Kartik Bharat Sonaghela

+91 8879684343 | kbskartik@gmail.com | LinkedIn | GitHub | Portfolio

TECHNICAL SKILLS

- **Programming Languages**: Python, SQL
- **Databases**: MySQL, VectorDB
- **Frameworks**: PyTorch, LangChain
- **DevOps, Cloud & Misc.:** Docker, Git, Postman, Flask, CI/CD, AWS, LLM-Inference
- **Data-Science & AI**: Machine Learning, Deep Learning, NLP, RAG, GenAI, LLMs, Fine-Tuning, AI-Agents

EXPERIENCE

Data Scientist, DeepNeura Technologies May 2022 – Present | Mumbai, India

Crop Trend Prediction:

- Developed a crop price prediction system using agricultural data from Agmarknet.
- Cleaned and pre-processed data, handling missing values and selecting key features for optimal performance.
- Fine-tuned model parameters to enhance accuracy and prediction reliability.
- Utilized the XGBoost algorithm to forecast crop prices for the upcoming 10 days, aiding farmers in decision-making.

Chatbot Development:

- Created a RAG-based application for Bayer, utilizing large language models to answer questions and provide support to farmers.
- Leveraged LangChain framework, MongoDB vector database with GPT-4o-mini model to build an interactive chatbot, handling diverse queries.
- Designed the chatbot to support multi-language communication, responding in any Indian language or English, and improved communication and decision-making for farmers by delivering tailored responses based on Bayer's internal data.

Jr Data Scientist, Care Risk Solution March 2021 - April 2022 | Mumbai, India

- Developed an Auto-ml Software which include entire machine learning cycle life starting from uploading data to predicting the data.
- Text Summarization of scraped news article into 200 words.
- Created a covid dashboard using Tableau during the second wave which shows information regarding vaccination, how many are positive and how many have recovered.

PROJECTS

YouTube Comment Sentiment Analysis Chrome Extension

- Developed a Chrome extension to analyze YouTube comment sentiment in real time, providing insights into audience reactions.
- Utilized MLflow for experiment tracking, DVC for dataset versioning, and built a XGBoost model with TF-IDF vectorization, SMOTE, and Optuna for hyperparameter tuning to ensure accurate sentiment classification.
- Deployed the backend using Flask APIs in Docker containers on AWS (S3, CodeDeploy, ECR), integrating with the Chrome extension frontend built with HTML, CSS, and JavaScript.

Fine-Tuned Llama for Indian laws

- Fine-tuned a 3.18B-parameter LLaMA 3.1 model on 35K legal records, including BNS sections and public laws, to enhance legal understanding.
- Constructed the dataset with instruction, input, and simplified explanations generated by the Gemma 2.0 flash model to demystify complex legal jargon.
- Optimized inference using the vLLM framework on an L4 GPU after fine-tuning model on an A100 GPU for 6 hours, enabling users to ask legal questions and understand legal documents like FIRs.

AI Code Assistant

- Developed a VS Code extension for code generation by fine-tuning the Phi-3 4-bit quantized model on the iamtarun/python_code_instructions_18k_alpaca dataset using LoRA with the Unsloth library.
- Optimized a small language model (SLM) for local execution, enabling efficient and accurate Python code generation within VS Code.

EDUCATION

University of Mumbai **Aug 2016 – Oct 2020**
Bachelor of Engineering in Computer Engineering Mumbai, India