

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 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

Product Funnel Optimization & Behavioral Modeling Simulator

- Simulated the full journey of 1M+ users through a product funnel using Python and SQL, mapping user actions across stages like profile setup, follows, likes, and posts to uncover drop-off points that limited conversion, leading to a 28% gain in simulated retention.
- Built churn and uplift models using XGBoost and causal inference (DoWhy) to identify users most likely to benefit from onboarding interventions, enabling targeted treatment strategies that improved 30-day retention by 10.4%.
- Developed an interactive Streamlit dashboard that consolidated funnel metrics, A/B test insights, and model outputs into a single view, allowing product teams to track engagement trends, interpret results, and debug real-time changes easily.

TelConnect Customer Churn Prediction

- Designed an end-to-end churn prediction pipeline using PySpark and SQL to process over 1M customer records, achieving 81% recall and allowing business teams to forecast risk and allocate retention budgets.
- Generated automated churn scores and cohort reports segmented by tenure, complaint history, and service usage, enabling real-time churn monitoring and pattern discovery for monthly reviews.
- Diagnosed the behavioral journey of churned users, identifying pattern drop-offs before exit and presenting actionable dashboard insights that informed targeted interventions to prevent customer defection and also be user-friendly for non-technical stakeholders.

Amazon Product Review Analysis using BERT & LSTM

- Trained deep learning models using BERT and LSTM on 1M+ product reviews to classify user sentiment with 91% accuracy, surfacing behavioral insights around product preferences and content engagement.
- Enhanced model performance by refining tokenization, masking, and input preprocessing steps, resolving inconsistencies caused by session-level variations and boosting model stability across review types.
- Deployed the models via AWS Lambda and SageMaker with alert-based logging for high-uncertainty cases, supporting real-time scoring, faster debugging, and model monitoring workflows.

SKILLS

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