RAHUL SAMANT

+91-9368884561 | rahulsamantcoc2@gmail.com | linkedin.com/in/ rahul-samant-kb37 | github.com/rahulsamant37 | Portfolio

OBJECTIVE

AI/ML Intern skilled in developing and fine-tuning machine learning models, with expertise in **transformer-based architectures, retrieval-augmented generation (RAG), and large language model (LLM)** optimization. Experienced in designing efficient **data pipelines**, applying advanced mathematical and statistical techniques, and integrating AI/ML solutions into production systems.

WORK EXPRIENCE

Python Developer Intern (DEMO)
Infosys Springboard | September – Present

- To develop a robust and efficient **web scraping application** to automate **data extraction from various e-commerce platforms**, reducing manual effort and improving data accuracy.
- **Technologies: BeautifulSoup, Playwright, Streamlit, Flask, LangChain, Hugging Face**, **RAG**
- Utilized **Playwright for web scraping**, BeautifulSoup for HTML parsing**, LangChain** for language model integration, and **ChatGenerativeAI** for extracting specific product details. Deployed the application using **Streamlit and LangServe for a user-friendly interface**.
- Implemented advanced scraping techniques to handle diverse web structures, achieving **an accuracy rate of 99%** in data extraction.
- Enhanced scraping speed by **30% using multithreading**, improving performance on high-traffic pages and reducing average load time.
- Enabled data extraction from over **50 unique websites** with complex HTML structures, leading to a **20% reduction** in manual data collection time for users.

PROJECTS

End-to-End Machine Learning Pipeline with DVC, MLflow and DagsHub | Lead Developer **(**DagsHub Link)

- Designed and implemented a **reproducible, collaborative, and scalable machine learning pipeline** leveraging DVC for version control, MLflow for experiment tracking, and DagsHub for seamless integration.
- **Technologies: Python, DVC, MLflow, Scikit-learn, DagsHub**.
- Developed a collaborative, scalable ML pipeline with **DVC for version control**, **MLflow for experiment tracking**, and **DagsHub for integration**, ensuring reproducibility, automated preprocessing, and transparent model evaluation.
- **Achieving 99.3% accuracy** using a **Random Forest Classifier** with hyperparameters: n_estimators=100, max_depth=5, and random_state=42**.**.
- Enhanced workflow reproducibility and collaboration by integrating version-controlled pipelines with automated re-execution**, reducing manual overhead by 80%.**.
- **Future Work:** Planning to implement enhanced visualization features and advanced CRUD operations while expanding AI capabilities through broader pgai integration.

Automating Weather Data Collection with Apache Airflow | Lead Developer (DEMO**)** **(**GitHub Link)

- Designed and deployed **an automated ETL pipeline** using **Apache Airflow** to extract real-time weather data from the '**wttr.in**' API, transform it, and load it into a PostgreSQL database hosted on AWS RDS.
- **Technologies: Apache Airflow, PostgreSQL, AWS RDS, Astronomer, DBeaver, Python**.
- Developed DAGs featuring automated table creation, data extraction via SimpleHttpOperator, transformation of weather attributes (temperature, humidity, observation time, and description), and data loading into PostgreSQL.
- Monitored and verified pipeline execution through **DBeaver** and deployed workflows via **Astronomer**, ensuring robustness and scalability for downstream analytics and visualization**.**
- Achieved seamless daily automation, **reducing** manual data collection **efforts by 90%**.
- **Future Work:** Implementing enhanced logging, **alerting mechanisms**, and expanding data collection to multiple locations.

EXTRA CURRICULAR ACTIVITIES & ACHIEVEMENTS

- Actively contributed to renowned repositories, including **Langchain, Gradio** and **Wagtail**, enhancing their functionalities and supporting community-driven development.
- Successfully completed **Hacktoberfest** by having **4 Pull/Merge Requests accepted**, demonstrating commitment to collaborative open-source contributions.
- Shortlisted among only 25 candidates for the prestigious Data Science program **at IIT Kharagpur**, a highly selective achievement in a competitive field**.**

TECHNICAL SKILLS

Languages & Tools: Python, SQL, Docker, Git, Flask, Streamlit, MLFlow, Apache Airflow, Astronomer , AWS, Github Actions
AI/ML: scikit-learn, XGBoost, TensorFlow, PyTorch, scikit-learn, Hugging Face Transformers, LangChain, llamaindex
Generative AI: RAG, RLHF, LLM Fine-tuning, Attention Mechanisms
Data Engineering: NumPy, Pandas, BeautifulSoup, Playwright
Statistical Techniques: Hypothesis Testing, Statistical Significance, Regression, Clustering, Model Optimization
Core Competencies: Model Training and Deployment**, Transformer Models, API Integration,** OOP, Feature Engineering
Soft Skills: Communication and Interpersonal Skills, Team Leadership, Analytical Thinking

EDUCATION

Punjab Engineering College Chandigarh
Bachelor of Technology in Computer Science Engineering 2022 – *Present*
City Convent School (Percentage: 95.00%) U.S. Nagar, Uttarakhand
Senior Secondary Education, CBSE *2020– 2021*
City Convent School (Percentage: 90.00%) U.S. Nagar, Uttarakhand
Secondary Education, CBSE *2018 – 2019*