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, enabling data-driven decisions across departments 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 using Statistical models, uncovering a 12% lift in session completion using SQL and Causal Inference techniques 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 Anomaly Detection in Financial Transactions 🔗
- Architected an end-to-end streaming pipeline simulating 6M+ financial transactions via Kafka and Spark Structured Streaming, lowering fraud detection latency by 80% and enabling real-time model inference.
- Integrated an interactive GCP-hosted Streamlit dashboard to visualize flagged anomalies in real time, empowering product and risk teams to investigate incidents faster and reducing false positives by 20%.
- Deployed a production-ready, low-latency fraud detection system by automating complete data-to-insight pipeline, decreasing false positive rates by 15% and flagging 200+ suspicious transactions daily.

TelConnect Customer Churn Prediction 🔗
- Constructed a predictive churn model (AdaBoost) on customer data, forecasting churn probability across future billing cycles and achieving 96% classification accuracy across key lifecycle attributes.
- Pinpointed primary drivers of customer attrition by analyzing contract types, tenure, and payment behavior, creating customer segmentations to inform retention strategies and mitigating potential revenue losses.
- Mapped customer outreach approaches to specific customer segments identified as high-churn risks, attaining high customer retention rate.

Amazon Product Review Analysis 🔗
- Built a multi-model NLP pipeline leveraging BERT and a custom LSTM to classify sentiment in 100K+ Amazon reviews, supporting both 3-class classification and star-rating regression with 90%+ accuracy.
- Enhanced ad ranking performance by 15% by incorporating sentiment scores into click-through rate prediction models.
- Operationalized real-time inference by deploying ML models on AWS Lambda and SageMaker, aligning scalability with business needs.

SKILLS & CERTIFICATION

- **Programming, Databases & Big Data:** Python, R, SQL, PySpark, SparkSQL, Kafka, Apache Spark, Airflow, AWS EMR, ETL Pipelines.
- **Machine Learning, Deep Learning & NLP:** Supervised (Linear/Logistic Regression, Decision Trees, Random Forest, AdaBoost, Naïve Bayes), Unsupervised (K-Means, DBSCAN, PCA), Ensemble Methods (Gradient Boosting, Bagging), SVM, CNN, RNN/LSTM, Time Series Forecasting, Reinforcement Learning, BERT, HuggingFace, Spacy, NLTK, TensorFlow, Keras, PyTorch, Scikit-learn.
- **Statistical Analysis & Experimentation:** Hypothesis Testing, A/B Testing, ANOVA, Chi-Square, Causal Inference, Correlation Analysis, Time-Series Analytics, Predictive & Descriptive Modeling.
- **Visualization, Deployment & Tools:** [AWS Certified Cloud Practitioner](#), Power BI, Tableau, Streamlit, EDA, GCP, Flask, Git.