

# ARYAN RATHORE

[LinkedIn](#) ◇ [GitHub](#) ◇ [LeetCode](#)

Data Scientist at Vellore Institute of Technology

+91-9685071745 ◇ aryanrathore13572002@gmail.com ◇ khatiwala tank, indore M.P. 452014

## SKILLS

Python, Machine Learning, Data Analysis, Data visualization, BigQuery, Pandas, NumPy, SQL, Matplotlib

## EDUCATION

**VIT University**

CGPA: 9.04

*B.Tech in Computer Science and Engineering*

*Expected May 2025*

**Divya Convent School, Indore**

Percentage: 67.40

*Class XII*

*07/2021*

**Agarwal Public School, Indore**

Percentage: 63.80

*Class X*

*07/2018*

## EXPERIENCE

**Artificial Intelligence Engineer, BharatPe**

*11/2024 – 12/2024*

- Resolved 99.5% of discrepancies amounting to 1 Billion INR Approx in 2022–23 financial data via reconciliation.
- Partnered cross-functionally to align reconciliation with business goals, boosting stakeholder trust.
- Detected 80% of 2020–21 discrepancies via edge-case analysis, improving anomaly detection by 40%.
- Automated SQL query optimization 2018–20, cutting runtime 90% 1hr to 6min and saving 100+ hours.

**Technology:** Python, SQL, BigQuery, GCP, Software Development Lifecycle, operational efficiency

**Data Science Internship, Jio**

*04/2023 – 07/2023*

- Built ArangoDB knowledge graph in Pandas analyzing 38K+ medicines, reducing prescription costs by 95%, 2463 INR saved in some cases.
- Engineered NLP ontology framework extracting 2000+ semantic relationships with 99% schema consistency.
- Prototyped scalable graphs through cross-functional collaboration, enhancing predictions.

**Technology:** Python, ArangoDB, proto3, Numpy, Pandas, Matplotlib, Seaborn, Sklearn, scikit-learn, GPT

## PROJECTS

**AQI analysis and model India**

*03/2023*

- Boosted predictive accuracy to R2: 0.92 vs. 0.81 baseline using XGBoost on 100K+ entries, slashing MAE by 24%, 21.22 vs. 27.9.
- Identified PM2.5/PM10 68% and NO2 as top pollutants amid 15% national AQI decline 2015–2020.
- Ranked Gujarat AQI 156 and Delhi 148 as most polluted; Mizoram 32 and Aizawl 34 as cleanest via 500+ station analysis.

**Technology:** Python, OOP, sklearn, matplotlib, NumPy, Pandas, seaborn, tkinter, supervised Learning

**Link and Results:** [Air Quality Analysis and Prediction for Indian Cities and States](#)

**Language-Agnostic Ontology Extension Framework**

*06/2023 – 11/2023*

- Built NLP-driven knowledge graphs to map 200+ multilingual terms into numerical IDs, enabling language-agnostic machine interpretation.
- Automated semantic relationship extraction achieving 99% consistency in cross-lingual data structuring.
- Designed ID-based vector system for real-time human-PC translation of unstructured text.

**Technology:** Python, NumPy, Pandas, natural language processing, Knowledge Graphs, ArangoDB, Large Language Models, transformers

**Link and Results:** [Language-Agnostic-Ontology-Extension-Framework](#)

## CERTIFICATIONS

- Applied Machine Learning with Python, University of Michigan *01/2023*
- Data Analysis with Python, IBM *02/2023*
- Fundamentals of Data Analytics, LEAPS by Analyttica *03/2023*

## EXTRA-CURRICULAR & ADDITIONAL INFORMATION

- Solved 450+ questions and ranked in the top 3% in 2023 and 7% in 2024.
- Among 5% 150/3000 in "Immersion 2021," a competitive DSA event at VIT Bhopal.
- Hobbies:** Fitness Enthusiast, Video Games, Anime, K-Dramas, Coding
- Languages:** Fluent in English, native Hindi speaker.