Ujjwal Gupta

Github | Linkedin
ujjwalgupta1302@gmail.com | Leetcode | GFG | Contact No. :+91-9958370742

EDUCATION

B.Tech Computer Science and Engineering
Maharaja Agrasen Institute of Technology, New Delhi, India (2022 - 2026)
GPA: 9.1/10

SKILLS

Tech Stack: Machine Learning, Deep Learning (AI/ML), Pandas, NumPy, Scikit-Learn, PyTorch, Transformers.
Languages: Python, JAVA, C/C++, HTML, CSS, JAVASCRIPT, SQL
Tools and Platforms: Git & GitHub, VS Code, Jupyter Notebook, Google Colab, MLflow, DVC, DagsHub, Apache Airflow, Docker, FastAPI, Flask, MongoDB

PROJECTS

Phishing URL Detection with MLOps Automation

|*Scikit-Learn, MLflow, DagsHub, Apache Airflow, Astronomer, Docker, GridSearchCV, FastAPI, MongoDB, Python*| *GitHub* | *DagsHub*

- Designed a production-grade MLOps pipeline using **Apache Airflow (via Astronomer in Docker)** to automate **data ingestion** from **MongoDB**, **validation**, **drift detection**, **model retraining**, and **deployment** for detecting phishng URLs entered by users.
- Configured **daily retraining** via scheduled **Airflow DAGs**, ensuring the model adapts to new phishing patterns.
- Integrated **GridSearchCV** to tune multiple classifiers (**Random Forest, AdaBoost, Gradient Boosting**, etc.) and saved the **best model** using **pickle** for deployment.
- Tracked experiments and models through **MLflow** and **DagsHub**, enabling remote **version control**, **reproducibility**, and **model comparisons**.
- Deployed the **latest model** using **FastAPI**, exposing a **/predict** endpoint that extracts features from **URLs** and returns phishing classification with the **current version** of the **model** it is **using**.
- Engineered **robust preprocessing** with custom **feature extraction**, **schema validation**, **KNN-based imputation**, and **modular logging** and **exception handling**.

Scalable MLOps Pipeline for Annual Health Premium Calculator

| *Scikit-Learn, MLflow, DVC, DagsHub, GridSearchCV, Python, Flask, Logging* | *Github* | *DagsHub*

- Developed a production-grade ML pipeline that **automates data ingestion (downloading, unzipping)**, **data validation**, **data transformation (EDA)**, **model training**, and **evaluation** to predict **annual health premium** based on user inputs**.**
- Integrated **GridSearchCV** for hyperparameter tuning and saved the best model using **joblib** for real-time predictions.
- Created a **Flask-based web UI** (with HTML) allowing users to **upload data**, **trigger entire model training pipeline (`/train`),** and perform **live predictions.**
- Implemented **MLflow** and **DVC** using **DagsHub** for **experiment tracking** and **reproducibility**, with **logging** and **exception handling** throughout the pipeline.

End-to-End Text Summarization *Pipeline*

| *Hugging Face, Transformer, T5 model, PyTorch, FastAPI, Logging, Python* | *Github*

- Developed an **end-to-end text summarization pipeline** using **Hugging Face's t5-small model** which was fine-tuned on the **SAMSum dataset**. It generates concise summaries of user-inputted conversations or text.
- Built an **automated pipeline** for **data ingestion** (downloading, unzipping), **data transformation, model training**, **model evaluation** (**ROUGE metrics**) and **model deployment**, all triggered via the **'/train'** endpoint in a **FastAPI interface**.
- '**/predict'** endpoint serves **real-time summaries** and **'/metrics'** displays **ROUGE evaluation metrics** to assess performance of the most **recently** trained model.
- The project is structured for **reproducibility** and **scalability** with **Git versioning**, **modular design**, **custom scaffolding script** and **integrated logging**.

CERTIFICATION

Neural Networks and Deep Learning (Grade 97.50 percent) Certificate Link — January 2024

Improving deep neural network (Grade 97.33 percent) Certificate Link — March 2024

Convolutional Neural Networks (Grade 96.50 percent) Certificate Link — May 2024