MONISHA PATRO

United States · monishaapatro@gmail.com · 812-345-4652 · LinkedIn . GitHub

EDUCATION

INDIANA UNIVERSITY BLOOMINGTON — United States
Masters: Data Science GPA: 3.7/4.0 — **August 2023 - May 2025**
Relevant Coursework: Machine Learning, Deep Learning, Natural Language Processing, Distributed Computing, Business Intelligence

VELLORE INSTITUTE OF TECHNOLOGY — India
Bachelor's: Computer Science and Engineering, GPA: 3.8/4.0. — **June 2019 - May 2023**

WORK EXPERIENCE

Candid — United States
Data Science Intern — **May 2024 - December 2024**

- Developed scalable ETL pipelines using Python and SQL Server, migrating over 10M records and reducing processing time by 30%.
- Built ML pipelines on AWS using PyTorch and TensorFlow, analyzing large-scale structured and unstructured data.
- Conducted A/B testing and statistical hypothesis testing to validate data integrity, ensuring high accuracy across data pipelines.
- Utilized PySpark and Power BI to perform in-depth analysis and create interactive dashboards uncovering key insights that informed strategic decision-making and improved data-driven initiatives across cross-functional teams.

eProtons — India
Data Science Intern — **October 2022 - February 2023**

- Optimized PostgreSQL databases and SQL queries to efficiently manage and retrieve large-scale energy consumption data, enhancing data processing speed by 25% and supporting advanced deep learning and predictive modelling initiatives.
- Built distributed data processing systems using AWS EMR and PySpark, achieving 5x acceleration in large-scale analytics.
- Designed and executed SQL queries for structured and semi-structured datasets, enhancing business intelligence operations.

Mukham Pvt Ltd — India
Data Analyst Intern — **June 2022 - November 2022**

- Developed anti-spoofing algorithms using convolutional neural networks (CNNs), reducing fraudulent access attempts by 50%.
- Integrated geolocation analytics and time-series analysis into access control systems, improving authentication accuracy by 35%.
- Implemented real-time fraud detection models, leveraging transfer learning to increase accuracy by 25% under diverse conditions.

PROJECTS

Google Ads Search Optimization:

- Analyzed 5M+ search queries and ad performance data using PySpark and SQL to identify trends in user engagement.
- Built an ML-based ad ranking system using NLP and statistical modelling to optimize keyword targeting.
- Assessed bidding strategy optimizations by modeling auction dynamics and user behavior improving ad relevance.

Real-Time Anomaly Detection in Fraud Transactions:

- Designed fraud detection algorithms using unsupervised ML techniques on large scale financial transaction dataset.
- Developed real time processing pipeline with Kafka and Spark Streaming demonstrating its feasibility for rapid fraud detection.
- Implemented a monitoring dashboard to visualize anomaly patterns and support decision making for risk assessment.

Amazon Product Review Analysis:

- Created sentiment analysis pipeline using BERT & LSTMs, classifying thousands of amazon reviews to access customer satisfaction.
- Engineered ML-driven Ad ranking model based on scores to recommend products and deployed on AWS Lambda & Sagemaker.

TelConnect Customer Churn Prediction:

- Developed a churn prediction model using AdaBoost, and SQL improving accuracy in forecasting customer retention.
- Identified key customer behavior trends influencing churn rates, supporting data driven retention strategies.

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.