Parth Jain

Data Scientist Specializing in Data Analysis and ETL Pipelines.

parth.hc.jain@gmail.com | linkedin.com/in/parthhjain | 848-437-1922 | Portfolio | GitHub | Sunnyvale, CA

EXPERIENCE

AI Engineer (Generative AI) Intern | **Foundational AI** | DE, USA (Remote) Oct 2023 - Present

- Designed platform architecture using Azure Cloud to automate business operations with customizable GenAI tools.
- Built FastAPI-based server side logic for no-code RAG building across many use cases for 5x user productivity.
- Executed advanced prompting and retrieval techniques with langchain, for better LLM output with 40% reduced cost.
- Fine-tuned and deployed ML model with Azure ML to categorize user queries, providing customer sales cycle status.
- Performed hyper-parameter tuning for the Intent model using Optuna, SetFIT, and PyTorch to boost accuracy by 10%.

Data Science Research Assistant | **RUCI lab** | NJ, USA Aug 2022 - Sept 2023

- Developed deep learning framework, boosting accuracy up to 10% for user-specific handwritten OCR applications.
- Built ETL pipelines and scraped data, to ensure high-quality data for machine learning and data analysis application.
- Conducted research on NLP topics i.e. machine translation and sentiment analysis to report findings in white papers.
- Utilized Python for data analysis and visualization on text and finance data, to deliver key insights for client queries.
- Designed Jupyter labs to give a hands-on approach to AI, for Prof. Jim Samuel's course, increasing enrollment by 50%.

Data Science (R&D) intern | WINLAB | NJ, USA June 2022 - Aug 2022

- Collaborated with engineers to build IoT-ML system to identify users with motion patterns, for safe use of medicines.
- Developed data pipeline to fetch data over MQTT protocol and load to S3 with Python cutting process time by 16%.
- Extracted features from time series data and tuned random forest model with sk-learn achieving accuracy of 91%.

Data Analyst Intern | Mahavir Coconut Industries | KA, India Dec 2020 - Feb 2021

- Built dashboard with Tableau to summarize daily user purchase patterns to aid stakeholders' decision-making.
- Interpreted business insights from trends resulting in budget optimization with the aid of data visualization.
- Restructured data models and optimized SQL query, leading to a 26% reduction in query time.

PROJECTS

Uber Data Pipeline

- Built data engineer solution for analytics processes on Uber dataset to process 1 million records with **Airflow**.
- Optimized query time to 7 seconds on **Google BigQuery** by building **Python ETL** pipeline for data ingestion.
- Designed interactive **Looker** dashboard to visualize cloud data to generate revenue-driving insights.

User Retention Analytics

- Undertook product analytics for Yammer dataset, identifying a 15% decline in user retention.
- Analyzed 10 potential causes, revealing mobile users exhibited 20% higher churn rate and observed 25% drop in digest mail click-through rates with **SQL** and **Tableau** producing actionable insights.
- Analyzed **A/B testing** results on search feature revealing it is underutilized and needs enhancements.

Netflix Movie Recommendation system

- Developed Neural collaborative filtering based movie recommendation system and deployed model at scale on **PySpark** and achieved RMSE=0.86 and MAE=0.9569, that outperforms baseline models (SVD, KNN, and SVDpp)
- Worked as data analyst, generating performance insights by analyzing impact of scaling of data on predictions.

EDUCATION

Rutgers University - New Brunswick Sept 2021 - May 2023
Master of Science: Computer Science GPA: 3.96/4.00
Courses: Data structure, Algorithms, Maths for Data Science, Databases, Massive Data mining, Machine Learning.

University of Mumbai Aug 2017 - June 2021
Bachelor's of Engineering**:** Computer Engineering GPA: 8.37/10.00

SKILLS

- **Programming Languages**: Python, R, Java, JavaScript, HTML, CSS | Querying in SQL
- **Python Libraries**: Pandas, NumPy, Django, Flask, scikit-learn, Plotly, Matplotlib, TensorFlow, NLTK, SpaCy.
- **Tools**: Tableau, Looker, ETL pipeline, Postgres, Git, Jira, AWS(EC2, RDS, S3, Lambda).

PUBLICATIONS

- Adaptive Framework for Handwriting Recognition of Numerical Digits, Information, 2023 [PDF](**First Author**).