

KRUTIK BAJARIYA

| krutikbajariya123@gmail.com | +1 (207) 409-4704 | [LinkedIn](#) | [Portfolio](#) |

PROFESSIONAL EXPERIENCE

COASTAL MEASURES | Data Scientist and Software Developer

May 2024 – Dec 2024

- Built end-to-end ETL pipeline using **AWS services** (S3, Lambda, Redshift) and Apache Airflow to ingest real-time data for U.S. Army Corps of Engineers, improving data **retrieval time by 37%** and processing **50K+** daily sensor readings
- Developed **REST APIs** for large file uploads (>10MB) with proper error handling and retry mechanisms; learned MATLAB integration to support non-technical stakeholder requirements and ensure seamless data ingestion workflows
- Enhanced tidal forecasting pipeline by engineering weather feature datasets and implementing logistic regression model, achieving **12% accuracy** improvement over NOAA's baseline model for coastal planning applications
- Implemented data integration solution using **FAISS** vector search to merge ERA5 weather and NOAA tidal datasets, applying entity resolution techniques to **improve forecast accuracy by 4%** and reduce data preparation time

TCS (Aviva Insurance) | Assistant Systems Engineer - Data Engineer

Jul 2021 – Jul 2023

- Provided production support for critical Qlik and PL/SQL data pipelines serving business analytics, **reducing job failures by 23%** through improved monitoring, debugging, and ensuring **100% adherence** to data governance standards
- Automated reporting workflows using Cognos, QlikView, and Python with CRON scheduling to **generate 16 quarterly business reports**, eliminating manual processes and enabling consistent data delivery to stakeholders
- Supported fraud detection ML pipeline by building CI/CD workflows and data preprocessing components, contributing to **7% reduction in fraudulent claims payouts (€15.6M savings)** and improving model deployment reliability
- Collaborated with cross-functional teams including front-end developers, business analysts, and brokers to understand requirements and deliver production-ready data solutions on schedule

TECHNICAL SKILLS

Data Engineering & ETL : Apache Spark (PySpark), ETL (DataStage), Snowflake, dbt, CRON, Airflow, Databricks (basic)

Cloud Technologies: AWS (S3, EC2, Lambda, Batch, SageMaker, Redshift, RDS, API Gateway, IAM, VPC)

Programming Languages: Python, SQL (PL/SQL, MySQL, PostGreSQL), Shell Scripting, NoSQL(MongoDB)

Data Processing & Analytics: Pandas, NumPy, PySpark, Boto3, Matplotlib, Seaborn

Agile Development Tools: Git (GitHub, BitBucket Stash), Jira, Confluence, Docker (basic), Linux/Unix

BI Tools: Tableau, Cognos, Qlik, Looker

ACADEMIC PROJECTS

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

May 2025

- Designed a **PySpark** pipeline with **Airflow** to process 55K+ stock records and [Tableau](#) dashboards for real-time analytics.
- Created financial analytics module calculating technical indicators (RSI, moving averages, volatility metrics) and identifying top-performing stocks like NFLX (+23.3%), enabling data-driven investment insights

ReMo Book Recommendation ([Post](#)) | Natural Language Processing, Recommendation System

Jun 2024

- Managed a team of 6 to develop a recommendation module using “**facebook/bart-large-mnli**” & “**all-MiniLM-L6-v2**” models, thereby increasing educator & student engagement by **42%**
- Preprocessed the 72K book synopsis data using SpaCy & NLTK and created a **51k x 51k topic modelled cosine similarity matrix**, which generated 25 recommendations based on previous reading history & genre

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