KARTIKEY SHAURYA

kartikeyshaurya1@gmail.com ◇ 918765882608 ◇ Linkedin/kartikeyshaurya ◇ GitHub

EXPERIENCE

Data Scientist July 2024 - Current
Primotech *Chandigarh,Punjab*

1. Developed and deployed AI agents(Langchain and Langgraph) on the Rejara platform to efficiently collect user data and support other platform agents by leveraging insights to enhance decision-making and functionality.
2. Worked on custom NLP models for text summarization(Pegasus Fine-tuned for conversation), topic classification(Bert Fine-tuned), and recommendation systems(ES based), which significantly decreased manual data processing time and enhanced the quality of data-driven decision-making.

Machine Learning Engineer Aug 2022 - July 2024
Sapio Analytics *Mumbai, Maharashtra*

1. Collaborated with the Assam Government to enhance police deployment efficiency by analyzing historical data, identifying crime patterns, and providing actionable insights for strategic decision-making
2. Independently built and managed a classical regression model predicting driver performance clusters, increasing driver income by 8-10% and reducing attrition by nearly 15%.
3. Created an LLM-based chatbot for job assistance, integrated with a recommendation system that earned 10,000 dollars in AWS credits.
4. Streamlined the Internal Data Lake (160GB), achieving a 40-50% boost in write speed, 20% data compression, and a 10-20% faster response time. Developed a PII Detection model to enable secure data sharing.

Machine Learning Product Engineer(intern) Jan 2022 - April 2022
Highradius *Bhubaneswar, Orissa*

1. Worked on a **React** application capable of generating invoices based on real-time data and implemented regression techniques to **identify** potential payment delays.
2. Increased payment delay detection accuracy by 15%. Additionally, introduced an ML pipeline to automate ML training and manage data drift effectively.

SKILLS

Technical Skills	Python, Machine Learning,NLP, LLM, Microservices, MLOps
Tools and Technologies	Scikit learn, Tensorflow, Pytorch, Streamlit, Flask, Git, Docker AWS, LangChain, Langgraph and Langsmith
Database	MySql, Postgres, Elastic Search, MongoDB, Pinecone, Redshift

EDUCATION

Bachelor of Computer Science, Lovely Professional University 2019 - 2023 (7.7/10)

PROJECTS

Samarth. Developed essential **analytical tools** on the Samarth platform catering to both job seekers and providers. Created and Enhanced the **search** and **recommendation system** and optimized data collection through the implementation of an **AI-powered Bot**. Established **planning** and data pipelines for Management Information Systems (**MIS**), along with crafting tools for monitoring data quality, refining data collection processes, parsing data effectively, and ensuring secure data storage.

Neural Ocean(Internal Data Lake). Optimized the analytical query by implementing DWH Architecture and airflow, resulting in a 40 percent reduction in response time, and it helped strengthen the internal profiling tool.

PERSONAL PROJECTS AND ACHIEVEMENTS

Data Anonymizer – Part of the West Bengal Government Initiative

1. Fine-tuned DistilBERT to detect personal information (e.g., names, addresses, organizations) in text and integrated YuNet for facial detection and anonymization in images.
2. Developed a platform that enables users to create custom regex patterns to redact sensitive information (e.g., emails, phone numbers) in text, securing 3rd place in the initiative competition.

Self Driving Car Udacity course curriculam

1. Created a custom model to clone the behaviour of moving car.(throttle , brake ,left,right etc)
2. Successfully implemented the model, ensuring the car navigated an empty road without crashing into barriers.

Real-Time Disaster Message Categorization Kaggle Self Exploration Project

1. Developed an NLP model to categorize disaster-related messages in real-time using a labeled dataset of tweets and messages from actual disasters.
2. Processed and merged datasets, storing the final data in an SQLite database for efficient handling.
3. Built a machine learning pipeline to preprocess, split, and train the data, exporting the final model as a pickle file.
4. Developed a web application to deploy the model and predict disaster message categories in real-time.

CNN-Based Early Stage Fire Detection Classifier

1. Developed a Convolutional Neural Network (CNN) model for early-stage fire detection in images.
2. Collected and preprocessed fire and non-fire image datasets for training the model.
3. Achieved high classification accuracy in identifying fire hazards in images, ensuring quick detection of potential fires.
4. Implemented the model in a real-time fire monitoring system for practical use in safety applications.