Haider Zainuddin Ali

Hali4@mail.yu.edu | (551) 344-4769

https://alihaider20.github.io/ | LinkedIn/Hali20 | Github/AliHaider20

EDUCATION

Katz School of Science and Health at Yeshiva University — May 2024
Master of Science in Artificial Intelligence – 3.3 GPA — New York, New York
- Relevant coursework: Numerical Analysis, Deep Learning, Predictive Modeling, Data Acquisition, and Management

University of Mumbai — April 2021
Bachelor of Science in Computer Science – 3.64 GPA — Mumbai, India
- Relevant coursework: C, Python, SQL, Statistics, Probability, AI, Discrete Mathematics, Linear Algebra

PROFESSIONAL EXPERIENCE

Python Developer — January 2022 – July 2022
Kan Innovations, A healthcare company focused on human biomechanics | ***Computer Vision*** — Mumbai, India
- Scraped, captured feet images and built 3000 images dataset. Applied Image processing and **Pix2Pix** to segment out feet from the images
- Predicted patient's feet deformities, such as flat feet and claw toes, using foot images scanned from a Plantar pressure mat
- Constructed **3D voxel** meshes of feet to help design insoles which resulted in reducing the patient's feet deformity
- Used **Mediapipe's BlazePose** model and sockets to control a game character using the shoulder joints of a person

Data Science and Business Analytics Intern — November 2020 – December 2020
The Sparks Foundation, A skill-building and mentorship platform | ***Data Analysis*** | ***Modeling*** — Pune, India
- Performed cluster analysis on the Iris dataset, reduced dimensions using PCA
- Classified Iris flowers into three categories using decision trees and **K-means**. Visualized the **Decision Tree** using Scikit-Learn. Found petal length and petal width to be the highest-level decision nodes
- Presentation available on **Youtube**

Deep learning Intern — July 2020 – August 2020
Indian Servers, A skill-building platform |***PyTorch*** | ***Image processing*** | ***OpenCV***| ***Deep learning*** — Mumbai, India
- Led a group of 3 members to identify the five levels of severity of diabetes in eyes using the **Diabetic Retinopathy** dataset on Kaggle
- Guided and worked with them on the latest architectures, applied different **augmentations** techniques, and read research papers
- Built a custom CNN model with an accuracy of **73% and EfficientNet**, which boosted the accuracy up to **85%**

RELEVANT PROJECTS

Thyroid Detection and Analysis | ***Data Analysis*** | ***Data Visualization*** | ***Pandas*** | ***Modeling*** — December 2022
- Dropped features with more than 90% missing values. Replaced NaN values using the correlated features method
- Found females and pregnant females have the highest chances of having a Thyroid disease
- Found most people aged between 60 and 75 have Hypothyroid and between 30 and 45 have Hyperthyroid
- **Linear Regression** with a 0.1 **threshold** performed better than the **K-means** and **logistic regression** algorithms

Income vs. Crime in NYC | ***Data Aggregation*** | ***Data Analysis*** | ***Hypothesis testing*** | **Modeling** — November 2022
- Scraped, aggregated, and analyzed 20000 data points of Crime and Income datasets using **API**, **SQL programming**, and **Pandas**
- Larceny and Harassment level 2 crimes are frequent and successful, and robbery and burglary are often caught red-handed
- Combined both datasets and found that crime is highly correlated with income

Automobile Customer Lifetime Value prediction – GreyAtom | ***Data Analysis*** | ***Presentation*** — October 2021
- Discovered car ownership patterns based on the **geolocation** features of customers
- **Segmented** customers based on their activeness, number of visits to the garage, service type, geolocation, and revenue generated.
- Utilized these insights to understand customer **behavior** and target marketing campaigns and rewards
- Delivered insights to the company, which led to a **5% increase** in the customer base

Protein Classification – Jovian | ***PyTorch*** | ***Augmentation*** | ***Transfer learning*** | ***Regularization*** — June 2021
- Identified **multiple** proteins present in a single-cell image using the Human Protein Atlas dataset from Kaggle
- Achieved **80% accuracy** (using F-score as an accuracy metric and **Resnet18** as a backbone model) in classifying ten types of proteins present in single-cell images

SKILLS & CERTIFICATIONS

- **Technologies:** Python, R, AWS, Neo4J, PostgreSQL, Git, Pytorch, OpenCV, NLTK, ONNX, Pandas, Seaborn, Matplotlib, Numpy, Scipy, C
- **Skills:** Data and Regression analysis, Statistical Modeling, Hypothesis testing, Predictive Modeling, Pattern Recognition, Visualization, RDBMS, Machine Learning, Decision Trees, GBM's, Classification, Cluster analysis, Image processing, Segmentation, Computer Vision
- **Competitions:** Won 2 Python coding competitions (400+ participants) and a data science competition (50+ participants)
- **Certifications:** Data Science Master's program (GreyAtom), Zero to GANs (Jovian), Statistical Inference by Johns Hopkins University