

Rohan Kumar Singh

📞 +91 9465287233

✉️ rohan0808singh@gmail.com

LinkedIn

Github

Education

Vellore Institute of Technology, Bhopal Integrated Mtech AI (CGPA: 8.17)	Sep 2022 – May 2027 Bhopal, Madhya Pradesh
BCM School Senior Secondary(12th), CBSE (Percentage: 80.2%)	June 2021 – July 2022 Ludhiana, Punjab
BCM School Secondary(10th), CBSE (Percentage: 82.8%)	June 2019 – July 2020 Ludhiana, Punjab

Experience

Data Analyst Intern <i>Mushin Innovation Labs</i>	Oct 2025 – Dec 2025 Gurugram, Haryana
<ul style="list-style-type: none">Executed Parquet data preprocessing in Databricks using PySpark, ensuring schema consistency, data integrity, and readiness for downstream analytics.Built Databricks-managed tables and implemented automated data ingestion into MySQL using Python-based connectors and SQL-driven extraction.Tackled complex data challenges to enhance problem-solving abilities.	

Projects

Covid-19 Data Pipeline Project <i>Data Engineering</i>	Apr 2025
<ul style="list-style-type: none">Achieved pipeline success rate consistently above 98% with robust error handling and retries.Ingested data from multiple sources into Azure Data Lake for centralized storage.Transformed raw data using ADF Data Flows and Databricks Notebooks, preparing it for analytics and reporting.Loaded processed data from Azure Data Lake to Azure SQL Database.	
Olympics Analytics <i>Data Analysis</i>	Aug 2025
<ul style="list-style-type: none">Developed and deployed an end-to-end data pipeline using Azure toolsAutomated data ingestion and cleaning from APIs and GitHub, organizing raw and processed datasets in Azure Data Lake.Improved data transformation accuracy to 99.8% by implementing validation checks and clean data pipelines for analytics.Increased workflow automation coverage by 75%, improving operational efficiency and reducing pipeline failures.	
Gold Price Prediction using Machine Learning <i>Machine Learning</i>	Dec 2025
<ul style="list-style-type: none">Built an end-to-end machine learning regression model in Python to forecast gold prices using historical market data and feature engineering.Performed data preprocessing, exploratory analysis, feature scaling, and outlier handling to improve model performance.Evaluated and optimized models including Lasso Regression, RandomForest, and XGBoost, achieving high predictive accuracy and feature importance insights.	

Technical Skills

Languages: Python, Java, HTML, CSS

Frameworks & Libraries: NumPy, Pandas, PySpark, Hadoop

Databases: MySQL

Tools & Others: Git, GitHub, Azure Data Factory, Azure Data lake, Schema design, CI/CD concepts

Achievements & Certifications

- Solved 350+ Data Structures & Algorithms (DSA) problems on LeetCode and GFG.
- Java Programming – Udemy, May 2023
- Cloud computing – NPTEL, Aug 2024
- Big Data Engineering Bootcamp – Udemy, Sep 2025