

Education

Northeastern University

Master of Science in Information Systems, AI, Machine Learning & Data Science

Sept 2022 - July 2024

Boston, MA

University of Mumbai

Bachelor of Engineer in Information Technology

May 2012 - May 2016

Mumbai, India

Work Experience

Graduate Research Assistant (Software Engineering & Data Science)

Feb 2024 - Present

Northeastern University | Tech Stack: HTML, CSS, Node JS, Python, Qualtrics, LimeSurvey, R, Prolific, XGBoost Boston, MA

- Designing **4-condition** online instrument to simulate digital dashboard for study on visual communication & spatial ability

Data Analytics Engineer Co-Op

May 2023 - Dec 2023

Novelis | Tech Stack: Power BI, Salesforce, Arize, **Azure** Data Workflows, Azure, MLOps, Snowflake, Databricks Atlanta, GA

- Crafted data analysis strategies by integrating Salesforce in Power BI dashboards elevating reporting efficacy by **90%**
- Streamlined operations by migrating R&D Accolade data to Snowflake, monitoring **15 key metrics** such as Budget, IP, Projects
- Implemented **CI/CD pipeline** using Data Factory for ingestion, Blob storage for storage, Databricks for model training, enhanced data warehousing with SQL and Spark for processing workflows with **40%** proficiency, saving **100+ manual hours**
- Built Azure **MLOps pipeline** via DevOps, App Service, Docker, GitHub Actions, CLI, cutting ML models rollout time to **30 mins**
- Constructed **35 out of 68** web-app modules using React, FastAPI, PostgreSQL, Alembic for alert generation system in **2weeks**

Graduate Research Assistant (Machine Learning)

Feb 2023 - May 2023

Northeastern University | Tech Stack: Computer Vision, CNN, DeepFace, Image Analysis, AWS Sagemaker, HPC Boston, MA

- Engineered features images via OpenCV and Keras, processing **1000+ records** to compute **20** attributes in **30 secs each**
- Deployed CNN with Deepface model on **AWS Sagemaker** for face recognition, **increasing** precision and F1 score by **25%** for **Marketing Academia Research**, while also sharing diversity insights using Tableau reporting

Data Engineer

Dec 2020 - Aug 2022

UST Global Xpanxion | Tech Stack: **AWS**, Databricks, Hadoop, Terraform, Colibra, Salesforce Pune, India

- Led Hadoop data prep, achieved **20% surge**, to cut generation time from **12 to 1 hour**, with tuned ETL jobs in Teradata
- Migrated to **serverless data pipeline** to AWS Glue, Redshift, Lambda, Step Functions, Terraform for **credit reporting agency client** elevating data influx **25x**, trimmed and saved daily execution time by **4 hours** in **event-driven architecture**
- Improved time series anomaly detection efficiency on **3TB unstructured data** by **30%** using AWS resources like TensorFlow, AutoML, precise hyperparameter tuning, conducting A/B testing for validation
- Managed sprint planning and execution of revamping credit product documentation for team of **5** data analyst, engineers and scientists, resulting in **50% reduction** in new-hire onboarding time

Other Work Experience

Senior Software Engineer Analyst

Mar 2019 - Dec 2020

Accenture | Tech Stack: ETL, ELT, Restful APIs, AWS (EMR, S3, Redshift), PySpark, SQL, Stored Procedures, UI/UX Mumbai, India

System Engineer

Mar 2017 - Mar 2019

Infosys | Tech Stack: Software Engineering, Core Java, React JS, MongoDB, GraphQL, Automation Testing Pune, India

Academic Projects

Predict Diabetes | Tech Stack: Prima Diabetes Dataset, Random Forest, Neural Network

Jan 2024

- Trained and evaluated Keras Neural Network model with 200 epochs, achieving 76.0% accuracy and 81.0% ROC-AUC score

Aircast - Air Quality Forecasting | LSTM Forecasting | Data Pipeline | Data Augmentation | MVP Architecture

Apr 2023

- Implemented Data-as-a-Service (DaaS) solution on AWS, integrating data scraping, Amazon RDS storage, and LSTM modeling for air quality forecasting (RMSE: 0.8)
- Orchestrated MLaaS, FastAPI, Airflow for historic batch and daily data processing, NGINX, Docker, Kubernetes, CloudWatch

Ride Fare Forecasting (Uber/Lyft) | Tech Stack: Decision Tree, Gradient Boosting Regressor

Mar 2023

- Analyzed datasets on Uber and Lyft for availability, car preferences, price-weather, distance correlation to predict fare prices

Technical Expertise

Languages & Packages: Python, R, Java, Scala, Scikit Learn, Keras, PyTorch, Tensorflow, CUDA, OpenMP (C++), Dask, NLP

Databases: MySQL, PostgreSQL, DBT, Hadoop, Teradata, Cassandra, Oracle SQL, AWS (RDS, Aurora, DynamoDB)

Data Science: Neural Networks, Deep Learning, Hugging Face, LLM, Transformers, Explainable ML, Bayesian Modeling, Survival model, Ensemble modeling, Gaussian Mixture, Multivariate Analysis, A/B Testing, Hypothesis Testing, Monte-Carlo Simulation

Big Data Tools: AWS (EC2), Azure, Kafka [Confluence], Astronomer, Kubernetes, Data Warehouse, Terraform, Looker, Arize

Achievements

Promotion to Senior SDE at Accenture

Mar 2020

Pat on the Back Award at UST-Xpanxion

Jan 2022

Certifications & Student Clubs

MS Azure Fundamentals AZ-900

Feb 2020

IBM Coursera Data Science Capstone SpaceX Project

Jan 2024