Varsha Hindupur

🏠 Boston, MA 📞 (857)-693-9108 ✉ hindupur.v@northeastern.edu 💼 varsha-hindupur 🌐 varshahindupur09 📖 varsha.hindupur

Education

Northeastern University	**Sept 2022 - July 2024**
Master of Science in Information Systems, AI, ML & Data Science	*Boston, MA*
University of Mumbai	**May 2012 - May 2016**
Bachelor of Engineer in Information Technology	*Mumbai, India*

Technical Expertise

Programming & Frameworks: Python, Django, R, Java SpringBoot, OpenMP, ReactJS, NodeJS, Spark Scala, FlaskAPI, Selenium
Databases: MySQL, PostgreSQL, DBT, Hadoop, Hive, Teradata, Cassandra, MongoDB, Oracle SQL, AWS (RDS, Aurora, DynamoDB)
AI, ML & Data Science: Regression, Classification, Pattern Recognition, CNN, RNN, Neural Networks, Deep Learning, NLP, Scikit Learn, Hugging Face, Object Detection, Large Language Models (LLM), AutoML, Tensorflow, Keras, PyTorch, CUDA, A/B Testing
Tools: AWS, Azure [Data Factory], Kafka [Confluence], GitHub Actions, Docker, Snowflake, Databricks, Airflow, Astronomer, Kubernetes, Collibra, Data Warehouse [RedShift], Tableau, Postman, Swagger, Terraform

Work Experience

Graduate Research Assistant (Software Engineering)   **Feb 2024 - Present**
Northeastern University | Tech Stack: HTML, CSS, Node JS, Python, Qualtrics, LimeSurvey, R, Prolific   *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 - An Aditya Birla Company | Tech Stack: Power BI, Salesforce, LLM, Arize, Azure, MLOps, Snowflake   *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
- Built Azure MLOps pipeline via DevOps, App Service, Docker, GitHub Actions, CLI, cutting ML models rollout time to **30** mins
- Molded 'EagleGPT' on top of ChatGPT, by improving conversational **AI** dialogues, model responses, and backend system
- Performed real-time IOT sensor data analysis using Databricks Streaming and Delta Lake for decision making & data integrity
- **Extra-Efforts:** Constructed **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, 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 gathering diversity data

Data Engineer   **Dec 2020 - Aug 2022**
UST Global Xpanxion | Tech Stack: GCP, AWS, Databricks, Hadoop, Terraform, Snowflake, Collibra, Salesforce   *Pune, India*
- Calibrated **ETL pipeline** using Kafka, NiFi, Snowflake, elevating data influx **25x**, trimmed daily execution time by **4 hours**
- Overhauled Hadoop data prep, achieved **20% surge**, to cut generation time from **12 to 1 hour,** with tuned ETL jobs in Teradata
- Developed ETL pipeline to AWS Glue, Redshift, Lambda, Step Functions, Terraform for **credit reporting agency client**
- Improved time series forecasts by **30%** using GCP's TensorFlow, AutoML, fine-tuned hyperparameters, conducted A/B testing

Senior Software Engineer Analyst   **Mar 2019 - Dec 2020**
Accenture | Tech Stack: ETL, Restful APIs, AWS (EMR, S3, Redshift), PySpark, SQL, Stored Procedures, UI/UX   *Mumbai, India*
- Created Scala data pipeline for **pharma client,** ingested **streaming** data into Cassandra NoSQL database, totaling **500+** hours
- Generated over **100** PySpark reports via queries, views, stored procedures, window functions to extract 7**5%** critical insights
- Configured **ETL Pipeline** using **AWS** EMR, S3, Redshift resulting **30% reduction** in data processing time in **3 weeks**
- Mentored **5** junior team members in sharing coding practices, KTs in **system architecture**, code reviews, problem-solving

System Engineer   **Mar 2017 - Mar 2019**
Infosys | Tech Stack: Java (SpringBoot), React, JavaScript, Automation Testing, NoSQL, GraphQL   *Pune, India*
- Enhanced **banking client**, **agile** project by building Spring Boot microservices, React UI using MongoDB and GraphQL
- Prepared testing framework with **30+ distinct cases** using Java, Selenium, TestNG, Karate, RestAssured for CI/CD pipeline

Academic Projects

Netflix Movie Recommendation Engine | *Machine Learning* | *Data Science* | Recommendation Engine   **Jan 2024**
- Transformed Netflix Prize Data from .txt files, scrapped genre from IMDB, included movie descriptions using ChatGPT 3.5 API, consolidated all into CSV, applied **Collaborative Filtering** ML model for personalized recommendations

Aircast - Air Quality Forecasting | *LSTM Forecasting* | *Data Pipeline* |*Data Augmentation* | *MVP Architecture*   **Apr 2023**
- Orchestrated **DaaS** with S3, **RDS**, informed Air Quality forecasts via **Hugging Face** LSTM model with RMSE 0.8 in **EC2**
- Integrated **MLaaS**, Streamlit, FastAPI, Airflow, Nginx, Kubernetes, Postgres, CloudWatch, Great Expectations data quality tool

Achievements

Promotion to **Senior** SDE at Accenture	**Mar 2020**
Pat on the Back Award at UST-Xpanxion	**Jan 2022**

Certifications and Student Clubs

MS Azure Fundamentals AZ-900	**Feb 2020**
IBM Coursera Data Science Capstone SpaceX Project	**Jan 2024**
Solidity and Web3 Framework '**Technical Expert**' at NEU Blockchain Student Organization	**Dec 2023 – Present**