

MONISHA PATRO

United States · monishaapatro@gmail.com · 812-345-4652 · [LinkedIn](#) · [GitHub](#)

EDUCATION

INDIANA UNIVERSITY BLOOMINGTON

Masters: Data Science GPA: 3.7/4.0

Relevant Coursework: Machine Learning, Statistics, Data Analytics, Natural Language Processing, Information Visualization.

United States

August 2023 - May 2025

VELLORE INSTITUTE OF TECHNOLOGY

Bachelor's: Computer Science and Engineering, GPA: 3.8/4.0.

India

June 2019 - May 2023

WORK EXPERIENCE

Candid

Data Science Intern

United States

May 2024 - December 2024

- Developed ETL pipelines with **Python** and **SQL** Server, migrating **10M+** records and shortening data processing time by **30%**.
- Created end-to-end ML solutions on Azure, leveraging **PyTorch** & **TensorFlow** to handle large-scaled structured & unstructured data.
- Conducted A/B Tests & hypothesis validation in Python(SciPy, Statsmodel) improving data validation accuracy by **25%** across ETL.
- Collaborated with cross-functional teams to create Power BI dashboards addressing stakeholder inquiries on non-profit data.

eProtons

Data Science Intern

India

October 2022 - February 2023

- Proposed and implemented indexing in **PostgreSQL**, analyzing vast energy data to increase query performance by **25%** while supporting predictive modelling.
- Engineered distributed data pipelines on **AWS EMR** and **PySpark**, achieving a **5x** acceleration in large-scale analytics tasks.
- Designed and executed SQL queries for large complex datasets, elevating business intelligence operations.

Mukham

Data Analyst Intern

India

June 2022 - November 2022

- Led a team to engineer CNN-based anti-spoofing algorithms, lowering fraudulent access attempts by **50%** in high-security systems.
- Incorporated geolocation and **time-series analytics** into access control, elevating authentication accuracy by **35%**.
- Implemented **fraud detection models**, utilizing **transfer learning** to bolster accuracy across diverse operating conditions.

PROJECTS

Real-Time Fraud Detection pipeline in Transactions:

- Developed an unsupervised ML pipeline on **6M+** transactions enabling real-time detection, cutting fraud response time by **80%**.
- Integrated **Kafka** & **Spark** streaming to process live data, minimizing detection latency from hours to near real-time.
- Deployed a dashboard for immediate visualization of suspicious events, refining risk assessment and decision-making.

Amazon Product Review Analysis:

- Built a sentiment analysis pipeline with **BERT** & **LSTMs** classifying **100K+** of amazon reviews to gauge customer satisfaction.
- Formulated an ML-driven ad ranking module, leveraging sentiment scores to improve product recommendation CTR by **15%**.
- Deployed solution on **AWS Lambda** & **Sagemaker**, ensuring scalable real-time inference and seamless e-commerce integration.

TelConnect Customer Churn Prediction:

- Orchestrated an AdaBoost-based **churn prediction** model boosting accuracy and model stability in forecasting at-risk subscribers.
- Uncovered pivotal user behavior patterns driving churn, enabling data-driven **retention** strategies that significantly decreased churn.

Google Ads Search Optimization:

- Evaluated **5M+** search queries and ad performance data via **PySpark** and **SQL**, pinpointing user engagement trends.
- Constructed a budget optimization module using time-series forecasting, reducing average CPC by **20%** & boosting conversion rates.
- Optimized bidding strategies by modelling auction dynamics and user behavior enhancing ad relevance and **ROI**.

SKILLS

- Languages & Databases:** Python, R, SQL, PySpark.
- Libraries, Frameworks & NLP:** TensorFlow, Pandas, NumPy, Scikit-learn, SciPy, Keras, NLTK, Spacy, HuggingFace, BERT.
- Visualization and Analysis:** Power BI, Tableau, Matplotlib, Seaborn, EDA, Statistical Modeling, Hypothesis Testing, ANOVA, Causal Inference, Chi-Square, Predictive/Descriptive Analytics.
- ML:** DL techniques (CNN, RNN/LSTM), Supervised Algos(Linear & Logistic regression, Decision Trees, Random forest, Boosting, SVM&Naïve Bayes), Clustering Techniques, time-series forecasting, A/B testing, Predictive modeling, RL, ETL pipelines.