**RISHABH INDORIA**

Boston, MA (Willing to relocate) • (732) 486-6359 • [indoria.r@northeastern.edu](mailto:indoria.r@northeastern.edu) • [LinkedIn](https://www.linkedin.com/in/rishabhindoria/) • [GitHub](https://github.com/rishabhindoria25)

**education**

**Northeastern University • Master of Science in Information Systems • GPA: 3.72** (April 2024) Boston, MA

Relevant subjects: Neural Networks, High-Performance Parallel ML and Artificial Intelligence, Big Data Technologies

**RELEVANT skills**

* **Programming Language/Shell Scripting:** JavaScript, Java, R, Python, PySpark, Pandas, SQL, NumPy, Linux
* **Tools/Frameworks:** Scikit-Learn, NLTK, HuggingFace, PyTorch, Keras, Google BigQuery, Snowflake, Excel, AWS SageMaker, Hadoop, Databricks, PowerBI, PostgreSQL, Kubernetes, Vector DB, Llama, GPT, Flask, Map/Reduce
* **Competencies:** Data Structures, Algorithms, Statistical Modelling, Mathematics, Regression, LSTM, CNN, RNN, PCA, Transformers, Predictive Models, Data Analytics, Compliance, Transfer Learning, Text Mining

**experience**

**Next Quarter** Boston, MA

*Data Scientist Intern* January 2024 - April 2024

* Fine-tuned a pre-trained **Large Language Model (LLM) Generative AI** tool, boosting earnings potential by 24% through go-to-market strategy development and healthcare claims trend analysis of 1B Salesforce records
* Created a **conversational AI**, enhancing response accuracy by 45% by integrating **LangChain** Retrieval augmented Generation (RAG) and **Pinecone** with Apache **Kafka** and Apache Spark pipelines within Docker containerization

**Northeastern University** Boston, MA

*Teaching Assistant - Advanced Data Science and Architecture* January 2024 - April 2024

* Implemented RAG (Retrieval-Augmented Generation) architecture to fine-tune GPT, enabling automatic generation of course materials and assessment questions, resulting in a 30% reduction in curriculum development time

**Urja.io** New Delhi, India

*Senior Machine Learning Engineer* April 2020 - August 2022

* Engineered a **recommendation** system, achieved 96% precision in product features pitch and increased sales lead conversion by 22% through Single Value Decomposition, Neural Collaborative Filtering, and **XGBoost** meta-learner
* Led a cross-functional backend **MLOps** team, achieved 94% accuracy, and increased the Customer Satisfaction Score by 23% through the implementation of a **real-time** IoT time series forecasting platform with **MLFlow** monitoring
* Streamlined business stakeholders' data-informed decision-making, saving 10 hours weekly and recovering $100,000 in potential losses by implementing Agile **A/B testing** and data visualization Tableau dashboards for data-driven KPIs
* Developed a CRM tool, improved customer retention by 79% and reduced churn by 15% for 50,000+ customers using role-based dashboards, **decision trees classification**, and **K-means clustering** data models with **Airflow** orchestration

**Société Générale Global Solution Centre** Bengaluru, India

*Data Scientist* July 2017 - March 2020

* Optimized end-to-end **AWS ETL pipeline** efficiency, reducing time to production by 30% by developing FastAPI REST APIs and implementing security cloud **Infrastructure as Code** via Terraform and GitLab CI/CD actions
* Deployed a distributed KYC and risk management system, enhancing **financial fraud/risk analysis** rates by 17% and preventing $500,000 in losses using **Pub-Sub** for real-time data ingestion and XGBoost for anomaly detection
* Automated investment banking **transcript redaction**, reducing error rate to 3.5% and manual review time by 75% using OpenCV for document extraction and **spaCy** (NLP) for Named Entity Recognition and Parts of Speech Tagging

**projects**

**Imbalanced Meteorological Data Analysis for Rainfall Prediction |** [**GitHub**](https://github.com/NeuraNoir/Imbalanced-Meteorological-Data-Analysis-For-Rainfall-Prediction)December 2023

* Enhanced Class-1 accuracy to 77%, improving weather forecasting reliability by 25% by fine-tuning **multimodal** computer vision ConvLSTM2D models on GPUs for Meteorological Data Analysis (satellite imagery and time series)

**Next Sentence Prediction (Natural Language Understanding/Generation) |** [**GitHub**](https://github.com/NeuraNoir/Next-sentence-prediction) November 2023

* Achieved 85% accuracy compared to BERT (SoTA) by leveraging TensorFlow for NLP next sentence **text generation** in different languages using n-grams, GloVe word embeddings, and tokenization for Next Sentence Prediction

**CERTIFICATIONs**

• [**Microsoft** Certified](https://learn.microsoft.com/en-us/users/rishabhindoria-5834/credentials/e9c3d93e5b13053f): Azure Data Scientist Associate (DP-100) • [Bloomberg](https://portal.bloombergforeducation.com/certificates/Ge3MctSLAcGvYc4aPnR7r92j) Market Concepts