

SHIVRAJ KUMAR

[in LinkedIn](#) | [790-338-7090](tel:790-338-7090) | shivraj कुमार2255@gmail.com | [GitHub](#) | [YouTube](#)

Skills

- **Programming, Scripting & Tools:** Python, SQL, PySpark, Docker, WinSCP, PuTTY, Git
- **Data Engineering & Big Data:** ETL Automation, Airflow, Kafka, Apache Oozie, Hadoop, Impala, Data Modelling
- **Data Analysis & Visualization:** Tableau (Desktop, Prep, Server, REST API), SAP BO, Excel, Power BI, EDA
- **Machine Learning & AI:** Supervised & Unsupervised Learning, Scikit-learn, Pandas, NumPy, Model Development
- **Soft Skills:** Communication, Collaboration, Storytelling with Data

Experience

Data Engineer Cotiviti *Pune, India* **06/2024 - Current**

- **Developed robust, scalable ETL pipelines** using **PySpark, Python, and SQL**, automating ingestion and transformation of large datasets from multiple sources, resulting in an 80%+ improvement in processing speed.
- **Designed modular and reusable PySpark scripts**, integrated with **Airflow** and **Oozie**, to orchestrate daily and weekly data workflows with automated retries, logging, and dependency management.
- **Integrated version control and CI/CD best practices** using **Git**, streamlining the deployment and testing of PySpark jobs across dev, QA, and prod environments.

Associate Data Analyst Cotiviti *Pune, India* **07/2023 – 05/2024**

- **Created dashboards** in **Power BI and Tableau**, delivering insights to 100+ clients monthly across various business units, driving improved data-driven decisions.
- **Embedded analytics solutions** into external websites and portals using **Power BI Embedded** and **Tableau JavaScript API**, enabling secure, real-time access to interactive reports for clients and stakeholders.
- **Designed and maintained robust workflows using Tableau Prep**, preparing and transforming complex datasets for visualization and analysis in Tableau dashboards.

Machine Learning Intern Internship Studio *Remote* **01/2022 - 02/2022**

- Implemented a predictive model to estimate YouTube ad view counts by analyzing video metrics such as duration, likes, comments, and engagement, incorporating feature engineering to enhance model performance.
- Implemented and fine-tuned machine learning models, including SVM, Decision Trees, Random Forest, and ANN, to optimize prediction accuracy and deliver reliable results.

Education

Bachelor of Technology NIT Raipur *8.28 CPI* **07/2019 - 04/2023**

Projects

SafeRoute AI — Safety Indicator

Tech Stack: Python, PySpark, Apache Iceberg, Hadoop, SQL, Git, Docker, Apache Kafka

- Built modular **PySpark** jobs to automate preprocessing, schema validation, null handling, and anomaly detection, enhancing the processing of crime and accident data.
- Leveraged **Apache Iceberg** on **Hadoop** for versioned, ACID-compliant data lakes, and containerized key components using **Docker**, while tracking code with **Git** to ensure reproducibility and scalability.
- Used **Apache Kafka** to build a fault-tolerant, distributed data pipeline for reliable real-time safety data processing.

Certifications

- Machine Learning Specialization (Coursera)
- Data Analysis with Python (Coursera)
- SQL Intermediate (Hackerrank)

Awards

Monthly Performance Award at Cotiviti

06/2024