

Suraj Raghuvanshi

Data Engineer

✉ surajraghuwanshi05@gmail.com ☎ 8120742013 📍 Bhopal 🔗 LinkedIn 🐙 GitHub

SUMMARY

A motivated Data Engineer with a year of hands-on experience in designing, building, and optimizing data pipelines and workflows. Proficient in Apache Spark and PySpark for big data processing, with a strong command of Python and SQL for data manipulation and analysis. Familiar with Git for version control and adept at building interactive dashboards using Streamlit. Has foundational knowledge of Machine Learning techniques and models, enhancing the ability to implement data-driven solutions change this summary for data analyst

EXPERIENCE

iDataMinds

Data Analytics and Engineering Intern | 07/2024 – Present

- Built hands-on experience with database management systems (DBMS), employing normalization techniques, data modeling, and schema design to create efficient, scalable data storage solutions.
- Utilized data warehousing techniques to develop large-scale data repositories, streamlining complex querying and reporting needs.
- Leveraged advanced SQL for data extraction, transformation, and analysis, supporting impactful business decisions with actionable insights.

Zidio Development

Data Science and Analyst Intern | 04/2024 – 06/2024

- Enhanced model accuracy through advanced statistical analysis and feature engineering, improving data quality for predictive modeling.
- Streamlined data handling processes by implementing engineering solutions in collaboration with cross-functional teams, facilitating faster data access for stakeholders.

PROJECTS

Sales Insights and Analytics Dashboard 📄

- Designed and implemented a Bronze, Silver, and Gold data architecture, improving data processing efficiency with clean, consistent data structures.
- Developed robust SQL procedures to automate routine data operations, streamlining workflows and saving the team substantial time on manual processes.
- Enhanced query-based insights with a Streamlit app, enabling flexible data querying and filtering that significantly boosted insight generation.

Employee Attrition Insights with PySpark 📄

- Designed and developed a scalable data ingestion pipeline in PySpark using Databricks to handle large-scale employee attrition data from multiple sources, ensuring high performance and reliability.
- Implemented advanced data transformation workflows, including indexing, one-hot encoding, and feature engineering, to standardize and prepare datasets for downstream analytics and machine learning in a collaborative Databricks environment.
- Documented the end-to-end pipeline architecture, data lineage, and transformation logic for operational transparency and maintenance.

Fraud Detection system (Self Work) 📄

- Strengthened fraud detection capabilities with a machine learning model, using data engineering techniques to identify patterns in transaction behavior.
- Implemented exploratory data analysis on **10,000+ transactions**, employing pandas and visualization libraries to uncover transaction patterns and fraud signals.

EDUCATION

B.TECH (Agricultural Engineering)

Jawaharlal Nehru Agricultural university
CGPA -7.69

2019 – 2023
Jabalpur

Class 12th

Samdariya public school
Percentage -75.8

2018
jabalpur

SKILLS

Data Analysis — Python, EDA, Feature Engineering, Predictive Modeling, pandas, Excel

Data Engineering — SQL, ETL, Data Architecture, Data Cleaning, Data Pipelines, Numpy, Spark, databricks

Data Visualization — Streamlit, Matplotlib, Seaborn, Power BI Dashboards

Machine Learning — Scikit-Learn, Predictive Modeling, Neural Networks, NLP

Additional Tools and Technologies — Git and GitHub, MongoDB, Flask, HTML, CSS, JavaScript,