HARSH BHATT

harsshbhatt0201@gmail.com | (857) 333-6608 | LinkedIn | GitHub | Google Scholar

EDUCATION

Northeastern University, Boston, MA **Sep 2023 – Jun 2025**
Master of Science in Data Science
Teaching Assistant for DS2500: Intermediate Programming for Data Science (Jan – Apr 2024, Jan – Apr 2025)

Nirma University, India **Aug 2019 – May 2023**
Bachelor of Technology in Computer Science and Engineering

PROFESSIONAL EXPERIENCE

Cohere Health **Jun – Dec 2024**
Data Scientist Co-op *Boston, MA, USA*

- **Developed and deployed 3 machine learning models** to automate prior authorization request approvals for a new clinical service area, **preventing $200,000+ in monthly operational costs**.
- Built integrated solutions utilizing **PySpark, Airflow and Sagemaker**, to **monitor data health**, predictive performance and **model drift** for a suite of 19 machine learning models, resulting in **faster fixes, instant Slack alerts and efficient performance tracking**.
- Engineered a **pipeline to calculate and visualize business metrics for 20+ live models** using **Tableau** dashboards, **boosting KPI tracking.**
- Enhanced existing models with improved scoring methods, conducted appropriate **model and data version control (DVC)**, **ensuring consistent performance** for new clients.
- Designed and executed **systematic subsampling experiments** to empirically determine the optimal data sizes for training and testing of new models, **establishing guidelines for efficient data collection, robust training and accurate testing**.

Sudeep Tanwar Labs [Article] **Jan – Jul 2023**
Student Researcher *Ahmedabad, Gujarat, India*

- Proposed and simulated a Federated Learning-based DL framework**,** achieving collaborative learning for clients without sharing data.
- Developed an aggregation algorithm leveraging optimizer weights, **resulting in 7% accuracy increase over conventional methods**.

SUNY Binghamton [Article] **Aug 2022 – Jan 2023**
Student Researcher *New York, USA*

- Utilized signal processing, thresholding and timestamp altering to **extract and analyze EEG signals** of autistic individuals.
- Developed a **novel CNN-based framewo**rk, **improving the F1-scores by 24%** enhancing the analysis of ASD individuals' attention spans.

D360 Technology Inc. **Jun – Aug 2022**
Machine Learning Intern *Surat, Gujarat, India*

- Engineered data transformation pipelines classifying data from 10+ vendors into a single structured schema.
- Developed an ensemble framework to classify new data into set features, **improving classification accuracy to 95% on unseen data**.
- Demonstrated strong technical and business acumen by expanding the statistical analyses, **resulting in insights into vendor patterns.**

PROJECTS

LLIME: Large Language model Integrated Medical feature Extractor [GitHub] **Oct – Dec 2024**

- Constructed a **patient note processing pipeline**, creating 1000+ high-quality samples to **fine-tune LLMs** for **medical keyword extraction**.
- **Fine-tuned Llama-2 7b** with **4-bit quantization** and **QLoRA,** enabling **4x more precise keyword extraction** from 25+ word inputs with an **average 0.4 improvement in ROUGE scores**.
- Conducted **prompt engineering** and **model ablation studies** to **boost fine-tuning** and **inference performance by 8%**
- Developed a user-friendly **Streamlit** app for **easy input, prompt construction, model loading and inference** with interpretable outputs.

CORAL-X: Contextual Risk Assessment for Loan Applications using LLMs for Explainability [GitHub] **Jan – Apr 2025**

- Engineered a **loan risk assessment framework** integrating an **XGBoost** model with a **RAG-**powered **Llama-3.2**, which leveraged regulatory documents retrieved from **Pinecone** to deliver **accurate, contextually relevant and auditable LLM-generated explanations**.
- Designed an **LLM-as-a-judge module**, scoring the generated explanations on **custom metrics for human monitoring and feedback**.
- Led **comprehensive ablation studies** to **benchmark model choices**, demonstrating the pipeline's superiority for explanation quality.
- Delivered a **modular workflow deployable on local and cloud GPU environments** with a **Streamlit** app for user interaction.

End-to-end MLOps Pipeline for Walmart Supply Chain Forecasting [GitHub] **Feb – Mar 2025**

- Engineered a **pipeline to transform raw data**, and **predict sales for 40+ stores**, **achieving 92% accuracy for weekly forecasting**.
- **Automated data validation, transformation, and storage workflows with S3 integration**, **ensuring quality real-time updates and DVC.**
- Integrated **MLflow** for **experiment tracking**, logging parameters and metrics, **improving decisioning and forecasting accuracy by 8%.**
- Deployed the pipeline on **AWS EC2** with **Docker** and **GitHub Actions**, **streamlining CI/CD processes** for scalable supply chain analytics.

SKILLS

Competencies: LLM fine-tuning, RAG pipelines, Prompt Engineering, ML, ETL, Data Pipelines, CI/CD, Experiment Tracking, Model Deployment
Programming Languages: Python, SQL, R, Java, C++
Tools: PyTorch, Langchain, PySpark, Scikit-Learn, WandB, MLflow, Docker, S3, Sagemaker, EC2, Athena, Airflow, Pinecone, GitHub Actions