UJJWAL SHUKLA

Machine Learning · MLOps · Software Engineering

+91 93479 13070 | shuklaujjwal.work@gmail.com | github.com/Phantom1270 | linkedin.com/in/ujjwal-s1757 | ujjwalshukla.online

EDUCATION

Vellore Institute of Technology, Chennai 2024 – 2028

B.Tech, Computer Science & Engineering | CGPA: 8.98 / 10

TECHNICAL SKILLS

Languages: Python, Java, C++, C, SQL, JavaScript

Machine Learning: Scikit-learn, Regression, SVM, KNN, Decision Trees, Ensemble Methods (AdaBoost, Gradient Boosting, XGBoost, CatBoost), K-Means, DBSCAN, PCA, Cross-Validation, GridSearchCV

Deep Learning & NLP: PyTorch, TensorFlow, Keras, ANN, CNN, RNN/LSTM, Transformers, NLTK, spaCy, TF-IDF, Word2Vec, Embeddings, RAG

MLOps & Cloud: Docker, MLflow, DVC, Git, GitHub, GitHub Actions, CI/CD, AWS (Elastic Beanstalk, CodePipeline, ECR, ECS, EC2)

Data & Tools: NumPy, Pandas, Matplotlib, Seaborn, Plotly, Flask, FastAPI, Streamlit, REST APIs, EDA, Statistical Analysis

PROJECTS

Student Performance Prediction — End-to-End ML, MLOps & Cloud Deployment

Python • Scikit-learn • Flask • MLflow • DVC • Docker • AWS (Elastic Beanstalk, CodePipeline, ECR)

- Engineered a modular, end-to-end regression pipeline with separate data ingestion, transformation, model training, evaluation and inference components, moving the workflow out of notebooks into a structured Python project.
- Evaluated multiple regression models — including XGBoost, CatBoost and AdaBoost — using GridSearchCV for hyperparameter tuning and model selection, with AdaBoost Regressor selected as the final model.
- Exposed the trained model through a Flask web application for real-time inference, supported by model serialization, logging and unit tests.
- Integrated MLflow for experiment and model tracking and DVC for data and model versioning, with Git/GitHub for source control.
- Automated deployment via a GitHub → AWS CodePipeline → Elastic Beanstalk CI/CD workflow, and built a separate containerized path by Dockerizing the service and pushing images to Amazon ECR for deployment on AWS container infrastructure.

Fact-Checking Assistant — Web Scraping + Retrieval-Augmented Generation (RAG)

Python • Web Scraping • Embeddings • Vector Store • RAG • LLM

- Built a web-based fact-checking assistant that scrapes published fact-check articles from sources such as AltNews and BOOM Live to form a domain-specific knowledge base.
- Implemented a RAG pipeline covering document processing, chunking, embedding generation, vector-store retrieval and reranking of retrieved passages before context construction.
- Generated grounded responses with an LLM over the retrieved fact-check sources, so user queries are answered from verifiable published articles rather than model memory.

Landroid — Land Health Assessment Platform · 2nd Place, National Hackathon

Python • Computer Vision • Sentinel-2 NDVI • GIS • REST APIs

- Secured **2nd place** at a national hackathon among 100+ competing teams, building an AI/ML land-health platform with a team of 4.
- Owned the land health scoring pipeline, integrating soil, terrain and geospatial APIs with Sentinel-2 NDVI imagery into composite land health scores.
- Developed the data visualization and map overlay layer to present land health and validation results to end users.

Hcare — AI-Powered Medical Cabinet & Diagnosis App

Python • Streamlit • Hugging Face • OCR • Image Recognition

- Built a Streamlit application that maintains a personal medical cabinet and flags conflicting medications through active-ingredient checks.
- Integrated a Hugging Face vision model for OCR-based medicine identification and image-based disease prediction.

LEADERSHIP & VOLUNTEERING

Member, Microsoft Innovation Club (MIC) — VIT Chennai 2024 – Present

- Helped organize and run 5+ large-scale tech hackathons, coordinating logistics, participants and event operations.