

# Harshit Arora

Navi Mumbai, Maharashtra | 589harshitarora.2020@gmail.com | 9215847003 | LinkedIn | Tableau Portfolio | GitHub

## Summary

Aspiring Data Science Professional experienced in **Python**, **SQL**, and **Java** with strong skills in **ML Modeling**, **Predictive Analytics**, and **Statistical Analysis**. Proficient in building end-to-end ML pipelines using **scikit-learn**, **Pandas**, **NumPy**, and **statsmodels**, with expertise in **EDA**, **feature engineering**, and **data visualization**. Hands-on with **PySpark**, **Hive**, **Tableau** and **Power BI** aiming to deliver scalable, data-driven solutions.

## Skills

- **Programming & Databases:** Python, SQL, Java, MySQL
- **Machine Learning:** Regression, Classification, Feature Engineering, scikit-learn
- **Data Analysis:** Pandas, NumPy, Exploratory Data Analysis (EDA)
- **Statistics:** Descriptive & Inferential Statistics, Hypothesis Testing
- **Big Data & Cloud:** PySpark, Hive, AWS, Docker, Linux, Git
- **Visualization:** Tableau, Power BI, Matplotlib, Seaborn

## Projects

### MaritimePort AI

*Python, Machine Learning, RAG, LLM, Docker, Streamlit*

- Developed ML models (Random Forest, LightGBM, XGBoost) to predict port congestion, vessel delays, and berth allocation feasibility.
- Built a policy aware intelligent agent using RAG and LLM based feature extraction to deliver regulation compliant operational recommendations.

### Netflix Viewership Analytics Pipeline

[Tableau Dashboard](#)

*AWS, Python, PySpark, Big Data, Tableau*

- Built an event driven AWS pipeline to ingest, clean, and transform Netflix viewership data using PySpark.
- Automated workflows and monitoring on AWS and delivered insights via Tableau dashboards across genres, ratings, and countries.

### Telecom Customer Churn Prediction Model

[Live App](#) – [GitHub](#)

*Python, Pandas, scikit-learn, Flask, Docker*

- Built ML pipeline on 7K+ records, improved accuracy by 15%.
- Trained Logistic Regression, Random Forest & XGBoost; achieved 80% accuracy.
- Deployed real-time prediction app using Flask & Docker

## Education

### Centre for Development of Advanced Computing

Aug 2025 – Feb 2026

*PG - Diploma in Big Data Analytics*

Aggregate: 83%

### Maharshi Dayanand University

2020 – 2024

*B.Tech. in Computer Science & Engineering*

Aggregate: 78%

## Experience

### Freelance Web Developer

Jun 2023 – Aug 2023

*Self-Employed*

- Built a responsive portfolio website using HTML, CSS, and JavaScript.
- Engaged directly with clients to customize site features based on iterative feedback.

## Certifications

- AWS Academy Graduate – Generative AI – Amazon Web Services – 2026
- AWS Academy Graduate – Data Engineering – Amazon Web Services – 2025
- AWS Academy Graduate – Cloud Foundations – Amazon Web Services – 2025
- Google Advanced Data Analytics Professional Certificate – Coursera – 2023
- Python for Data Science – NPTEL (IIT Madras) – 2023

## Notable Achievements

- GATE 2025 Qualified – Computer Science
- Built and deployed multiple end-to-end Machine Learning projects.