

MONISHA PATRO

+1-812-345-4652 ✉ monishaapatro@gmail.com [LinkedIn](#) [GitHub](#) [Portfolio](#)

EDUCATION

INDIANA UNIVERSITY BLOOMINGTON

Master of Science in Data Science

VELLORE INSTITUTE OF TECHNOLOGY

Bachelor of Technology in Computer Science

United States

August 2023 – May 2025

India

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 data delivery by 25% for 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 & engineering teams to translate stakeholder requirements into Power BI dashboards, informing product development and strategy, contributing to a 15% increase in product adoption while tracking key KPIs.

EProtons | *Real – Time analytics for EV stations*

October 2022 – February 2023

Data Science Intern

- 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 utilizing data models, uncovering a 12% lift in session completion using SQL to control for location-based confounders.

Mukham | *AI – driven attendance and geofencing*

June 2022 – November 2022

Data Analyst

- Spearheaded development of CNN-based facial authentication models, cutting spoofing incidents by 50% across security endpoints.
- Augmented fraud detection performance utilizing geolocation & 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

eBay Product Strategy Analytics [🔗](#)

- Engineered a dataset of 15,000+ sold items from eBay using both the Browse and Marketplace Insights APIs, combining real-time product metadata and sales performance to create a rich foundation for marketplace analysis.
- Built a full data pipeline to clean and transform raw NDJSON into structured CSV/Parquet files, enabling seamless analysis of price trends, listing behaviors and category-level performance across product types.
- Analyzed how auction vs. fixed-price formats and single vs. multi-variation listings affect sales outcomes, finding that multi-variation listings led to 12% higher sell-through rates and auction formats underperformed for low-demand items.

TelConnect Customer Churn Prediction [🔗](#)

- Designed an end-to-end churn prediction pipeline on 1M+ customer records using machine learning techniques, selecting AdaBoost for high recall (81%) and reliability in identifying at-risk users.
- Generated automated churn scores and cohort reports segmented by tenure, complaint history, and service usage, supporting real-time risk tracking and strategic retention planning.
- Diagnosed the behavioral journey of churned users, identifying key drop-off patterns and delivering insights through a dashboard tailored for non-technical stakeholders, enabling targeted interventions.

Amazon Product Review Analysis [🔗](#)

- Fine-tuned BERT and LSTM models on 1M+ Amazon product reviews to classify sentiment, achieving 91% accuracy, and enabling insight into how language tone, product category, and phrasing patterns reflect customer experience.
- Refined preprocessing pipeline with advanced tokenization, attention-masking, and review truncation logic, improving the model's ability to handle domain-specific slang, inconsistent punctuation, and variable review lengths across product types.
- Analyzed linguistic misclassification cases to identify patterns in sarcastic, mixed-tone, or repetitive reviews, using these insights to optimize input handling and improve model robustness across nuanced sentiment classes.

SKILLS

- Programming & Databases:** Python, SQL, R, PySpark, ETL.
- Statistics & Experimentation:** A/B Testing, Causal Inference, Hypothesis Testing, Forecasting, Correlation Analysis, Time-series Analytics, Predictive & Descriptive Modeling, Experimentation.
- ML & Visualization:** Power BI, Tableau, Regression, Classification, Gradient Boosting, Bagging, SVM, TensorFlow, PyTorch.