Rakesh Chaudhary

Data Analyst/Engineer | Python Developer

✉ arcsle09plus@gmail.com

📞 9998006038/7016731487

👤 Profile

An accomplished technology professional having about 5 years of experience in all phases of agile software development lifecycle. Passionate about Data Engineering technologies and delivery of effective solutions through creative problem-solving.

🎓 Education

Master of Science - Financial Mathematics, *Maharaja Sayajirao University of Baroda* ⧉
Jun 2016 – Jun 2018

📇 Certification

- Certified SAFe® 6 Scrum Master ⧉
- Google Cloud Skills ⧉
- Databricks Lakehouse Fundamentals ⧉
- HackerRank ⧉

💼 Professional Experience

Senior Analyst, *NielsenIQ*
Nov 2018 – present

- An Integral Part of Global Innovation Team for Client.
- Communicate with onshore product team to understand the problem statement and requirements.
- Design and develop the data engineering solutions.
- Review codes of peers and junior developers.

🧠 Technical Skills

Programming Languages ● ● ● ● ○
Python | SQL

Databases ● ● ● ● ○
MySQL | Oracle | Snowflake

Data Transformation Tools ● ● ● ● ●
Pyspark | Pandas | Polars

DevOps Tools ● ● ● ● ○
Git | Bitbucket | Docker | Jenkins

Cloud Services ● ● ● ● ○
Databricks | Blob | ADF | EC2 | S3 | BigQuery

Project Tracking Tools ● ● ● ● ○
JIRA and Confluence

Machine Learning Models ● ● ● ● ○
Tree-Based Classifiers

📁 Projects

Retailer Sales Data Classification
  Tech: Python, Sklearn, polars, pandas & Oracle SQL

- Acting Lead Developer to implement classifier for highly imbalanced binary classification problem.
- Wrote analytical SQL queries to extract the long historical time series data.
- Performed EDA and used various techniques to optimized the model performance for each country dataset.
- Reduced ~50% manual validation efforts by deploying the models for each country.

Global Item Xcode App
  Tech: *Databricks, Pyspark, Azure-Blob, python-polars*

- Developed the data extraction and transformation pipelines to build country level product info vocab using pyspark using databricks platform
- Developed the string match module to produce top 4 matches for respective user input product description parsed from Django UI.
- Reduce the manual item coding efforts to 30%

Ecommerce Data Acquisition
  Tech: *Python, Scrapy, Selenium, RapidFuzz, EasyOCR, Pandas, S3 & EC2*

- Developed Various Web Scrapers to scrape product images, price, sales, units and categories data (Retailers and Google sites).
- Optimized the OCR process of images using python multiprocessing.
- The web scrapers process & store data for ~2 million products every week.