Rutvik Dhopate

857-753-7456 dhopate.r@northeastern.edu LinkedIn GitHub Boston, MA - Open to Relocation

Education

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

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

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

Experience

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

- Developed entity text extraction models with 90% precision and 70% recall using Multi-Question Machine Reading Comprehension (NER-MQMRC) architecture to identify missing product information on Amazon across 19 countries and 500 product categories to optimize customer experience.
- Resolved unit of measurements mismatch across 2 million product families by re-integrating NLP model architectures to predict unique dimensions for each product family.
- Launched research on Large Language Models (LLMs) to enhance image-based data, improving training data quality by 10%.
- Elevated attribute extraction performance by 2% using Sentiment Analysis for specific product types in the US marketplace.

Northeastern University Jan 2025 – Present
Graduate Teaching Assistant - DS 5230/CS 6620 Unsupervised Machine Learning and Data Mining *Boston, MA*

- Mentored 80+ graduate students addressing Mathematical and Python programming problems related to Unsupervised Machine Learning like PCA, t-SNE, Latent Dirichlet Allocation, etc.
- Presented effective in-person code demonstrations on clustering, EM Algorithms, and knowledge-based QA graphs.

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

- Engineered 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.
- Designed and incorporated a parallel batch processing system in Python to explore 30 data files, testing the hypothesis that lamp shutter timing significantly impacts peak variability in Laser Spike Annealing (LSA) signals.

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

- Reduced insurance fleet premiums by 2.5% using Vector Auto-Regressive (VAR) models.
- Ideated Pay-as-You-Drive scheme to introduce variable premiums in correspondence with miles driven.

Projects

AI Agent for Statistical Modeling | *Artificial Intelligence, Large Language Models, Chat-Bot, Business Insights*

- Deployed a fine-tuned Code-LLaMA to perform statistical analysis on any structured dataset that achieved 94% success rate verified across toy and real data.
- Built an interactive User-Interface that allows users to input datasets and receive dynamically generated Python scripts for statistical modeling and visualization.

Recommendation System using DeepCoNN | *Deep Learning, User Reviews, Cold-Start Recommendations*

- Re-implemented 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.

Measuring Financial Risk Using Extreme Value Theory | *Gumbel and Pareto Distributions, Financial Metrics*

- Investigated market data using Extreme-Value Distributions to assess 2% expected shortfall and value-at-risk metrics.

Technical Skills

Programming: Python, PyTorch, SQL, R, MATLAB, Linux, TypeScript, SAS, HADOOP, NoSQL, Airflow, Apache, XML
Technologies: Spark, TensorFlow, Pandas, Numpy, Matplotlib, MS Excel, Tableau, KPIs, CSS, AWS Frameworks, Git