# RISHABH INDORIA

732.486.6359 | indoria.r@northeastern.edu | LinkedIn | GitHub | Portfolio

## EDUCATION

**Northeastern University**                                                                                         Boston, MA
*Master of Science in Information Systems (Teaching Assistant – Advances in Data science and Architecture)*     Expected April 2024
**Manipal Institute of Technology (MIT)**                                              Manipal, Karnataka, India
*Bachelor of Technology in Information Technology*                                                                  2013 - 2017

## SKILLS

- **Software Programming and Shell Scripting:** Python, Spark, SQL, Kafka, Algorithms, Linux, Data structures, Algorithms, APIs
- **Machine Learning:** Supervised learning, Scikit-Learn, NLTK, HuggingFace, Statistical modeling, OpenAI, PyCaret, GenAI
- **Deep Learning:** TensorFlow, Keras, PyTorch, Computer vision, Unsupervised modeling, RNN, CNN, GAN, GNN, Transformers
- **Cloud computing:** AWS (Ec2, IoT, S3, Lambda, Redshift), GCP, Git, Airflow, Terraform, SageMaker, Docker, Azure, Dataiku
- **Data Warehouse:** Snowflake, Redis, Pinecone, InfluxDB, Telegraf, Hadoop, Databricks, Postgres, MongoDB
- **Data Visualization and Business Intelligence:** PowerBI, Tableau, Grafana, Excel, SAP Analytics cloud

## WORK EXPERIENCE

**NEXT QUARTER | *AI Developer Intern***                                                            February 2024 - Present
- Delivered an Artificial Intelligent solution for account-planning and **price predictive model**, increasing client's earnings potential by 23% and enhancing **sales forecast** accuracy by 20% through predictive growth strategy forecasting and effective **trend analysis**
- Proposed an **automated risk management** LLM framework POC, achieving a 37% increase in clients' customer engagement

**URJA.IO | *Lead Data Scientist***                                                                 April 2020 - August 2022
- Spearheaded **ML Ops** team in creating an end-to-end analytical IoT dashboard tool, resulting in a 25% increase in user experience
- Reduced costs by 23% with a 94% accurate machinery-failure **prediction system** using **Prophet** for **IoT time-series forecasting**
- Streamlined stakeholder decision making by 10 hours biweekly through efficient **Tableau reporting** of **A/B test** progress
- Contributed to **SAFe Agile** roadmap planning by collaborating with **cross-functional** teams to implement **product strategies**
- Saved $100,000 by leading **data-driven business solutions** and **monitoring product metrics** to steer **product innovation**

**SOCIETE GENERALE GLOBAL SOLUTION CENTRE | *Data Scientist***                                       July 2017 - March 2020
- Enhanced **AWS IaC ETL** pipeline deployment speed by 30% through **REST APIs**, **Terraform** and **GitHub CI/CD** optimization
- Increased **fraud detection** in investment banking by 17% with **XGBoost** models for identifying outlier transaction volumes
- Boosted **risk management** efficiency by 21% using **Isolation Forest** to spot high-value transactions exceeding forecasted pricing

## PROJECTS

**IMBALANCED METEOROLOGICAL DATA ANALYSIS FOR RAINFALL PREDICTION | GITHUB**                          December 2023
- Enhanced prediction accuracy in an **imbalanced dataset** using **SMOTE, feature engineering, and optimizer tuning**
- Trained multimodal image and time-series **ConvLSTM** model for Rainfall Forecasting Achieving **77% Class-1 Accuracy**

**BIOMEDICAL TEXT TAGGING USING BIOBERT | GITHUB**                                                    December 2023
- Automated **NER** and **POS** tagging of genes and proteins in biomedical data using **transformer based BioBert** model using **spaCy**
- Integrated pipeline with **Dataiku** enabled streamlined model training and deployment, achieving a **validation accuracy** of 95%

**SPEECH-TO-TEXT WITH LLM | GITHUB**                                                                  August 2023
- Integrated emotion detection, scalable **Snowflake** data storage, **GenAI prompt engineering** for contextual feedback, **speech-to-text** logging, and transformer-based BERT summarization, deployed within Docker containers, increasing user retention by 20%

**FINANCE SEARCH OPTIMIZATION | GITHUB**                                                              May 2023
- Engineered an NLP Q&A text summarization platform, ingesting streaming transcripts through **Kafka** pub-sub, storing in Redis, and integrating **LangChain**, **RAG**, and **Pinecone vector similarity** for 50% improved information quality, alongside **Apache Airflow** and **Apache Spark** data processing pipelines deployed in **Docker** for efficient processing

**SENTIMENT ANALYSIS: DECIPHERING RESTAURANT REVIEWS | GITHUB | YOUTUBE**                             April 2023
- Curated dataset with **GPT-4 Large Language Model** (LLM) using **Reinforcement Learning from Human Feedback (RLHF)**
- Implemented **tokenization, TF-IDF Vectorization**, and **Naive Bayes** for real-time sentiment analysis, achieving 91% accuracy

**PRODUCT RECOMMENDATION SYSTEM | GITHUB**                                                            April 2023
- Enhanced recommendations by incorporating pre-trained **GloVe** embeddings with clustering into customer segmentation strategies

**ECONOMICS OF HAPPINESS | GITHUB**                                                                   November 2022
- Leveraged SHAP for hyperparameter tuning and **causal inference, logistic regression, and decision trees** in **exploratory data analysis (EDA)**, boosting predictive accuracy by 9% in linking happiness index and economic indicators across 150 countries

## CERTIFICATIONS

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