KAKARA DURGAPRASAD

durgaprasadkakara29@gmail.com 9493744785 Hyderabad Linkedin Github

Portfolio

Professional Summary

Detail-oriented Data Science professional with 1+ year of practical experience in data analysis, machine learning, and model deployment. Began career as a Graduate Engineer Trainee in EHS, where extensive data work inspired a transition into data science. Skilled in Python, ML, NLP, time series forecasting, and deep learning (ANN), with hands-on expertise in Scikit-learn, TensorFlow, FastAPI, and MLflow. Completed a professional data science internship and currently delivering end-to-end freelance projects, including NLP, forecasting, and RAG-based applications. Actively seeking a full-time opportunity to apply expertise and grow as a data scientist.

Work Experience

Self-driven, *Freelance Data Science Projects* 04/2025 – Present | Remote

- Designed and implemented end-to-end machine learning projects, from data preprocessing to deployment using FastAPI
- Built and trained deep leaming models (ANN) and applied advanced time series forecasting techniques for real-world prediction tasks
- Integrated MLflow into projects for model tracking, version control, and performance monitoring
- Developed a RAG-based Smart FAQ Assistant leveraging ChromaDB for semantic retrieval and LLMs for accurate, context-aware responses

Data Science Intern, *Al Variant* 07/2024 – 01/2025 | Hyderabad

- Cleaned and preprocessed real-world datasets using Python (Pandas, NumPy) for machine learning tasks
- Performed exploratory data analysis (EDA) and visualized insights using matplotlib, seabom
- Built and evaluated ML models for classification and regression using Scikit-learn and XGBoost
- Applied natural language processing (NLP) techniques and deployed models via Streamlit and Flask for demo applications

Graduate Engineer Trainee - EHS [Data Focussed], 06/2023 – 07/2024 | Jaisalmer

Renew Private Limited

- Collected, cleaned, and analyzed EHS data including incident logs, safety inspections, and audit reports using Python (Pandas)
- Generated monthly safety reports and visualized patterns in incident types, root causes, and near misses using matplotlib and seaborn
- Contributed to internal presentations by preparing data summaries and visual reports for safety performance reviews
- Supported EHS audits by tracking environmental KPIs such as emissions, waste disposal, and energy usage, contributing to compliance reviews

Education

Rajiv Gandhi University of Knowledge and Technologies, 05/2023 | Srikakulam

Mechanical Engineering

Achieved academic excellence with a stellar 92.9% in graduation.

Soft Skills

Problem solving • Analytical thinking • TeamCollaboration • Attention to detail

Technical Skills

Programming & Database
Python, SQL, Pandas, NumPy

Deep Learning & NLP
ANN, LLMs, Hugging Face Transformers, Fine-Tuning, BERT, GPT architectures, Prompt Engineering

Data Analysis & Visualization
Data Cleaning, EDA, Matplotlib, Seaborn

Computer Vision
OpenCV, Object Detection[Yolo]

Machine Learning
Supervised & Unsupervised Learning, Model Evaluation, Forecasting

Generative AI & RAG
Vector Databases (ChromaDB), RAG Pipelines, OpenRouter API

Frameworks & Tools
Scikit-learn, TensorFlow, PyTorch, FastAPI, Flask, Streamlit, MLflow

Projects

Patient's Condition Classification Using Drug Reviews
- Objective: Created a machine learning model to analyze patient reviews and classify medical conditions and to recommend the most effective drugs.
- Approach: Applied Natural Language Processing (NLP) techniques, including TF-IDF, to preprocess and vectorize the text data. Built several classification models such as Random Forest, KNN, XGBM, LGBM and SVM to predict patient conditions from the reviews.
- Outcome: Deployed the final model using Flask, providing a user-friendly interface for real-time condition classification and drug recommendations based on patient feedback.

Tools & Technologies: Python, Pandas, Scikit-leam, TF-IDF, NLP, Flask

Agent Auto-Assignment Engine for Ticket Routing
- Objective: Developed a machine learning engine to automatically assign incoming support tickets to the most suitable agents based on contextual and performance-based features.
- Approach: Built and evaluated multiple classification models using historical ticket and agent performance data. Features included ticket priority, sentiment, agent load index, skills, and resolution history. The best-performing model, XGBoost Classifier, was deployed via FastAPI and tracked using MLflow.
- Outcome: Reduced manual ticket routing, improved response time, and enhanced overall agent productivity through intelligent, data-driven assignment.

Tools & Technologies: Python, Scikit-learn, XGBoost, MLflow, FastAPI, Pandas

Smart FAQ Assistant (RAG-based Chatbot)
• Objective: Built a RAG-based chatbot to answer company FAQs instantly, reducing time spent on manual searching.
• Approach: Ingested FAQs into ChromaDB for semantic retrieval, combined top results with user query, and generated responses using LLM (OpenRouter API). Designed FastAPI backend for API integration with strict prompt engineering for accuracy.
• Outcome: Delivered a scalable, API-driven assistant that improves employee/customer efficiency and can extend to HR, IT, or customer support domains.
Tools & Technologies: Python, FastAPI, Streamlit, ChromaDB, OpenRouter API, NLP, RAG, Prompt Engineering

Certificates

Data Science Certification - Excelr 2024

Languages

English • Hindi • Telugu