Rajat Sharma

+91-7451989340 rajatraj9896@gmail.com

Work Portfolio| GitHub Profile | LinkedIn Profile |Kaggle

Professional Summary

Working as a data scientist with over four years of experience and using data to solve real world problems. I have a good grasp on Python and SQL and working knowledge of Tableau. My typical day involves collaborating with clients to understand their pain points and going through the technical details to give them the best possible solution.

Skills and Tools

- Programming Languages: Python (Data Analysis & Visualization, AI/ML)
- Databases: SQL (Data Extraction & Manipulation & Validation, Feature Creation)
- Data Visualization: Tableau, MS Excel, MS Power Point
- Data Science: Sklearn, Pytorch, Keras, Natural Language Processing, Large Language Models
- Gen AI: Gemini, RAGs, Agents and Agentic AI
- Other Tools: Git (GitHub, GitLab), Jira, Azure Cloud, Data robot

Experience

Senior Analyst/Data Scientist | Merkle, Dentsu | Sept 2024 – Present

The key responsibilities revolved around:

- Connecting with clients to understand the use case and formulate a plan and blueprint.
- Feature creation and data gathering and from vast sources(billions) of data.
- Data cleaning, processing and feature engineering to finally building AI/ML models.
- Showcasing results and benefits achieved from our models before deploying them.

Associate Data Scientist | Factspan Analyst | June 2022 – Aug 2024

The key responsibilities revolved around:

- Connecting with clients to understand the use case.
- Data gathering from various data sources using SQL skills.
- Data cleaning, processing and feature engineering to finally building AI/ML models.
- Taking sign off from client and giving the product as a tableau Dashboard.

Subject Matter Expert (Chegg India) | Dec 2018 - Jun 2020 (Freelance Electrical & Electronics Eng)

The key responsibilities revolved around:

- Helping students with their doubts and assignments (A freelance online work profile).

Discussion of Use Cases Worked (2022 – Present)

Severity Calculation for Customer Pain Points(Classification, GenAI, Ranking)

- Aim: Tackle the most severe pain points to customer on priority hence give weightage accordingly.
- Value: Adds to better CX hence customer retention and better handling of issues.
- Process: Attaching pain points to the contacts made from Level_1(very broad) to Level_5(most granular) this was done using fine-tuned MPnet model for the pre decided pain point levels (given by business owner).
- Process: Creating metrics like cancellations and refunds, # contacts, span of contacts, # touches, customer feedback score, escalation, sentiment to compare and calculate severity of a pain point. Then making their severity scores at L5 levels using equal weightage to all the features and then giving weightages based on feature importance, finally discussing with business.

Text Extraction from Images(GenAI, OCR)

- Aim: Extract SKU descriptions and quantities from unclear customer notes, map to system SKUs, and export to Excel.
- Value: Streamlines order processing by automating text extraction and SKU matching.
- Process: Pre-processed images, enhanced prompts with Gemini 2.5, converted results to Data Frame, and performed similarity matching with internal SKUs, returning top 2 matches.
- Result: Model achieved automated extraction with more than 95% accuracy.
- Challenge: Illegible notes with symbols, short hands, numeric, and dimensions complicated text recognition and matching.

Customer Clustering:

- Aim: Segment customers into cohorts based on behaviour to identify fraud and focus on profitable ones.
- Value: Eliminates fraud, enhances targeting of high-value customers.
- Process: Monitored purchasing habits, progressed to full-fledged modelling.
- Result: Model in progress, accuracy not finalized.
- Challenge: Detecting fraud customers creating new accounts post-ban remains unresolved.

Turn Around Time Prediction(Regression):

- Aim: Predict time between surgeries in operation rooms.
- Value: Boosts team efficiency in surgical scheduling.
- Process: Used extreme gradient boosting for regression on a complex dataset.
- Result: 70% accuracy, with 10–20-minute error margin.
- Challenge: Managing 300+ features, millions of records, and hierarchical variables.

Refill Request Volume Forecast(Time Series):

- Aim: Forecast daily pharmacy refill request volume.
- Value: Improves efficiency, inventory, and customer experience.
- Process: Applied SARIMA for time series forecasting with feature engineering.
- Result: 92% accuracy, MAE of 60 for ~600 daily requests.

- Challenge: Simple dataset required extensive feature engineering.

Pyxis Min Max Levels Prediction(Time Series):

- Aim: Predict min/max drug levels for Pyxis dispensing machines.
- Value: Optimizes inventory for high-cost medicines.
- Process: Used FB Prophet for top 50 medicines, after testing moving averages.
- Result: 85% mean accuracy with 90% confidence intervals.
- Challenge: Handling predictions for over 2,000 medicine series.

Radiology Care and Coordination(GenAI)

- Aim: Detect patient severity and incidental findings from radiology reports, provide alerts, and trend analysis.
- Value: Enhances patient care, automation saves costs.
- Process: Used Open AI endpoint with data masking for privacy, built a dashboard.
- Result: Delivered dashboard with severity trends and cost savings.
- Challenge: Ensuring data privacy and accurate keyword-based severity detection.

Education:

- **Masters** - MTech, NIT ALLAHABAD, MNNIT ALLAHABAD
- **Bachelors –** BTech, AKTU, ABES Engineering College

Extra Activities and Interests

- An active super-host on Airbnb with multiple listed properties.
- Stock market analysis and active investing/trading.
