

# RISHABH INDORIA

732.486.6359 | [indoria.r@northeastern.edu](mailto: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

Relevant subjects: Data Science Engineering Methods, Neural Modelling, Parallel ML and AI, Big Data Intelligent Systems

### 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, MATLAB, Linux, APIs
- **Machine Learning:** Scikit-Learn, NLTK, HuggingFace, OpenAI, PyCaret
- **Deep Learning:** TensorFlow, Keras, PyTorch, Computer vision, RNN, CNN, GAN, GNN, Transformers
- **Cloud computing:** AWS (Ec2, IoT, S3, Lambda, Redshift), Git, Airflow, Terraform, SageMaker, Docker, Azure
- **Data Warehouse:** Snowflake, Redis, Pinecone, InfluxDB, Hadoop, Databricks, PostgreSQL
- **Data Visualization and Business Intelligence:** PowerBI, Tableau, Grafana, Excel, SAP Analytics cloud

## WORK EXPERIENCE

### NEXT QUARTER | AI Developer Intern

February 2024 - Present

- Extracted sales data using **RAG** for an AI account-planning tool in **Salesforce**, achieving a 23% increase in the earnings potential
- Performed **statistical trend analysis** on historical growth strategies for **clustering**, enhancing sales forecast accuracy by 20%
- Proposed an automated risk management **NLP LLM** framework, achieving a 37% increase in clients' customer engagement

### URJA.IO | Lead Data Scientist

April 2020 - August 2022

- Led the ML Ops team in creating an end-to-end analytical **IoT dashboard** tool, resulting in a 25% increase in user experience
- Implemented a machinery-failure prediction system using Prophet for real-time IoT **time-series** forecasting, reducing costs by 23%
- Streamlined stakeholder decision-making through efficient **Tableau** reporting of **A/B** test progress, saving 10 hours biweekly
- Led **data-driven** business solutions and **monitored product metrics**, saving \$100,000 by steering product innovation

### SOCIETE GENERALE GLOBAL SOLUTION CENTRE | Data Scientist

July 2017 - March 2020

- Enhanced **AWS IaC ETL** pipeline infrastructure deployment speed by 30% through REST APIs, Terraform, and GitHub CI/CD
- Increased **fraud detection** in investment banking by 17% with **XGBoost** model for identifying outlier trading transaction volumes
- Boosted **risk management** efficiency by 21% using **Isolation Forest** to spot high-value transactions exceeding forecasted pricing
- Employed **NER** and **POS** tagging techniques to **redact** financial transcripts with **spaCy**, achieving a low error rate of 3.5%

## PROJECTS

### IMBALANCED METEOROLOGICAL DATA ANALYSIS FOR RAINFALL PREDICTION | [GITHUB](#)

December 2023

- Enhanced prediction accuracy in an imbalanced (**biased**) dataset by utilizing **SMOTE** for **oversampling**, applying **feature engineering** techniques, and **tuning optimizers**, resulting in 10% improved model performance for rainfall forecasting
- Trained a **multimodal ConvLSTM** model, integrating image and time-series weather data, achieving a Class-1 accuracy of 77%

### IMAGE SEGMENTATION WITH CONTEXTUAL CAPTIONING | [GITHUB](#)

October 2023

- Implemented CNN with **PyTorch** and **OpenCV** for efficient image segmentation, achieving an F1-Score of 0.92
- Developed RNN in **TensorFlow** and **NLTK** for generating context-aware captions, with a 70 Perplexity score from segmented data

### 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** within **Docker** containers, increasing user retention by 20%

### FINTech CONTEXTUAL Q&A SEARCH OPTIMIZATION | [GITHUB](#)

May 2023

- Engineered a Q&A text summarization platform, ingesting transcripts and storing in **Redis**, refining information quality by 50%
- Used **LangChain**, **Retrieval-augmented generation (RAG)**, and **Pinecone vector similarity**, increasing query handling by 40%
- Deployed **Airflow** and **Spark data processing** pipelines on **Docker**, resulting in a 50% reduction in information retrieval time

### SENTIMENT ANALYSIS: DECIPHERING RESTAURANT REVIEWS | [GITHUB](#) | [YOUTUBE](#)

April 2023

- Curated a dataset with **GPT-3.5 turbo LLM** using **RLHF**, implemented **tokenization**, **TF-IDF Vectorization**, and **Naive Bayes** for real-time **sentiment analysis**, achieving 91% accuracy

### ECONOMICS OF HAPPINESS | [GITHUB](#)

November 2022

- Boosted predictive accuracy by 9% in analyzing the happiness index and economic indicators across 150 countries, by leveraging **SHAP** for **hyperparameter tuning** and **causal inference**, and employing **logistic regression** and **decision trees** in **EDA**

## CERTIFICATIONS

- **Microsoft Certified:** Azure Data Scientist Associate: [Certificate](#)