

MONISHA PATRO

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

EDUCATION

INDIANA UNIVERSITY BLOOMINGTON

Master's: Data Science | GPA: 3.7/4.0

Relevant Coursework: Data Visualization, Statistics, Applied Machine Learning, Visual Analytics, Database Technologies

United States

August 2023 – May 2025

VELLORE INSTITUTE OF TECHNOLOGY

Bachelor's: Computer Science and Engineering | GPA: 9.2/10

India

June 2019 – May 2023

WORK EXPERIENCE

Candid

New York, US

Data Science Intern

May 2024 – December 2024

- Streamlined data processing workflows for 10M+ records by building ETL pipelines with Python, SQL and Airflow, boosting execution speed by 30%.
- Validated experiment results using A/B testing and hypothesis testing with Statistical Models, raising data validation accuracy by 25%.
- Collaborated with cross-functional product and engineering teams to translate stakeholder requirements into an intuitive Power BI dashboard, enabling data-driven decisions across departments and contributing to a 15% increase in product adoption.

EProtons

Hyderabad, India

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, by orchestrating parallel data processing workflows, yielding a 5x acceleration in large-scale analytics tasks.
- Developed interactive dashboards using Tableau, providing real-time insights into energy consumption patterns, empowering stakeholders to optimize resource allocation and dropping operational costs by 15%.

Mukham

VIT, India

Data Analyst

June 2022 – November 2022

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

January 2025 – February 2025

- 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

November 2024 – December 2024

- Constructed a predictive churn model (AdaBoost) on customer data, achieving 96% classification accuracy across key lifecycle attributes.
- Pinpointed primary drivers of customer attrition by analyzing contract types, tenure, and payment behavior, informing retention strategies, mitigating potential revenue losses.
- Mapped unique customer outreach approaches to specific customer segments identified as high-churn risks, attaining high customer retention rate.

Amazon Product Review Analysis

October 2024 – November 2024

- Developed a multi-model NLP pipeline using 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.