Rakesh Chaudhary

Senior Analyst

✉ arcsle09plus@gmail.com
📞 9998006038/7016731487
↖ Detailed CV

👤 Profile

An accomplished Technology & Research Professional and Certified Scrum Master(SAFE) having more than 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 and Research Team.
- Communicate with onshore product team to understand the problem statement and requirements.
- Designed and Developed Data Engineering solutions.
- Developed the technical POCs for large scale data operations and validation activities.
- Well versed with NielsenIQ Data (POS,CPS,PROMO) facts and Dimensions.
- Review codes of peers and junior developers.

🧠 Technical Skills

Programming Languages ●●●●○
Python | SQL

Databases/DWH ●●●●○
MySQL | Oracle | Snowflake

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

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

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

Project Tracking Tools ●●●●○
JIRA and Confluence

Machine Learning Models ●●●●○
Tree-based Models | Azure ML (StackEnsambler)

Data Visualization Tools ●●●○○
PowerBI | Plotly

📂 Projects Summary

Retailer Sales Data Classification Tech: Python, Sklearn, Polars, Pandas, Oracle SQL & Azure ML

- 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 Retailer Data Health Dashboard Tech: Oracle SQL, ADF, *Databricks(Pyspark), Azure-Blob & PowerBI*

Used 2 different data sources (oracle database and remote on-prem storage) for data ingestion using ADF.

Developed ,deployed & schedule rule-based transformation pipelines using Databricks - Pyspark.

Stored the output of the pyspark pipelines using parquet format in azure blob storage.

Built store level and item level KPIs for Retailers FMCG Data.

The ETL data pipelines processes 20 million records weekly.

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

- Developed and schedule the data extraction and transformation pipelines to build country level product info vocab using pyspark in databricks.
- The data loaded/written to azure blob as parquet files.
- Developed the string match module using polars and jaro-winkler algorithm to produce top 4 matches for respective user input product description parsed from end users via Django Web App
- Wrote unit test cases for string match module.
- The country level data gets auto-refreshed every week.
- Reduce the manual item coding efforts to 30%

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

- Developed,deployed and scheduled 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.
- wrote unit test cases for OCR and webscraping modules.
- The web scrapers process & store data for ~2 million products every week.
- Developed and deployed various webscrapers for Italy based retailer websites in order to download the promotional flyers(Avg 1000 flyers per week).