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%.
- Designed data warehousing solutions to improve data accessibility, supporting analytics for financial insights.
- 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 — Remote
Data Analyst 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.
- Conducted statistical modeling and regression analysis to identify business patterns and inform strategic decisions.

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

- Developed advanced anti-spoofing algorithms using convolutional neural networks**,** reducing fraudulent access attempts by 50% and enhancing security for client organizations**.**
- Integrated geolocation analytics and time-series analysis into access control systems, improving location-based authentication accuracy by 35% and reducing unauthorized access incidents by 20%.
- Implemented facial recognition algorithms using transfer learning techniques, increasing recognition accuracy by 25% under diverse conditions and reducing false positives.

PROJECTS

- **Amazon Product review Analysis:** Develpoed an LLM-powered sentiment analysis model using BERT and LSTMs to analyze Amazon product reviews. Designed an ad ranking system leveraging sentiment scores, improving click-through rates by 15%. Built cloud-deployed ML pipelines using AWS Lambda and SageMaker to optimize inference speed.
- **TelConnect Customer Churn Prediction:** Developed a machine learning-based churn prediction model to identify high-risk customers, optimizing retention strategies. Used PySpark, SQL and AdaBoost to process records, improving prediction accuracy and reducing retention costs through data-driven customer targeting.
- **Google Ads Search Optimization Project:** Focused on improving Google Ads' search algorithms to enhance targeting accuracy and ad performance that involved analyzing over 5 million search queries and ad data records using Spark and SQL to refine ad targeting and bidding strategies, thereby optimizing ad relevancy.
- **Real-Time Anomaly detection in financial transactions:** Designed fraud detection algorithms leveraging unsupervised learning techniques deployed real-time detection models using Kafka and Spark streaming with dashboards for transaction monitoring.

SKILLS

- **Languages & Databases**: Python, R, SQL, PySpark
- **Libraries and Frameworks:** TensorFlow, Pandas, NumPy, Matplotlib, Scikit-learn, SciPy, Keras
- **Visualization and Analysis:** Power BI, Tableau, Gephi, Matplotlib, Seaborn, EDA, Statistical Modeling
- **ML:** Deep Learning Techniques (CNN, RNN/LSTM), Supervised Learning Algorithms (Linear Regression, Logistic Regression, Decision Trees, Random Forest, Boosting, SVM, and Naive Bayes), Clustering Techniques, and Time-series Forecasting Techniques, A/B Testing, Hypothesis Testing. Predictive Modelling, Reinforcement Learning, ETL Pipelines, Anomaly Detection
- **NLP:** NLTK, Spacy, HuggingFace, BERT
- **Statistical Analysis:** Hypothesis Testing, ANOVA, Chi-Square, Predictive/Descriptive Analytics