RISHABH INDORIA

(732) 486-6359 | indoria.r@northeastern.edu | LinkedIn | GitHub | Portfolio

WORK EXPERIENCE

AI Developer Intern, Next Quarter | Boston, MA — February 2024 – April 2024

- Engineered a **Retrieval-augmented Generation Generative AI** tool in **Salesforce** to devise new business growth strategies and perform **trend analysis** using **DBSCAN clustering** algorithms, boosting earnings potential by 24%
- Modeled a **customer churn** tool using a **random forest classifier** with a 77% retention rate in consulting business
- Innovated a **conversational AI** by ingesting real-time financial transcripts from **Kafka**, integrating **LangChain**, RAG, and **Pinecone** vector similarity, deploying **Airflow** and **Spark** pipelines on **Docker**, reducing hallucination by 30%

Senior Machine Learning Engineer, Urja.io | New Delhi, India — April 2020 - August 2022

- Resolved machinery failure costs by leading a cross-functional ML team to **forecast time-series energy** data using an experiment with **Prophet** and **SARIMAX** models (accuracy: 94%), increasing Customer Satisfaction Score by 23%
- Streamlined decision-making by communicating **Agile A/B testing** progress in **Tableau**, saving stakeholders 10 hours weekly, and recovered $100,000 in potential losses through **data-driven** analysis of product development metrics
- Collaborated to visualize **IoT** time-series data using **Grafana dashboards**, which increased user engagement by 16%

Data Scientist, Société Générale Global Solution Centre | Bengaluru, India — July 2017 - March 2020

- Improved data **Extract – Transform – Load (ETL)** pipeline efficiency by developing **REST APIs, Infrastructure as Code** with **Terraform**, and **CI/CD** pipelines using **GitHub**, reducing time to production on **AWS** by 30%
- Deployed **risk management** pipeline by implementing a **XGBoost** classification model (AUC: 97%) to identify outlier trades through statistical methods, **pattern analysis** and feature selection, enhancing **fraud detection** rates by 17%
- Achieved a 3.5% error rate in redacting financial transcripts using **NER** and **POS** tagging with **spaCy**

PROJECTS

Movie Recommendation System | GitHub — May 2024

- Developed a hybrid **SVD** and **Neural Collaborative Filtering** model integrated with an **XGBoost** meta-learner, achieving 88% of the SOTA benchmark **RMSE** on the MovieLens dataset through **ensemble learning**

Kidney Disease Detection and Classification from CT scans | GitHub — April 2024

- Implemented a pre-trained **VGG16** model to **classify** kidney diseases from CT scans using data sources identified via PubMed literature, integrating **MLflow** and **DagsHub** for monitoring, achieving 93% diagnostic precision in healthcare

Imbalanced Meteorological Data Analysis for Rainfall Prediction | GitHub — December 2023

- Enhanced Class-1 accuracy prediction of multimodal time-series and satellite images dataset to 77% by using **SMOTE** for oversampling, feature engineering, and fine-tuning a **PyTorch** Convolutional - **LSTM** model with **CUDA** on GPUs

Speech-To-Text with LLM | GitHub — August 2023

- Integrated **sentiment analysis**, a scalable **Snowflake** data lake and **Automatic Speech Recognition** with **OpenAI whisper**, and **BERT summarization** within **Docker** containers, boosting information retrieval time by 20%

Economics of Happiness | GitHub — November 2022

- Boosted predictive accuracy by 9% by analyzing statistics of econometric features across 150 countries, leveraging **SHAP** for **hyperparameter tuning** and **causal inference**, employing **logistic regression** and **decision trees** after **EDA**

EDUCATION

Northeastern University | GPA: 3.72 — Boston, MA

Master of Science in Information Systems | Teaching Assistant - Advanced Data Science and Architecture — April 2024

Relevant subjects: Neural Modelling, High-Performance Parallel ML and Artificial Intelligence, Big Data Systems

CERTIFICATION

- **Microsoft** Certified: Azure Data Scientist Associate

SKILLS

- **Programming Language/Shell Scripting:** Java, Python, R, Spark, SQL, Kafka, MATLAB, Linux, SAS
- **Tools/Frameworks:** Sklearn, NLTK, HuggingFace, TensorFlow, Keras, Azure, GCP, Kafka, Power BI, Snowflake, Redis, InfluxDB, Hadoop, Databricks, SageMaker, BigQuery
- **Competencies:** Linear Regression, Support Vector Machines, Gradient Boosted Machines, Partial Dependence Plots, Principal Component Analysis, RNN, CNN, GAN, Transformers