

MONISHA PATRO

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

EDUCATION

INDIANA UNIVERSITY BLOOMINGTON

Masters: Data Science GPA: 3.7/4.0

Relevant Coursework: Machine Learning, Statistics, Data Analytics, Natural Language Processing, Data 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 by handling ingestion and transformation on 10M+ records with Airflow, accomplishing a 30% reduction in overall execution time.
- Conducted A/B tests and hypothesis validation using Scipy and Statsmodel, improving data validation accuracy by 25% across ETL.
- Collaborated with cross-functional teams to create Power BI dashboards and reports by capturing stakeholder requirements and designing interactive visualizations, enhancing insights into advanced analytical inquiries.

eProtons

Data Science Intern

India

October 2022 - February 2023

- Proposed and implemented indexing in PostgreSQL by optimizing database schema and analyzing vast energy data, increasing query performance by 27% while supporting predictive modelling.
- Engineered distributed data pipelines on AWS EMR and PySpark, by orchestrating parallel data processing workflows, achieving a 5x acceleration in large-scale analytics tasks.
- Executed SQL queries for large, complex datasets utilizing Hadoop for distributed storage and processing, elevating BI operations.

Mukham

Data Analyst Intern

India

June 2022 - November 2022

- Led a team to engineer CNN-based anti-spoofing algorithms, by applying deep learning frameworks, 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 by leveraging anomaly detection algorithms with real-time data ingestion, resulting in an 80% reduction in fraud response time.
- Integrated Kafka & Spark streaming by building a real-time pipeline on Google Cloud, cutting detection latency from hours to secs.
- Deployed a dashboard for suspicious event visualization leveraging GCP for immediate data retrieval and a dynamic front end, enhancing decision-making with instant insights.

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 by integrating these outputs into product recommendation algorithms, improving product recommendation CTR by 15% and deployed solution on AWS Lambda & Sagemaker.

TelConnect Customer Churn Prediction: [🔗](#)

- Orchestrated an AdaBoost-based churn prediction model by tuning ensemble algorithms and hyperparameters, securing 96% accuracy and boosting model stability in forecasting at-risk subscribers.
- Uncovered churn drivers by conducting EDA and correlation analyses, enabling targeted retention strategies reducing churn.

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.