

# KRUTIK BAJARIYA

| [krutikbajariya123@gmail.com](mailto:krutikbajariya123@gmail.com) | +1 (207) 409-4704 | <https://www.linkedin.com/in/krutikbajariya> |

## PROFESSIONAL EXPERIENCE

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

- Developed an end-to-end real-time ETL pipeline using AWS (S3, Lambda) to ingest data into a centralized data lake for the U.S. Army Corps of Engineers, improving data accessibility by 37% and supporting faster decision-making
- Improved tidal forecasting accuracy by 12% compared to NOAA's in-house staff, by engineering weather feature datasets and developing a Logistic Regression model, enabling more precise coastal and maritime planning
- Prototyped RAG pipelines using FAISS to combine ERA5 weather and NOAA tidal data, boosting context-aware tidal forecast accuracy by 4% and cutting report time by 16%.

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

- Provided production support by resolving legacy PL/SQL & Unix (Linux) script failures, improving data pipeline reliability, reducing job failures by 8%, and ensuring 100% compliance for business-critical analytics
- Automated quarterly reporting infrastructure using Cognos (business intelligence tools) & CRON (automation tools) to generate and mail 16 business statistics reports, enabling data-driven decision-making for business stakeholders
- Developed CI/CD pipeline for fraud detection by supporting the "Fraud Claim Detection" Machine Learning project for client Aviva Insurance, contributing to improved risk management and 7% (€15.6 Million) reduction in payouts

## TECHNICAL SKILLS

**Data Engineering & ETL :** Apache Spark (PySpark), ETL (DataStage), SQL Developer (Oracle), Airflow

**Cloud Technologies:** AWS – Storage & Compute (S3, EC2, Lambda, Batch, Sagemaker), Data Services (Redshift, RDS, API Gateway), Security (IAM, VPC), GCP – BigQuery, Looker

**Programming Languages:** Python, SQL (PL/SQL, MySQL), Shell Script

**Data Processing & Analytics:** Numpy, Pandas, PySpark, Dask, Boto3, Matplotlib, Seaborn

**Agile Development Tools:** Git (GitHub, BitBucket Stash), Jira, Confluence (Documentation)

**Additional Technologies:** Scikit-Learn, PyTorch, Tensorflow, Transformers, Hugging Face, R, MATLAB

## ACADEMIC PROJECTS

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

- Designed PySpark pipeline to process 55K+ stock records via Alpha Vantage API with automated data ingestion
- Created financial features – risk metrics, RSI, and moving averages; optimized for real-time signal generation
- Deployed end-to-end pipeline with [Tableau](#) visualization, identifying top performers (NFLX: +23.3%) and enabling data-driven portfolio optimization

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

- Built data pipeline for 72K+ book records, with 95% completeness – automated imputation and metadata enrichment
- Implemented data quality checks and cleaning for 3,900+ records, and applied MultiLabelBinarizer to encode multi-label categorical fields, optimizing the dataset for analysis and reducing data duplication
- Applied unsupervised-learning(K-means, t-SNE); segment data into 4 distinct genres, enhancing data interpretability

## 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