HRISHIKESH GHOLE

+91 9004041292 hrishikeshghole@gmail.com linkedin.com/in/hrishikesh-ghole github.com/hrishi1998

PROFESSIONAL SUMMARY

Aspiring Data Engineer proficient in Python, AWS, and Big Data technologies. Experienced in data analysis and real-time processing, passionate about solving complex business problems and driving strategic decisions through meaningful data insights.

EDUCATION

PG-Diploma in Big Data Analytics • (CDAC) **Sep 2023 - Feb 2024**
Python, SQL, Big data Technologies, Cloud, ML,Data Visualization

Master in Science, Information Technology • Mumbai **Nov 2020 - Apr 2022**
Machine Learning & AI, Data Science, Statistics, Cloud Computing, Linux

Bachelor of Science, Computer Science • Mumbai **Jun 2017 - Nov 2020**
C, Java, Data Structures, Algorithms, OS, Version Control, DBMS, Networks

TECHNICAL SKILLS

Cloud Platforms: Amazon (S3, EC2, IAM, RDS, CloudWatch, CloudFormation, EMR, Athena, SQS)
Big Data Tools: ETL Pipeline, Big Data, Hive, Spark, Hadoop, MapReduce, Confluent Kafka
Languages: Python, Java, SQL, Go, Bash/Shell Scripting
Libraries: Pandas, NumPy, Scikit-learn, PySpark, Streamlit
Version Control: Git, GitHub Actions (CI/CD)

PROJECTS

Real-time Fraud Detection and Analytics Pipeline **1 Month**
Kafka, Spark, AWS EMR, Tableau, Python, GitHub Actions

- Developed a real-time fraud detection system for credit card transactions using **Kafka** for data streaming and **Apache Spark** for processing large-scale transactional data.
- Implemented **Amazon EMR** for efficient data processing, validating data quality, and applying transformations for fraud detection.
- Classified transactions as **fraudulent or genuine** based on predefined rules and stored processed data for further analysis.
- Utilized **Tableau** to visualize transaction data, creating interactive dashboards for trend analysis and anomaly detection.
- Automated pipeline tasks using **Python scripts (PyScript)** and **GitHub Actions (CI/CD)** for continuous deployment and testing.
- Deployed and managed infrastructure with **AWS CloudFormation**, ensuring scalable and repeatable deployment processes.

Calorie Burning Prediction Using Machine Learning **1 Month**
Python, Pandas, NumPy, Scikit-learn, Streamlit, RandomForestRegressor

- Developed a Random Forest Regressor model to predict calories burned based on age, height, weight, duration, heart rate, and body temperature, and trained it on 15,000 entries with a mean absolute error of 2.03.
- Deployed the model as a Streamlit web application for user input and prediction.

Flower Classification using Deep Learning **1 Month**
Python, Pandas, NumPy, Scikit-learn, Streamlit, RandomForestRegressor

- Developed a deep learning model to classify images of five flower types (Rose, Tulip, Sunflower, Daisy, Dandelion) with 80predictions.

CERTIFICATIONS

- Technical Certification: MTA: Security Fundamentals - Certified