Rutvik Dhopate

(857) 753-7456 | Boston, MA - Open to Relocation | dhopate.r@northeastern.edu | Github | LinkedIn | Website

Education

Northeastern University **Sep 2023 — Apr 2025**
Master of Science, Data Science (GPA: 4.00 / 4.00) *Boston, MA*

NMIMS University **Jul 2020 — May 2022**
Master of Science, Applied Statistics and Analytics (GPA: 3.73 / 4.00) *Mumbai, IN*

Savitribai Phule Pune University **Jun 2017 — Apr 2020**
Bachelor of Science, Statistics (GPA: 3.71 / 4.00) *Pune, IN*

Work Experience

Northeastern University **Jan 2025 — Apr 2025**
Graduate Teaching Assistant - DS 5230/CS 6620 Unsupervised Machine Learning *Boston, MA*

- **Guided 80+ graduate students** addressing Mathematics and Python programming problems related to **Unsupervised Machine Learning,** like PCA, t-SNE, Latent Dirichlet Allocation, etc.
- Delivered 15-minute in-person **code walkthroughs** for **clustering, EM, and QA graphs**, enhancing conceptual clarity and hands-on skills.

Veeco Instruments Inc. **Jun 2024 — Dec 2024**
Data Scientist Co-Op *San Jose, CA*

- Developed a **Python-based executable application** to compute **Gaussian statistics** and beam profile metrics, reducing report consolidation time by over 75% compared to the previous MATLAB workflow.
- Invented and programmed a circular **Convolutional Neural Network** architecture to model resistance uniformity in Boron wafers, reducing manufacturing errors to 5% and **saved $1,200** per wafer in costs.
- Led a PoC comparing **Snowflake, Databricks, and Azure Migrate** for a **unified data lakehouse**, improving data accessibility by 30%.

Amazon **Jan 2022 — Aug 2023**
Data Scientist *Bangalore, IN*

- Engineered entity text extraction models with **90% precision** and **70% recall** using NER-MQMRC architecture in **Python** and **PyTorch.**
- Uncovered missing product information on Amazon Website across **19 countries** and 500 product categories to optimize customer experience.
- Spearheaded **LLM** research to enhance image-derived data, increasing labeled training data accuracy by 10%.
- Boosted attribute **extraction accuracy** by 2% via **sentiment-based classifiers** for niche US product types.
- Optimized **SQL** pipelines, **reducing data fetch latency** by 30% and accelerating preprocessing tasks.
- Designed an **anomaly detection workflow** by removing outlier dimensions for 800,000 products, increasing precision by 4%.

Tata AIG General Insurance Company **May 2021 — Jun 2021**
Data Analyst Intern *Mumbai, IN*

- Minimized insurance fleet premiums by 2.5% using **time-series forecasting** Vector Auto-**Regressive (VAR) models**.

Projects

CodeSynth: XAI Agent for Statistical Modeling, Artificial Intelligence Chat-Bot, Business Insights, LLMs **Mar 2025 — Apr 2025**

- **Deployed a Generative AI** model to perform statistical analysis on any structured dataset that achieved 88% success rate verified across benchmark DS1000 data.
- Built an **interactive User-Interface** that allows users to input datasets and receive dynamically generated Python scripts for data analytics and visualization using **prompt engineering**.

Recommendation System using DeepCoNN, Deep Learning, User Reviews, Cold-Start Recommendations **Jan 2025 — Apr 2025**

- Refactored **end-to-end** DeepCoNN PyTorch module to **jointly model** user behavior and item properties from review text, reaching an MSE of 1.57 to predict and recommend products to users.
- Integrated user and item embeddings to **recommend user-desired products** using a **MLOps**, which reduces re-training time by 4 hours.

The Delivery Cube, SQL, MySQL, Functions, Stored Procedures, Views, DQL **Oct 2023 — Dec 2023**

- Constructed 10+ tables and applied DDL, DML, and DQL operations to manage customers, orders, deliveries, and inventory efficiently.

Skills

- **Programming**: Python, PyTorch, SQL, R, MATLAB, Linux, TypeScript, SAS, HADOOP, NoSQL, Airflow, Apache, XML, Agile
- **Technologies**: Spark, Databricks, Pandas, Numpy, Matplotlib, MS Excel, Tableau, SAP, MySQL, PostgreSQL, Power BI, Azure, BigQuery