

## VARSHA HINDUPUR

(857)-693-9108 | [hindupur.v@northeastern.edu](mailto:hindupur.v@northeastern.edu) | [linkedin.com/in/varsha-hindupur-4aaa83175](https://www.linkedin.com/in/varsha-hindupur-4aaa83175) | ([website](https://github.com/varshahindupur09)) [varshahindupur09.github.io](https://github.com/varshahindupur09)

### EDUCATION

#### Northeastern University

Master of Science in Information Technology, ML Engineering & Data Science

Sept 2022 - July 2024

Boston, MA

#### University of Mumbai

Bachelor of Engineer in Information Technology

May 2012 - May 2016

Mumbai, India

### TECHNICAL SKILLS

**Programming:** Python, SQL, DBT, Spark, Java (SpringBoot), R, C/C++, JavaScript (Angular, React, Node), Scala, Julia

**AI, ML & Data Science:** Data Modeling, Deep Learning, Regression & Classification, LLM, Hugging Face, NLP, LLM, Scikit-Learn, AutoML, Tensorflow, Keras, PyTorch, SAS, Hadoop, Hive, FlaskAPI, FastAPI, Terraform, Teradata, Selenium

**Cloud:** AWS, Azure, Kafka [Confluent], GitHub Actions, Docker, Snowflake, Databricks, Airflow, Astronomer, Kubernetes

**Certification:** MS Azure Fundamentals AZ-900 (Feb 2020), **Publication:** Data Transfer Through Human Body (Jan 2016)

### WORKS EXPERIENCE

#### Data Analysis Co-Op

May 2023 – Dec 2023

Novelis - An Aditya Birla Company | Tech Stack: Hugging Face, Arize, Azure, Power BI, Salesforce, Snowflake

Atlanta, GA

- Executed predictive modeling with Hugging Face, **validating equipment reliability** in diverse manufacturing plants
- Built Azure MLOps pipeline via DevOps, App Service, Docker, GitHub Actions, CLI; cutting models rollout time to **30 mins**
- Built **"EagleGPT"** on top of ChatGPT, by improving conversational AI dialogues, model responses, & backend system
- Built KPIs - budget allocated, project completed, IP generated in Snowflake to optimize projects & foster invention
- **Extra-Efforts:** Built & integrated **51.47%** (35 out of 68) web-app modules - React, FastAPI, PostgreSQL, Alembic, Postman

#### Graduate Research Assistant (Machine Learning)

Feb 2023 - May 2023

Northeastern University | Tech Stack: Computer Vision, Image Recognition & Engineering, AWS

Boston, MA

- Conducted image feature engineering via OpenCV, computing **20** attributes in **30 secs** for each **1000+ records**
- Performed identity verification, customer profiling, fraud detection via Deepface model on AWS SageMaker; achieved **25% boost in precision, recall, and F1 score** for age, race, gender, and emotion extraction for Marketing Research Academia

#### Data Engineer

Dec 2020 - Aug 2022

UST Global (Xpanxion) | Tech Stack: Apache Kafka, NiFi, Restful APIs, Airflow DAGs, GCP, Docker, Snowflake

Pune, India

- Architected ETL pipeline using Kafka, NiFi, Snowflake; elevated data influx **25x**, trimmed daily execution time by **4 hours**
- Designed APIs to integrate Collibra & Salesforce data via Databricks; optimized data processing for reuse in **6 projects**
- Amplified model accuracy by **30%** with GCP's TensorFlow & AutoML optimizing hyperparameter tuning, data augmentation, ensemble learning, batch normalization, pruning, and learning rate scheduling

#### Senior Software Engineer Analyst

Mar 2019 - Dec 2020

Accenture | Tech Stack: UI/UX, ETL, Hadoop, Kafka, NiFi, Restful APIs, Airflow DAGs, AWS

Mumbai, India

- Re-architect multi-page app with product managers, optimizing user experience, enabling scalability for **1000+ users**
- Championed creation of in-house UI library, building reusable components & saving **125+ hours per month**
- Overhauled Hadoop data prep, achieved **20% surge**, cut creation time from **12 to 1 hour**, with tuned ETL jobs in Teradata
- Configured ETL Pipeline using AWS EMR, S3, Redshift achieving **30% reduction** in data processing time in **3 weeks**

#### System Engineer

Mar 2017 - Mar 2019

Infosys | Tech Stack: PySpark, SQL, Angular, JavaScript, Automation Testing | Java

Pune, India

- Produced **100+ PySpark reports**, attained **75% reporting efficacy** with queries, views, stored procedures, windows functions
- Created online banking, accounts, payment processing features via AngularJS (frontend) & Flask APIs (backend)
- Built testing framework with **30+ distinct cases** using Java, Selenium, TestNG, Karate, RestAssured for CI/CD pipeline

#### Business Intelligence Intern

Nov 2016 - Jan 2017

Rendered Ideas | Tech Stack: Sentimental Analysis, Microsoft Excel, SAS, ETL

Mumbai, India

- Revamped player engagement analysis using **SAS** platform, contributed in reducing ETL time by **8 hours** and **10% revenue growth** through initial exposure to predictive modeling, and participation in orchestrating complete BI life-cycle
- Optimized sentimental analysis by dataset construction via feedbacks, creating **10x surge in game downloads** worldwide

### ACADEMIC PROJECTS

#### Aircast - Air Quality Forecasting | MLOps Engineering

Apr 2023

Tech Stack: Python Streamlit, Docker, Airflow DAGs, NGINX, PostgreSQL, Postman, FastAPI, AWS (CloudWatch, RedShift, EC2)

- Crafted air quality forecasting with **LSTM**, achieving **0.85 RMSE**; orchestrated end-to-end solution with **MVC architecture** principles; incorporated data-as-a-service and machine-learning-as-a-service paradigms

#### Predictive Analytics for F1 Race | Machine Learning | Data Science | Tech Stack: Rapid Miner, Java, Folium

Apr 2023

- Leveraged **Data Mining techniques** (Apriori algorithm, Naive Bayesian) with Folium for comprehensive analysis of 2022 Formula 1 statistics; crafted predictive model for accurately anticipating match outcomes

#### Meeting Intelligence Application | Data Infused Software Development | Tech Stack: Great Expectations

Mar 2023

- Adopted **OpenAI's DALL-E** for TEXT to IMAGE conversion, transcribed audio data using **Whisper API**, and combined ChatGPT APIs for answering meeting-related queries. Hosted this application on Amazon **EC2** instance

#### Ecommerce Analysis for Online Retail Store | Machine Learning | Tech Stack: K-Means Clustering Algorithm

Mar 2023

- Conducted in-depth analysis on online store data, calculating customer **Recency, Frequency, Monetary** values