transformers, Predictive Modeling, Model Validation, Model deployment, CNN & RNN, LSTM, BERT.

Tools	Jupyter Notebook, Arduino IDE, MS Excel, Tableau, VS code, Postman, n8n.
Technologies	Fundamentals (Hive) , Langchain, LLM, Large Language Models, IoT, Onnx, Docker, Kubernetes.
Cloud	AWS EC2, AWS IAM, AWS Beanstalk, AWS Bedrock, AWS RDS, AWS Sagemaker (Fundamentals), AWS APIGateway, AWS Lambda, AWS S3, AWS ECR, GCP

PROJECTS

Omnibot - LLM RAG Retrieval, [Langchain, LLM, Genai, Word Embeddings] [Link]   **March 2024**
- Developed a streamlined Vector Storage leveraging Google word embeddings, resulting in a notable 50% reduction in processing time by harnessing Google's backend infrastructure for execution.
- Implemented Faiss document retrieval from Meta for efficient text data extraction, prioritizing speed and enabling rapid nearest neighbor search and clustering operations. Furthermore, integrated advanced language models such as GEMMA and Gemini with Faiss to ensure superior performance and accuracy in data processing tasks.
- Deployed the system on Amazon EC2 Ubuntu AMI (Amazon Machine Image), ensuring scalability and reliability while leveraging Amazon's robust cloud infrastructure for seamless operation and maintenance.

HUMAN VS AI TEXT CLASSIFICATION, [NLP, LLM, BERT, Transformers, Deep Learning] [Link]   **January 2024**
- Implemented XLMBert, a variant of the BERT architecture, using the Keras NLP framework for natural language processing tasks, capable of handling more than 100,000 instances.
- Conducted fine-tuning of the model for Adaptation to Domain and Data using a custom classification dataset, employing techniques such as stopword removal and word embeddings for preprocessing.
- Achieved 99% testing accuracy in discerning human-generated text from machine-generated text, demonstrating the model's effectiveness.

MRI Image Enhancer, [Deep Learning, CNN, Image Processing][Link]   **December 2023**
- Implemented TensorFlow's CNN framework to enhance low resolution medical MRI images using the FSRCNN model increased image clarity by 35% enabling extraction of crucial details for accurate diagnosis and treatment planning.
- Developed a Streamlit environment for interactive deployment, facilitating easy input of images and immediate visualization of enhanced outputs. This streamlined workflow led to a 14% improvement in real-time predictions.
- The FSRCNN model with PReLU activation performs better compared to bicubic and ReLU by 20% and 2%, respectively.

ADDITIONAL ACTIVITIES

Participations : 1) Co-Headed a team of 6 in the Smart India Hackathon 2023 (SIH1338), delivering a pioneering solution to a critical problem In Mininstry of Mining Department identified by the Central Government of India; showcased exceptional problem-solving abilities.
          2) Headed a Team of 4 in the DARPG Hackathon 2024 which is orgnized by Government of India.
Seminars : Delivered Hands on Seminar on basic data preprocessing for machine learning for Data Analytics Students from the Department of Computer Application in Bharathiar University, positively impacting 39 Students.
Kaggle : Actively contributing to Kaggle competitions, consistently engaging in data science challenges to stay abreast of industry trends and enhance problem-solving skills.

COURSES AND CERTIFICATIONS
- Neo4j Certified Professional, **Neo4j**
- AWS Certified AI Practitioner AIF-C01, **Udemy**
- AWS Certified Cloud Practitioner CLF-C02, **Udemy**
- Cypher Fundamentals, **Neo4J**
- Database Management System, **NPTEL**