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

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.

Vogue Vision – Automated Fashion AI System   **GitHub Link**

- **Tools Used:** Python, PyTorch, Torchvision, FastAPI, Uvicorn, Streamlit, Scikit-learn, Pandas, NumPy, Pillow
- Architected a multi-output CNN in PyTorch to automate product attribute tagging (type, color, season, gender); achieving **90.14% test accuracy** and streamlining the product listing process for e-commerce.
- Engineered a **full-stack** solution by serving the model via a **FastAPI REST API** and an interactive **Streamlit web application**.

Streamly – Content Strategy and User Retention Analysis for Streamly   **GitHub Link**

- **Tools Used:** Python, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn
- Engineered a predictive model using a Random Forest Regressor to forecast user retention based on content attributes, achieving a high **R-Squared value of 0.89**.
- Provided actionable insights for content strategy by analyzing financial performance (ROI) and genre trends, helping to **guide future acquisition** and **production decisions**.

Skills

- **Languages:** Python, SQL, R, C
- **Data Science:** PyTorch, TensorFlow, Scikit-learn, Pandas, NumPy, XGBoost, Computer Vision (OpenCV), LLMs (LangChain)
- **MLOps and Deployment:** Docker, Kubernetes, FastAPI, Streamlit, Git
- **Big Data and Databases:** Spark, Kafka, Hadoop, SQL
- **Web Development:** Next.js, React, Node.js, JavaScript/TypeScript, HTML/CSS, TailwindCSS

Achievments

- IBM Data Science Professional Course by Coursera
- Secured 7th place out of 364 teams in a Kaggle Data Analytics Hackathon
- Participated in HackNight organised by ACM and contributed to Open-Source