Aashi Chahal

Navi Mumbai, Maharashtra | aashi.cdacmum.aug25@gmail.com | 8791627341 | Github | LinkedIn

Summary

Strong understanding of Big Data and Machine Learning concepts, with hands-on experience in Python, SQL, PySpark, Hive, and AWS. Strong foundation in data analysis, feature engineering, and machine learning through academic projects. Worked on an AI-Based Maritime Port Intelligence System, a Netflix Viewership Analytics Pipeline, and an NLP-based Duplicate Question Detection system using Random Forest, LightGBM, and XGBoost. Passionate about learning data-driven technologies and applying analytical skills to solve real-world problems.

Technical Skills

- **Programming:** Python, SQL
- **Machine Learning:** Regression, Classification, Feature Engineering, Model Evaluation, scikit-learn
- **Big Data & Cloud:** PySpark, Spark, Hive, AWS, Linux, Git
- **Data Analysis:** Pandas, NumPy, Exploratory Data Analysis
- **Statistics:** Descriptive & Inferential Statistics, Hypothesis Testing
- **Generative AI:** Retrieval-Augmented Generation (RAG), LLM-based feature extraction
- **Visualization:** Tableau, Power BI, Matplotlib

Academic Projects

AI-Based Maritime Port Intelligence System
Python, Machine Learning, RAG, LLMs, Streamlit, Power BI

- Developed ML models (Random Forest, LightGBM, XGBoost) to predict port congestion, vessel delays, and berth allocation feasibility.
- Built end-to-end ML workflow including preprocessing, feature engineering, model training, and inference.
- Implemented policy-aware decision making using RAG with TF-IDF and cosine similarity.
- Designed intelligent agent pipeline combining ML predictions with regulatory context for operational recommendations.
- Deployed trained models in Streamlit application for real-time scenario evaluation.

Netflix Viewership Analytics Pipeline
AWS, Python, SQL, PySpark, Big Data, Tableau

- Designed a scalable AWS analytics pipeline to ingest and transform large datasets using PySpark.
- Performed data aggregation and quality validation to prepare clean datasets for analytics and future ML workflows.
- Built an event-driven ETL workflow using S3, Lambda, Glue Workflows, and Athena, with SNS email notifications triggered when new data is uploaded to S3.
- Automated pipeline execution and monitoring using cloud-native AWS services.
- Delivered Tableau dashboards analyzing content trends by genre, rating, and country.

Duplicate Question Detection Using NLP
Python, NLP, Machine Learning

- Built NLP-based classification system to detect semantic similarity between question pairs.
- Performed feature engineering including similarity scores, word overlap, and length-based metrics.
- Trained Random Forest, LightGBM and XGBoost models evaluated using accuracy and F1-score.

Education

PG Diploma in Big Data Analytics – CDAC Mumbai — Aggregate: 78%

B.Tech in Computer Science & Engineering – AKTU — CGPA: 7.5

Certifications

- AWS Academy Graduate – Data Engineering (2025)
- AWS Academy Graduate – Cloud Foundations (2025)
- AWS Academy Graduate – Generative AI Foundations (2026)
- Python Programming with DSA – YBI Foundation