HARSH BHATT

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

EDUCATION

Northeastern University, Boston, MA **Sep 2023 – Jun 2025**
Master of Science in Data Science

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

PROFESSIONAL EXPERIENCE

Cohere Health (3rd on LinkedIn 2024 Top Startups) **Jun – Dec 2024**
Data Scientist Co-op *Boston, MA, USA*

- Developed and deployed 3 machine learning (ML) models to automate prior authorization request approvals for a new clinical service area, saving $200,000+ in monthly operational costs.
- Built integrated solutions utilizing PySpark, Airflow and Sagemaker, for data version control (DVC), and monitoring model drift for 19 ML models, resulting in consistent performances and instant failure alerts**.**
- Engineered data pipelines and Tableau dashboards to calculate and monitor KPIs for 20+ live models, boosting performance tracking**.**
- Enhanced existing ML models with improved scoring methods, reducing false approvals by ~2% while maintaining true approval rates.
- Designed and executed systematic subsampling experiments to determine optimal training and testing data sizes for new models, establishing guidelines for efficient data collection and future model development.
- Conducted ad-hoc analysis of Provider programs and Clinical notes, assisting Product and Clinical teams to make data-driven decisions.

Northeastern University **Jan – May 2025, Jan – May 2024**
Teaching Assistant: Intermediate Programming for Data Science *Boston, MA, USA*

- Mentored 300+ students in Data Science theory such as statistics, data handling, NLP, supervised and unsupervised ML.
- Spent 150+ hours assisting students in Python, Scikit-learn & NLTK code debugs, significantly reducing supervisor workload.

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

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

SUNY Binghamton [Peer Reviewed Publication] **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 framework, 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 schema, improving classification accuracy to 95%.

PROJECTS

Facebook Advertisement Campaign Analytics [GitHub] **Apr – May 2025**

- Analyzed 1,100+ Facebook ads, identifying key factors to achieve maximum ROAS and conversion rates (CR).
- Utilized A/B tests across interest groups, uncovering significant ROAS and CR driving differentiators across 3 ad campaigns.
- Developed a Tableau dashboard to visualize CTR, CPC, and other KPIs, enabling rapid identification of high-performing segments and actionable optimization opportunities.

Multi-Touch Attribution Analysis and Budget Optimization for Marketing Channels [GitHub] **Apr – May 2025**

- Implemented multi-touch attribution models on ~600,000 customer interactions across 5 marketing channels, enabling accurate measurement of channel conversion rates and ROI.
- Developed data transformation pipelines to convert raw impression data into structured customer journey paths, facilitating advanced attribution analysis for 250,000 unique users.

- Constructed a Streamlit application for dynamic simulation of optimal budget allocation for varying scenarios.

End-to-end MLOps Pipeline for Walmart Supply Chain Forecasting [GitHub] **Feb – Mar 2025**
- Built an ML pipeline to transform data and predict sales for 40+ stores, achieving 92% accuracy for weekly forecasting.
- Automated data validation, transformation, and storage workflows with S3, ensuring quality real-time updates and DVC.
- Integrated MLflow for experiment tracking, logging parameters and metrics, increasing model improvement tracking.
- Deployed the pipeline on AWS EC2 with Docker and GitHub Actions, streamlining CI/CD processes for scalable supply chain analytics.

End-to-end ML-based Ad Slot Reserve Price Prediction System [GitHub] **May –Jun 2025**
- Built a scalable ML pipeline using PySpark, MLlib and XGBoost to predict ad slot reserve prices using ~500k real-time auction records, improving reserve price matching accuracy.
- Automated data ingestion and analytics using AWS S3, Glue, and Athena, enabling real-time querying and data management.
- Integrated DVC and experiment tracking using WandB, ensuring workflow reproducibility and continuous model improvements.
- Developed interactive Grafana dashboards for real-time model health and KPI monitoring, delivering actionable model performance insights and supporting data-driven decisions for ad inventory management.

PHIlter: Agentic AI system for Personal Health Identifier (PHI) detection [GitHub] **Jun – Jul 2025**
- Developed a modular, multi-step, agentic system using LangGraph to process clinical documents and accurately extract PHI instances.
- Engineered dynamic file processing with OCR pipelines, to successfully handle image, unstructured documents and tabular inputs.
- Integrated GPT-4o LLM agents to automate PHI recognition and categorization, producing verifiable outputs to assist HIPAA compliance.
- Utilized LangSmith tracing to monitor and debug workflow runs, successfully handling node failures, edge cases and output processing.

LLIME: Large Language model Integrated Medical feature Extractor [GitHub] **Oct – Dec 2024**
- Constructed a clinical note processing pipeline, creating 1000+ high-quality samples to fine-tune LLMs for medical keyword extraction.
- Fine-tuned Llama-2 with 4-bit quantization and QLoRA, increasing keyword extraction precision by 4x and ROUGE scores by ~0.4.
- 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 full-end system utilizing XGBoost models for loan risk decisions and Llama-3 for generating decision rationale.
- Built a RAG-pipeline, using policy documents stored in a Pinecone vector store to generate source-driven, auditable LLM explanations.
- Used SHAP values for each case as part of LLM prompt, achieving case-specific feature importance interpretation and explanation.
- Designed an LLM-as-judge module, scoring the generated explanations on custom metrics for human audit and feedback.
- Delivered a modular workflow deployable on local and cloud GPU environments with a Streamlit app for user interaction.

Fashion Recommendation System using Hybrid Filtering [GitHub] **Feb – Apr 2024**
- Web-scraped a fashion products website to gather data on 9000+ products, 100,000+ user and 250,000+ product reviews.
- Designed a hybrid recommendation system consisting of rating-based collaborative filtering, user body fit-based collaborative filtering and content-based filtering, resulting in 3x more relevant recommendations for a new user.
- Implemented SVD, PCA and UMAP for up to 40% reduction in dimensionality, resulting in 1.5x faster compute, 10% increase in recommendation precision, and up to 30% increase in recommendation diversity.

SKILLS

Competencies: LLM fine-tuning, RAG pipelines, Prompt Engineering, Agentic Workflows, ML, CI/CD, Experiment Tracking, Model Deployment, Hypothesis testing, Forecasting, Data Wrangling, Data Mining, ETL pipelines
Programming Languages: Python, SQL, R, Java, C++
Tools: PyTorch, Langchain, LangGraph, LangSmith, Scikit-Learn, WandB, MLflow, Docker, Sagemaker, EC2, Airflow, Pinecone, GitHub Actions, PySpark, Tableau, Grafana, Mllib, S3, Athena, Glue