Mayur Jain

Long Beach, CA (Open to Relocation) | 562-254-7817
mayurjain333@gmail.com | LinkedIn | GitHub

SUMMARY

Results-driven data engineer with 4+ years of experience leading cross-functional teams to build scalable data platforms in healthcare and finance. Known for strong collaboration with product, engineering, and business teams to turn complex data into actionable insights and drive strategic decisions.

SKILLS

Programming: Python, SQL, PySpark, Databricks, Java, REST APIs, Git, Shell
Data Engineering: Dataiku, Airflow, Azure Data Factory, Feature Engineering, Data Modeling
Cloud & DevOps: AWS (S3, EC2), Azure (Monitor, DevOps), GCP, Docker, Jenkins, Kubernetes, CI/CD
Big Data & Storage: Apache Spark, Kafka, Hadoop, Hive, Delta Tables, Azure Data Lake
Machine Learning: Neural Networks, Clustering, Anomaly Detection, NLP, MLFlow, TensorFlow
Data Visualization: Tableau, A/B Testing, KPI Dashboards, Reporting Automation
Management: Agile, JIRA, Confluence, Stakeholder Communication, Team Mentoring, Cross-functional Alignment

PROFESSIONAL EXPERIENCE

Data Engineer | Cigna Group, USA   July 2024 – Present
- Automated data transformation workflows, reducing manual intervention by 40%; improving operational efficiency.
- Optimized data models and schemas to enhance analytics performance, reducing query runtimes by 30%.
- Proactively monitored and resolved ETL pipeline issues with logging and alerting mechanisms, decreasing downtime by 25%.
- Conducted rigorous data quality assessments, maintaining high accuracy and compliance with HIPAA and GDPR regulations.
- Partnered with cross-functional teams to define data governance practices, improving stakeholder satisfaction by 20%.
- Delivered analytical-ready data formats for Medicare analytics, enabling faster patient outcome tracking.

Senior Data Engineer | ZS Associates, USA   September 2022 - July 2023
- Designed ETL pipelines in Dataiku + PySpark to optimize HCP targeting; boosted marketing ROI by 30% across U.S. regions.
- Implemented clustering and neural nets in Zaidyn to predict HCP behavior; improved campaign conversion by 25%.
- Used NLP to guide HCP messaging; integrated outputs into Tableau dashboards for cross-functional visibility.
- Championed data governance initiatives using Collibra to ensure compliance and traceability across 100M+ healthcare records.
- Applied optimization, anomaly detection, and KNN for physician-insurer matching; enhanced targeting accuracy by 20%.

Data Engineer | Tata Consultancy Services, UK   July 2019 - September 2022
- Built fraud detection models using clustering, anomaly detection, and decision forests on Azure Databricks pipelines.
- Processed terabytes of transaction data in real time; flagged fraud with 92% accuracy using tuned ML classifiers.
- Managed end-to-end ML pipeline lifecycle from LLD/HLD, ADF integration to deployment via Azure DevOps + Docker.
- Delivered 10+ ETL pipelines for real-time alerting and classification; reduced false positives by 15%.
- Developed UAT-ready financial models with feature selection and OAT; maintained 99.5% data integrity.

EDUCATION

California State University-Long Beachaster   August 2023 - May 2025
Master of Science(MS), Computer Science   *Long Beach, CA*
- **Achievements:** Member of Career Development Center (CDC), GRE: 329/340
- **Coursework:** Adv Analysis of Algorithm, Adv Software Engineering, Adv Artificial Intelligence, Data Visualization

PROJECTS

Fraud Detection | https://github.com/mayurjainf007/Credit_Card_Fraud_Detection
- Developed event-based alerting system with Spark and Airflow; reduced false positives.
Heart Disease Detection | https://github.com/mayurjainf007/Heart_Disease_Detection
- Streamed vitals using Kafka and scored real-time risk using SparkML.
Sentiment Analysis | https://github.com/mayurjainf007/Apache-Airflow-and-Machine-Learning
- Monitored brand sentiment via Twitter API and NLP pipeline in real time.

CERTIFICATIONS

- **Microsoft Certified:** Azure Data Engineer Associate
- **Microsoft Certified:** Azure DevOps Engineer Expert
- **Published Author on Medium (@mayurjain007):** Topics on Scalable Data Platforms and Real-Time Analytics