

UTKARSH SARAOGI

857-426-9744 | saraogi.u@northeastern.edu | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

EDUCATION

Northeastern University, Boston, MA

Exp. May 2027

Master of Science in Computer Science

- **Coursework:** System Design, Database Management Systems, Algorithms, Machine Learning, NLP

Manipal Institute of Technology, Manipal, KA

May 2022

BTech in Electronics and Communication

- **Coursework:** Data Structures & Algorithms, Data Science, Python Programming, Cloud Computing

TECHNICAL SKILLS

Programming: Python, SQL, Java
Data Science: A/B Testing, Statistical Analysis, Causal Inference, Matplotlib, Seaborn, Pandas
Machine Learning: Scikit-learn, SMOTE, Cross Validation, Evaluation Metrics
Data Engineering: MySQL, MongoDB, Airflow, Spark, dbt, Snowflake
Cloud & DevOps: Git, Jenkins CI/CD, Docker, Jenkins, AWS (IAM, SageMaker, S3, Lambda), REST/GraphQL API

PROFESSIONAL WORK EXPERIENCE

HEWLETT PACKARD ENTERPRISE

Aug. 2022 – Aug. 2025

Cloud Developer I

Bangalore, India

- **Data Pipeline Observability:** Monitored switch data flow across 15+ distributed services using real-time logs; **identified pipeline failures** by pinpointing exact service breakpoints, data anomalies, and root causes—**reducing mean time resolution by 40%**.
- **Regression Optimization:** Automated data validation and regression testing across 1,200+ test cases; **reduced test-cycle times by 55%** through parallel execution and optimized integration testing processes.
- **ML-Powered Triage System:** Contributed to the training of an AI assistant to automate failure classification by sourcing and labeling data through **4,000+ failure scenarios** (test, feature, server), achieving **65% improvement in classification accuracy** and reducing manual triage time.
- **CI/CD & Deployment:** **Owned end-to-end testing lifecycle**—developed smoke, integration, and regression test suites; automated execution via Jenkins CI/CD pipeline; validated production releases through post-deployment testing.

HEWLETT PACKARD ENTERPRISE

Jan. 2022 – Jul. 2022

Graduate Intern

Bangalore, India

- **Test Automation & QA:** Executed daily automation suites, identified software defects, and contributed fixes to maintain test framework health for the Switch Data Monitoring platform.
- **Code Quality Automation:** Created automated PyLint validation scripts integrated into the CI pipeline, eliminating production rejections due to lint errors and reducing pre-release debugging time.

PERSONAL PROJECTS

SoundStream Analytics | *ELT, dbt, Airflow, duckDB, Pyspark, SQL* | [Link](#) | [Live](#)

- Built a **production-grade ELT pipeline** processing 46M+ music records using **medallion architecture**, implementing raw ingestion, deduplication and star schema layers with **Airflow orchestration**.
- Optimized data quality and performance: Removed **5,043 duplicates**, handled malformed data, reduced storage by **60%** through Parquet format with region/date partitioning.

Credit Risk Strategy Engine: Revenue Recovery | *A/B Testing, Scikit-Learn* | [Link](#)

- Developed a fraud-resistant risk engine using HistGradientBoosting and Counterfactual A/B Testing to replace rigid threshold policies with a probabilistic scoring framework; recovered \$7.8M in potential revenue by identifying low-risk applicants and reduced false rejections by 35% while maintaining FCRA compliance via an Explainable AI module.

Stock Portfolio Management System | *Python, MySQL, REST-APIs, ETL* | [Link](#)

- Built an **end-to-end data pipeline**; extracted real-time market data from the Finnhubb API, transformed via SQL stored procedures, loaded normalized schema into MySQL supporting portfolio analytics; served it to frontend using REST API.

ACHIEVEMENTS & RESEARCH ACTIVITIES

- Researching LLM sycophancy at Northeastern—analyzing model responses across 3 models using hidden state probes to detect when models abandon correct answers under social pressure. Implementing a guardrail to prevent flips.
- Mentored 3 interns at HPE, creating comprehensive documentation and held live sessions that reduced onboarding time by 1-2 months.