MONISHA PATRO

+1-812-345-4652 monishaapatro@gmail.com LinkedIn GitHub Portfolio

EDUCATION

INDIANA UNIVERSITY BLOOMINGTON — **United States**
Master of Science in Data Science — **August 2023 – May 2025**
Relevant Coursework: Data Visualization, Statistics, Applied Machine Learning, Visual Analytics, Database Technologies

VELLORE INSTITUTE OF TECHNOLOGY — **India**
Bachelor of technology in Computer Science — **June 2019 – May 2023**

WORK EXPERIENCE

Candid | ***Nonprofit intelligence via data integration***
Data Science Intern — **May 2024 – December 2024**

- Developed and implemented scalable SQL-based ETL processes within the data warehouse platform to process and standardize non-profit data from govt. publications (~10M+ records), accelerating internal data delivery by 25% for downstream analytics and product teams.
- Partnered with Data Services and API engineering teams to integrate cleaned and mastered datasets into public-facing APIs as a data as service, enhancing data accessibility for 10K+ external users while maintaining backward compatibility using SQL.
- Collaborated with cross-functional product and engineering teams to translate stakeholder requirements into Power BI dashboards using SQL, informing product development and strategy, and contributing to a 15% increase in product adoption while tracking key KPIs.

EProtons | ***Real – Time analytics for EV stations***
Data Science Intern — **October 2022 – February 2023**

- Reconfigured PostgreSQL indexing strategies to improve query performance by 27% validated via logs, on high-volume energy datasets used in forecasting models.
- Engineered distributed data pipelines on AWS EMR and PySpark using Python, orchestrating parallel data processing workflows, yielding a 5x acceleration in large-scale analytics tasks.
- Designed and evaluated an A/B test comparing flat-rate and dynamic pricing models across EV charging stations, informing decisions based on KPIs and utilizing data models, uncovering a 12% lift in session completion using SQL to control for location-based confounders.

Mukham | ***AI – driven attendance and geofencing***
Data Analyst — **June 2022 – November 2022**

- Spearheaded development of CNN-based facial authentication models, cutting spoofing incidents by 50% across high security endpoints.
- Augmented fraud detection performance utilizing geolocation and time-series signals into predictive models, increasing precision by 35%.
- Established image processing pipeline for facial data, improving image quality for 80% of enrolled users and minimizing the number of support tickets related to image failures.

PROJECTS

Real-Time Fraud Detection in Financial Transactions

- Built a fraud detection pipeline using Kafka and Spark Structured Streaming to process 6.3M+ PaySim transactions in real time, triggering anomaly flags within ~1s of ingestion.
- Applied unsupervised models (Isolation Forest and Autoencoders) to detect suspicious TRANSFER and CASH_OUT flows based on features like balance deltas and frequency shifts, achieving a 60% fraud recall on flagged events.
- Deployed real-time dashboards via Streamlit on GCP, allowing teams to explore flagged users, transaction paths, and fraud types across time windows.

Marketing Campaign Uplift Evaluator

- Simulated an A/B experiment across 10 geo-segments to evaluate campaign impact using uplift modeling, logistic regression, and difference-in-differences, isolating a 12.3% increase in weekly revenue post-treatment.
- Controlled for seasonal variation and baseline trends using pre-post analysis, ensuring statistical validity ($p < 0.05$) while identifying the top 3 regions for campaign effectiveness.
- Delivered dashboards comparing treated vs control trends and uplift by segment, helping stakeholders prioritize marketing investments based on measured treatment impact.

TelConnect Customer Churn Prediction

- Developed a churn prediction model using PySpark and SQL pipelines on 1M+ telecom records, identifying high-risk customers based on tenure, complaints, usage gaps, and regional churn trends.
- Achieved 81% recall on churned customers using a tuned Random Forest classifier, helping simulate a retention strategy that reduced predicted churn by 18% in monthly forecasts.
- Generated automated risk flags and summaries for high-risk user segments, supporting monthly churn reviews and business strategy alignment through simple reports and cohort visualizations.

SKILLS

- **Programming & Tools:** Python, SQL, R Spark, PySpark, Kafka, DBT, Power BI, Tabelau.
- **Statistical & Modeling Techniques:** Anomaly detection, Hypothesis testing, Uplift modeling, Logistic regression, Causal inference.
- **Fraud & Risk Analytics:** Behavioral pattern analysis, Fraud detection models, Risk scoring, Treatment effect evaluation, Metric tracking.