Sumukh Acharya

Bangalore | sumukh.acharya@gmail.com | 9972454072 | sumukh-acharya.vercel.app | GitHub | LinkedIn

Education

PES University, B-Tech in Computer Science and Engineering 2022-2026

Coursework: Machine Learning, Databases, Deep Learning, Software Engineering, Cloud Computing, Big Data, Data Analysis, Statistics for Data Science, C/Java, Data Structures and Algorithms, OS, Computer Networks

Experience

CODMAV, Data Science Intern | June 2024 - August 2024 **GitHub Link**

- **Tools Used:** pandas, numpy, scikit-learn, Matplotlib, seaborn, XGBoost, CatBoost, SVM, KNN, PCA, RFE, SelectK, BSO, Random Forest
- Led the development of a **lung cancer risk prediction system** using advanced feature selection (**PCA, BSO, RFE, SelectKBest**) and machine learning models (**XGBoost, CatBoost, SVM, KNN**), achieving **98.75% accuracy** and **96.25% recall**; co-authored on a research paper accepted for publication on **IEEE Xplore**.

Projects

Senior Vision AI – Elderly AI Assistant **GitHub Link**

- **Tools Used:** Python, Flask, Gunicorn, JavaScript, HTML/CSS, Google Vision API, Gemini API, Render, Vercel
- **Architected** a **full-stack AI** solution to assist the elderly and visually impaired by scanning product labels, generating simplified summaries, and vocalizing the results.
- **Deployed** a resilient backend API on **Render** and a responsive frontend on **Vercel**, featuring a modern and user-friendly UI.

DFOS – Distributed File Orchestration and Synchronization **GitHub Link**

- **Tools Used:** socket, TCP Protocol, ThreadPoolExecutor, Multi-Threading, Client-Server Architecture
- **Engineered** a secure, **multi-client file server** with user authentication, enabling upload, download, preview, deletion, and listing of user-specific files with **concurrent handling** and graceful shutdown support.

Dynamic Fare Engine – Ensemble Forecasting Model for Price Optimization **GitHub Link**

- **Tools Used:** Python, Pandas, NumPy, Scikit-learn, Statsmodels, XGBoost, Matplotlib, Seaborn
- Developed a **dynamic pricing strategy** by building an ensemble model (SARIMAX, XGBoost, VAR) to forecast fares for a multi-vehicle ride-sharing service, achieving a **SMAPE score of 3.27** using three years of historical time-series data.

Skills

- **Languages:** Python, MySQL
- **Data Science:** PyTorch, TensorFlow, Scikit-learn, Pandas, NumPy
- **Big Data:** Kafka, Spark
- **MLOps and Deployment:** Docker, Kubernetes, FastAPI, Streamlit, Git

Achievments

- IBM Data Science Professional Course by Coursera
- Secured 7th place out of 364 teams in a Kaggle Data Analytics