

ARIF SARFARAZ WAGHBKRIWALA

Massachusetts, USA | +1 (617) 417 3729

arifwaghbakriwala97@gmail.com | [LinkedIn](#) | [Github/Portfolio](#) | [Medium](#)

EDUCATION

Northeastern University, Boston

Masters of Science (MS) in Data Science (3.7)

Graduation Date: Dec 2023

Coursework: Algorithms, Database Management Processing, Unsupervised/Supervised ML & Data Mining, NLP, Deep Learning & AI, Statistics

Gujarat Technological University, Ahmedabad

Bachelors of Engineering (BE) in Computer Engineering (8.77)

Graduation Date: Aug 2020

Coursework: Differentials, Calculus & Linear Algebra, Programming (C, C++, Java, Python), Data Mining & BI, Big-Data, DataStructures, MySQL

SKILLS

Programming Skills: R, C, C#, C++, Java, SQL, MatLab, Python [Certified](#)

ML/Web Scraping/GUI: Sklearn, PyTorch, Tensorflow/Keras, Darts, ADTK, sktime, Beautifulsoup, Scrapy, requests, Flask, Tkinter, Streamlit

Data Manipulation/Visualization: Pandas, Dask, Polars, Numpy, celluloid, Manim, Matplotlib, Seaborn, Plotly, Bokeh, PIL

Database/Frameworks: MySQL, NoSQL, MongoDB, JSON, XML, Visual Studio IDE, Power BI, Cloud Computing

Cloud Services/Devops: AWS, Azure (ML Studio, DataFactory, DataLake, DataBricks), GCP, Github, Docker, Unix/Linux, Huggingface

EXPERIENCE

Northeastern University | Boston, MA – Graduate TA

Jan 2023 – Dec 2023

- Graduate Teaching Assistant at Northeastern University for CS 7150: *Deep Learning* and DS 4400: *Machine Learning and Data Mining*
- Restructured CS 7150 to emphasize practicality of *Neural Networks*, including *Transformers* and developed *Diffusion* homework from scratch. Assisted in delivering DS 4400, integrating theory with practicum and simplified complex concepts through office hours
- Utilized **Python**, **PyTorch**, Keras, **Scikit-learn**, **Pandas** and **Numpy** to implement course materials and assignments. Orchestrated panel discussions on Popular research in AI to foster 90% increase in student engagement, improving proficiency and understanding of ML

Puma North America, Inc | Somerville, MA – Data Scientist

July 2022 – Dec 2022

Time Series Forecasting System

- Built Forecasting System utilizing tools like **Python**, *Azure ML Studio*, *Azure DataFactory*, Azure Data Lake, SSIS and *Time Series* models (XGBoost, ARIMA, Random Forests, RNNs, FbProphet), to predict demands, resulting in ~10% Cancellations, increasing revenue by ~\$1M+
- Developed SQL Server Integration Services (SSIS) Package using Azure Feature Pack in Visual Studio to migrate data from SQL Server to Azure Cloud and Orchestrated Pipeline in Azure Data Factory to facilitate Data Wrangling, enabling ETL operations to merge files in Azure
- Applied rolling statistics using **Polars** & **Numpy**, Formed Ensemble of Models from **PyTorch** achieving 80% accuracy with 95% confidence

Anomaly Detection Application

- Developed "Anomaly Detection Application" using **Docker**, **Python** and leveraged *Machine Learning*, *Time Series Algorithms* to identify anomalies in sales data, simplifying month-end analytical reporting process and reducing the need to manually flag outliers by 90%
- Implemented automated Excel refreshes for Data Extraction and advanced *statistical transformations* using libraries like 'openpyxl', **Pandas** and **Numpy**. Trained models for every statistic and achieved 60% ensemble accuracy, displaying flagged outliers on Tkinter GUI
- Evaluated KPIs with 95% confidence and conducted *hypothesis testing* (paired t-test for effectiveness of Ensemble Method performance)

Sarvathi Solutions Private Limited | Surat, GUJ – Python Developer

Aug 2020 – Aug 2021

- Designed & implemented ETL (Extract Transform Load) processes in **Python** & utilized requests, **bs4** & scrapy for sourcing data, enhancing data accessibility & streamlining workflows, resulting in improved data integrity and reduced manual interventions
- Implemented *CI/CD Pipelines* & automated configurations with Version Control Systems (Github) via webhooks, analyzed existing processes and identified bottlenecks to optimize software delivery, drastically reducing deployment time by 50% for faster delivery
- Wrote comprehensive unit test scripts in **Python** for all modifications, leveraging **unittest/PyTest** module to ensure robust code quality

PROJECTS

Commodities Trading Web Application [Link](#)

Jan 2024 – Present

- Deployed Streamlit **Python** Web App on **Huggingface** with **Docker**, allowing users to fetch US Imports/Exports data with **RestAPIs** built on the fly; Visualize Sankey, Donut & **Time Series** charts; **Forecast** near-future valuations of particular commodities to make trade decisions, using **ARIMA**, **SARIMA** & **XGBoost** (*Walk-Forward Validation*) & adjust Forecasting Horizon; Utilized **Pandas**, **Numpy**, **sklearn**, etc

Sales Analysis [Link](#)

Jan 2022 – May 2022

- Conducted sales EDA using **Python**, **pandas**, **matplotlib** and **Dask**, performed **ETL** & *data modeling* on 180k+ rows to uncover buyer behavior patterns. Feature engineered *time series* variables, extracted *frequent item-sets* and generated intuitive visuals, formulating *recommendations* hypothesized to potentially increase total revenue by 26%

Diabetic Classification [Link](#)

Jan 2022 – May 2022

- Built Non-Invasive Diabetic *Classification tool* using **Docker** and **Python**, conducted *data wrangling*, preparation and geospatial analysis with **Pandas** & **Numpy** on 1.8M anthropometric records. Optimized hyper-parameters with *Grid Search CV* to implement and *fine tune* custom *Classification Algorithms*. Also, addressed class imbalance issues with *SMOTE*, boosting accuracy from **13%** to an impressive **72%**

Visual Grounding [Link](#)

Jan 2023 – April 2023

- Implemented *Visual Grounding*, built an **ELT** pipeline to train *Transformer Neural Networks*, LLMs using **Python**, **PyTorch** and *Transfer Learning*, predicting **75%** IoU Bounding-Box with **58%** Accuracy, resulted in effective multi-modal processing for applications in image/text