

KRUTIK BAJARIYA

| bajariya.k@northeastern.edu | +1 (207) 409-4704 | <https://www.linkedin.com/in/krutikbajariya> |

PROFESSIONAL EXPERIENCE

COASTAL MEASURES | *Intern - Data Scientist and Software Developer* May 2024 – Dec 2024

- Enhanced data accessibility for the US Army Corps of Engineers by developing an end-to-end ETL pipeline using AWS and a centralized data lake architecture, increasing data usability by 37%, enabling better decision-making.
- Saved \$2,000 annually by eliminating reliance on external weather APIs and implementing NOAA Data Models - ERA5 scraping system, ensuring continuous and cost-effective data solutions.
- Improved tidal forecasting accuracy by 12% compared to NOAA by developing a Logistic Regression model using weather features, enabling more precise coastal and maritime planning.

TCS | *Assistant Systems Engineer - Data Warehouse Executive* Jul 2021 – Jul 2023

- Upgraded fraud detection for FinTech client Aviva Insurance Ireland by leading the "Fraud Claim Detection" Machine Learning project, improving risk management and reducing fraudulent payouts by 7%.
- Refined reporting efficiency by generating 16 quarterly business statistics reports using IBM Cognos & Qlik Sense in Excel format, enabling data-driven decision-making for business stake-holders.
- Reduced job failures by 8% and ensured 100% compliance by resolving legacy PL/SQL & Unix (Linux) script failures, improving data integrity for business-critical reports and data analytics.

TECHNICAL SKILLS

Scripting / Programming Languages: Python, C++, Shell Script, PowerShell, Java, PHP, Javascript, MATLAB, R

DataBase Tools & Technologies: SQL Developer (Oracle), MySQL, PL/SQL, DataStage, Databricks, MongoDB

Data Science Libs: Numpy, Pandas, Matplotlib, Seaborn, PySpark, Dask, Boto3, GCloud, PyTest

Cloud Services: AWS (S3, EC2, RDS, Lambda, CloudWatch, API Gateway, VPC, IAM), GCP, Looker, Salesforce

ACADEMIC PROJECTS

Real-Time Stock Market Data Pipeline | *PySpark & Technical Analysis* May 2025

- Designed automated stock trading pipeline processing 55K+ records using PySpark and Alpha Vantage API, generating technical indicators and risk metrics that identified top performers (NFLX: 23.3% return) and underperformers (AMD: -14.8% return) for data-driven portfolio optimization.
- Built PySpark-based trading signal system calculating RSI and moving averages, generating actionable BUY/SELL/HOLD recommendations through [Tableau](#) for portfolio optimization.

Remo Data Analysis | *Data Management and Preprocessing* Aug 2024

- Formulated a clustering model using *K-means* and *t-SNE* on a dataset of over 72K books, successfully grouping books into 4 distinct genres. Achieved 95% data completeness by resolving missing values using imputation and external sources.
- Established a data pipeline for book metadata, binarizing multi-label genres with *MultiLabelBinarizer*. Enhanced clustering performance by reducing data redundancy, cleaning 3,898 records, and achieving clear separations in visualized clusters.

EDUCATION

Northeastern University Jun 2025

Master of Science in Computer Science GPA: 4.0 / 4

University of Mumbai Jun 2021

Bachelor of Engineering in Information Technology GPA: 8.45 / 10