NIRAJ MOHABEY

+1-774-502-4709 | Worcester, MA | nmohabey31@gmail.com | LinkedIn | Portfolio | GitHub

EDUCATION

Master of Science in Data Science, Worcester Polytechnic Institute **12/2025 (Expected)**
GPA: 3.88/4.0
Relevant Coursework: Machine Learning, Natural Language Processing, Big Data Analysis, Data Visualization

SKILLS

Languages & Libraries: Python, SQL, R, scikit-learn, TensorFlow, PyTorch, Hugging Face Transformers
Big Data & Engineering: Apache Spark, Apache Kafka, Apache Airflow, Docker, Git
Cloud & Databases: AWS (EC2, S3, Lambda), PostgreSQL, Snowflake
Visualization & BI: PowerBI, Streamlit, Plotly, Grafana
Core Expertise: Machine Learning, NLP, Predictive Modeling, Real-Time Data Analytics, Model Deployment, MLOps

EXPERIENCE

MITSUBISHI UFJ FINANCIAL GROUP | *Data Science Intern* **01/2025 – 05/2025**
- Engineered a fixed-income ETF pricing engine using Python, SQL, ETL pipelines and Streamlit, modeling 5,000+ bonds with sub-0.5% deviation from Bloomberg YAS, enhancing trade execution confidence for investors.
- Deployed a Streamlit dashboard with 1,000+ ISIN-level yield shock curves, enabling interactive scenario modeling for interest rate risk analysis across diverse portfolios for strategic investment planning.
- Collaborated with quant analysts to deploy production-ready dashboards with real-time pricing logic, streamlining portfolio risk decisions and boosting trader efficiency in high-frequency trading environments.

SUNRISE GROUP USA LLC | *Data Science Intern* **06/2022 – 11/2023**
- Modeled a loan status prediction model using SVM in Python, achieving 83% test accuracy on a preprocessed dataset, improving lending decisions for financial risk mitigation in banking operations.
- Advanced a Linear Regression model for medical insurance cost prediction, attaining 0.74 R^2 score, with a Streamlit interface for user-friendly predictions and stakeholder engagement across cross-functional teams.
- Launched PowerBI dashboards to visualize financial data patterns, supporting cross-functional teams in anomaly detection and data-driven decision-making processes for enhanced financial strategy formulation.

INNODATATICS.INC | *Data Science Intern* **08/2020 – 01/2021**
- Built a sentiment analysis pipeline for TripAdvisor reviews using NLP and Python, achieving 98% accuracy in classifying customer feedback for actionable business insights to drive service improvements.
- Crafted an intrusion prediction model for network traffic data, leveraging feature engineering and scikit-learn, enhancing cybersecurity measures for enterprise networks in real-time threat detection.

PROJECTS

BigDocBot – AI-Powered Code Documentation Tool **01/2025 – 05/2025**
- Developed a Streamlit tool to auto-generate summaries, docstrings, and code quality reports for Python and JavaScript using LLMs (DeepSeek-Coder, CodeT5+), optimized with AMP, batching, and 8-bit quantization.
- Integrated an end-to-end data engineering pipeline to parse GitHub repositories, run LLM inference, and export structured PDF reports, rigorously validated output on curated code corpora for quality assurance and reliability.

AI-Powered Customer Churn Prevention System **01/2025 – 05/2025**
- Designed an XGBoost model to predict churn with 90% precision on 10,000-customer Telco data, enhancing retention strategies for telecom business operations and maximizing customer lifetime value.
- Created a PostgreSQL analytics layer and Streamlit dashboard to visualize churn trends, reducing detection latency by 40% using Kafka pipelines for real-time business decision-making.

Conversational AI Product Discovery Platform **08/2024 – 01/2025**
- Innovated a semantic search platform using BERT and Sentence Transformers, improving query accuracy by 30% for e-commerce product discovery and user engagement across diverse retail platforms.
- Implemented a PostgreSQL catalog with SQL filtering and React.js frontend, deployed on AWS for sub-second query response across 1,000+ daily users, enhancing online shopping experiences.

Real-Time Financial Fraud Detection with Graph Intelligence **08/2024 – 12/2024**
- Constructed a GNN-based fraud detection system using PyTorch Geometric, identifying 95% of fraud rings in PaySim data for enhanced financial security and regulatory compliance assurance.
- Optimized a Kafka-Flink streaming pipeline and Grafana dashboard, achieving 100ms latency for real-time fraud insights and investigation efficiency in high-stakes financial environments.