Katta Hari Charan

Email: hccharankatta@gmail.com
Portfolio: https://charanportfolio-theta.vercel.app/
Mobile: +91-7036696045
Github: https://github.com/haricharnbytes
https://www.linkedin.com/in/haricharankatta/

Education

- **M. S. Engineering College (VTU)** — Bengaluru, India
  Bachelor of Engineering - Computer Science and Engineering — *August 2019 - June 2023*
  Courses: *Operating Systems, Data Structures, Artificial Intelligence, Machine Learning, Databases*

Skills Summary

- **Languages**: Python, SQL
- **Data Analysis & Visualization**: Pandas, NumPy, Matplotlib
- **Machine Learning & Deep learning**: Scikit-Learn, XGBoost, Keras, TensorFlow, PyTorch
- **Generative AI & NLP**: Hugging Face, LangChain, Transformers, Large Language Models
- **Tools**: Jupyter Notebook, Git, GitHub, VS Code, Google Colab, Streamlit

Experience

- **CSIR - Fourth Paradigm Institute, GOI** — Bengaluru, India
  Project Associate — *Dec 2023 - Present*
  - Developed a UAV-based chlorophyll estimation framework using Micasense Altum multispectral imagery and in-situ SPAD meter data for precision agriculture applications in paddy fields.
  - Processed and calibrated UAV data and derived key vegetation indices using empirical methods to enhance crop health assessment.
  - Implemented and compared multiple Machine Learning models (SVR, Random Forest, Extra Trees, XGBoost, AdaBoost, Gradient Boosting) with 80:20 train-test split, leveraging Gaussian Mixture Models (GMM) for data augmentation to improve generalization.
  - Assessed predictive performance using R^2, RMSE, and MAPE, and applied SHAP analysis to interpret feature importance.

- **Aditi Software** — Hyderabad, India
  Machine Learning Intern — *Oct 2023 – Dec 2023*
  - Conducted comprehensive Exploratory Data Analysis (EDA) to understand data distributions, detect outliers, analyze feature correlations, and identify key factors influencing house prices.
  - Trained and evaluated multiple machine learning models for house price prediction, comparing their performance to select the most accurate and robust approach.
  - Assessed model performance using evaluation metrics such as RMSE, MAE, and R^2 score to identify the best-performing model for dependable property price prediction.

Projects

- **RAG-Based Intelligent Document QA System (RAG)** — GitHub
  Developed an end-to-end RAG system for intelligent question answering in a document-based setting. Implemented text chunking, embedding creation, and semantic vector indexing with FAISS Vector Database for similarity search and retrieval. Integrated the LangChain framework with Large Language Models (LLMs) to produce context-aware and accurate responses..
  Tech: LangChain, LLMs, FAISS, Vector Databases, Embeddings

- **Natural language to SQL AI Application (Generative AI)** — GitHub
  Developed an AI-powered database assistant using LangChain, ChatGroq, and Python, reducing manual query analysis time by enabling automatic schema-based query generation. Designed dynamic prompts and a modular response pipeline to deliver accurate, step-by-step, schema-aware answers. Enabled non-technical users to query structured databases efficiently while maintaining full data privacy.
  Tech: Python, LangChain, ChatGroq, OpenAI

- **YouTube Video Summarizer using Large Language Models** — GitHub
  Built a Streamlit web application to automatically extract and preprocess YouTube video transcripts. Implemented intelligent text chunking and AI-driven summarization using large language models to generate concise, structured notes with real-time progress updates.
  Tech: Python, Streamlit, YouTube Transcript API, Large Language Models.

- **Diagnostic Breast Cancer Prediction (ML)** — GitHub
  A machine learning–based system that predicts breast cancer diagnosis using clinical and diagnostic data to support early and accurate detection. Achieved 95% classification accuracy, streamlining diagnostic screening processes.
  Tech: Python, Pandas, Scikit-Learn, XGBoost