KARTHIKEYAN K

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

Data Analyst / Data Scientist | Github/kkarthik3 | LinkedIn/kkarthick-k |Portfolio| kaggle/kk3103 |

EDUCATION

Bharathiar University, Coimbatore

Masters in Data Analytics, Cumulative CGPA : 7.6/10 **May 2024**

Relevant Courseworks: Data Analysis, Machine Learning, SQL, Deep Learning, Statistics

Ramakrishna Mission Vivekananda College, Chennai **June 2021**

Bachelors in Physics, Cumulative CGPA : 8/10

Relevant Coursework: Internet Of Things (IoT)

SKILLS

Libraries	Python (Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Tensorflow, Keras, NLTK, Streamlit), R (Dplyr, ggplot)
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, R Studio, MS Excel, Tableau, VS code, Power BI.
Technologies	Fundamentals (PySpark, Hive) , Langchain, LLM, Large Language Models, IoT, Flask, FastApi.
Cloud	AWS EC2, AWS IAM, AWS Beanstalk, AWS Bedrock, AWS RDS, AWS DocumentDB, AWS Sagemaker

EXPERIENCE

Machine Learning Engineer Intern [ISPG Technologies Pvt.Ltd, Kochin, Kerala] **May 2024 – Present**

- **RAG Technology Integration:** Integrated Retrieval-Augmented Generation (RAG) into the CRM database, enabling the seamless translation of user NLP queries into Cypher queries and converting Cypher responses into human-readable formats. This integration enhanced user interaction and operational efficiency by 40%.
- **LLM Evaluation:** Conducted comprehensive evaluations of various large language models, including GPT-4, GPT-3.5, LLAMA3, Mistral, and Groq, to identify the most suitable models for domain-specific use cases. This effort led to a 20% improvement in model selection accuracy.
- **AI-Based E-Commerce Query System** : Developed an AI-based e-commerce query system by embedding all product root paths into vector format and storing them in a .pkl file. Achieved query response times within 1000 ms, significantly enhancing user experience.
- **RAG-Based Product Manual Chatbot** : Created a RAG-based product manual chatbot by embedding approximately 5000 HTML files into vector format and storing them in a .pkl file. Enabled users to query their doubts naturally, eliminating the need to manually search through all 5000 HTML files.

PROJECTS

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.

SENTIMENT ANALYSIS ON AMAZON PRODUCT REVIEWS [Transformers, HuggingFace,NLP**]** **December 2022**

- Implemented web scraping techniques with Python's BeautifulSoup library to extract Amazon reviews data efficiently, enabling the collection of large volumes of diverse user feedback from various product listings on the platform.
- Leveraged a Pretrained ROBERTA model from Hugging Face, a state-of-the-art natural language processing (NLP) framework, to analyze the sentiment of the scraped reviews. By utilizing a pre-trained model, the need for manual training of a sentiment analysis model from scratch was eliminated By 100%
- Utilized ROBERTA model to analyze scraped reviews and classify sentiments (positive, negative, or neutral), extracting valuable insights on customer opinions and preferences toward the products.

OTHER 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 6 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

- AWS Certified Cloud Practitioner CLF-C02, **Udemy**
- Cypher Fundamentals, **Neo4J**
- Python (Basic), **HackerRank**
- SQL (Basic), **HackerRank**
- Text Summarization Using Bert, **Great Learning**
- Fundamentals of Visualization With Tableau, **Coursera**
- Deep Learning Fundamentals, **Eduonix**
- Database Management System, **NPTEL**
- Data Analysis With Python, **Cognitive Class**