

# Aashi Chahal

Navi Mumbai, Maharashtra

[aashi.cdacmum.aug25@gmail.com](mailto:aashi.cdacmum.aug25@gmail.com) | 8791627341 | [linkedin.com/in/aashichahal23](https://www.linkedin.com/in/aashichahal23)

## Professional Summary

---

Motivated analytics and machine learning professional with hands-on experience in data analysis, machine learning, big data processing, and cloud-based pipelines. Strong foundation in Python, SQL, PySpark, and statistical techniques, with project experience spanning logistics analytics, NLP, and large-scale data pipelines. Skilled in feature engineering, model development, exploratory data analysis, and Generative AI using RAG and LLM-based extraction. Passionate about transforming raw data into actionable insights and solving real-world problems through data-driven solutions.

## Technical Skills

---

**Programming:** Python, SQL

**Machine Learning:** Regression, Classification, Feature Engineering, Model Evaluation, scikit-learn

**Deep Learning:** Neural Network fundamentals, TensorFlow (basic exposure)

**Big Data:** PySpark, Spark, Hive

**Data Analysis:** Pandas, NumPy, Exploratory Data Analysis

**Statistics:** Descriptive & Inferential Statistics, Hypothesis Testing

**Cloud & Tools:** AWS, Linux, Git

**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 scalable AWS analytics pipeline to ingest and transform large datasets using PySpark.
- Performed data aggregation and quality validation for downstream analytics and ML workflows.
- Automated workflows and monitoring using cloud-native AWS services.
- Delivered Tableau dashboards analyzing 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 Logistic Regression and XGBoost models evaluated using accuracy and F1-score.

## Machine Learning & Data Workflow Experience

---

- Data preprocessing and feature engineering

- Model training, evaluation, and optimization
- Large-scale data processing using PySpark
- AWS-based pipeline orchestration
- Exploratory Data Analysis
- Applied Generative AI using RAG
- Translating business problems into data-driven solutions

## Education

---

**PG Diploma in Big Data Analytics** – CDAC Mumbai  
**B.Tech in Computer Science & Engineering** – AKTU

Aggregate: 78%  
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