

# 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

*Relevant courses: Advanced Data Science (Teaching Assistant), Neural Networks, Big Data Intelligent Analytics, Cloud Computing*

### Manipal Institute of Technology (MIT)

Manipal, Karnataka, India

Bachelor of Technology in Information Technology

*Relevant courses: Relational Database, Data Structures, Algorithms, Operating System*

## SKILLS

- **Software Programming and Shell Scripting:** Python, Spark, SQL, Kafka, Algorithms, Linux, Javascript, Unix
- **Machine Learning:** Supervised learning, Scikit-Learn, NLTK, HuggingFace, Statistical modeling, OpenAI, PyCaret
- **Deep Learning:** TensorFlow, Keras, PyTorch, NLP, Unsupervised modeling, FFN, RNN, CNN, GNN, Transformers
- **Cloud computing:** AWS (Ec2, IoT, S3, Lambda, Redshift), GCP, Git, Airflow, Terraform, SageMaker, Docker
- **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

### URJA.IO | Lead Data Scientist

April 2020 - August 2022

- Spearheaded development team in creating an end-to-end analytical dashboard tool, resulting in a 25% increase in user retention
- Reduced costs by 23% with a 94% accurate machinery-failure **prediction system** using **Facebook Prophet** for **forecasting**
- Streamlined stakeholder communication by 10 hours weekly through efficient **Tableau reporting** of production A/B test progress

### SOCIETE GENERALE GLOBAL SOLUTION CENTRE | Data Scientist

July 2017 - March 2020

- Enhanced deployment speed by 30% with Infrastructure as Code (IaC) through Terraform and GitHub CI/CD optimizations
- Increased **anomaly detection** in investment banking by 17% using **XGBoost** models for identifying irregular transaction volumes
- Boosted **risk management** efficiency by 21% using **Isolation Forest** to spot **outlier** transactions exceeding forecasted pricing
- Employed **Named Entity Recognition** and POS tagging techniques to redact financial transcripts, achieving an error rate of 3.5%

## PROJECTS

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

December 2023

- Enhanced prediction accuracy in an **imbalanced dataset** using **SMOTE**, **class weight adjustments**, and **optimizer tuning**
- Developed and trained multi-model **ConvLSTM** Model for Rainfall Forecasting Achieving **77% Class-1 Accuracy**
- Trained integrated **RNN-TimeDistributed CNN** model for dual time-series and image data analysis, achieving **93% F-1 score**

### BIOMEDICAL TEXT TAGGING USING BIOBERT | [GITHUB](#)

December 2023

- Automated NER for gene tagging in biomedical text using the **BioBert transformer** model, achieving validation accuracy of 95%

### ENHANCED AUDIO JOURNALING WITH LLM | [GITHUB](#)

August 2023

- Implemented containerized **audio processing** algorithms, focusing on emotion detection for accurate **sentiment analysis**
- Leveraged **Snowflake** for efficient and scalable data storage, facilitating data integrity and robust data management
- Utilized GPT-3.5 **prompt engineering** for generating personalized feedback, achieving **20% improvement** in user insight

### FINANCE SEARCH OPTIMIZATION | [GITHUB](#)

May 2023

- Engineered a Q&A and **text summarization** platform, ingesting streaming transcripts from **Kafka pub-sub** and storing it in **Redis**
- Used **Langchain** traditional filters, **RAG** and **Pinecone** for vector similarity to improve information quality by 30%
- **Dockerized Airflow** data pipelines and **Spark** processing strategies, resulting in a **50% reduction in information retrieval time**

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

April 2023

- Developed a dataset of restaurant reviews through **prompt engineering** using **OpenAI's Davinci** model
- Applied text preprocessing techniques (**tokenization**, **lemmatization**) followed by **TF-IDF Vectorization** and **Naive Bayes** algorithm for sentiment analysis on real-time data, achieving a classification accuracy of 91%

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

November 2022

- Utilized **logistic regression** and **decision trees** for exploratory analysis of the relationship between economic indicators and happiness index across 150 countries, enhancing model insights by 15% through **hyperparameter tuning** and **SHAP analysis**

## ACHIEVEMENTS

- Presented at AI Skunkworks' "Generative AI Deep Dive" technical workshop with the Institute for EAI ([Watch Video](#))
- Microsoft Certified: Azure Data Scientist Associate: [Certificate](#)