Selim Shaalan

547 Riverside Drive, New York, NY 10027| | selim-shaalan| Github: SelimUCSD
(619)-918-9650 | ss6952@columbia.edu

EDUCATION

Columbia University New York, NY
M.S. in Data Science Dec 2024
University of California San Diego San Diego, CA
B.S. in Math-Economics, Minor in Data Science, GPA: 3.92/4 Jun 2023

- Relevant Courses: Practice of Data Science, Supervised Machine Learning, Data Structures and Algorithms
- Honors: Provost honors x10, Cum laude

WORK EXPERIENCE

University of California San Diego San Diego, CA
Data Science tutor Mar 2022 - Jun 2023

- Prepared worksheets and organized discussion sessions with upwards of 50 students weekly
- Hosted Office hours for students to engage with new material and provided assistance with assignments, projects, and exams for 20+ students weekly
- Engaged in weekly staff meetings to provide updates on students' progress and understanding of material, and input course direction
- Collaborated with Professor to establish practice website with archived exam materials from past 5 years for students
- Developed midterm and final project from scratch by segmenting publicly available datasets and created analysis questions using Python

UnitedHealthcare San Diego, CA
Data analysis Intern Jun 2022 - Aug 2022

- Produced analytic reports on massive call center datasets (1,000,000+ entries) using Python
- Cleaned and cross referenced massive datasets with existing databases using Python leveraged RegEx querying
- Performed sentiment analysis on transcribed recorded phone calls utilizing UHC's systems and produced reports weekly

Banque Misr Cairo, EG
Data Science Research Intern Jun 2021 - Aug 2021

- Built exchange rate forecast model using Autoregressive Integrated Moving Average (ARIMA) in Python
- Researched feasibility of implementing dedicated data science division within Banque Misr
- Arranged frequent meetings with research head and CFO regarding infrastructure and skills needed for implementation of division

PROJECTS

"Predicting the Usage of Urban Bike Share Systems from Date" Jan 2023 - Mar 2023

- Cleaned and Combined Capital Bikeshare's dataset with 2 meteorological datasets regarding weather conditions and humidity
- Trained variety of regression models on dataset (Linear, KNN, Random Forest and SVM) in Python
- Executed extensive Hyperparameter tuning as well as 5-fold cross validation for each model, and examined various error metrics
- Created various visualizations using Matplotlib to plot performance of error metrics and models against each other

"Soccer Player Performance versus Market Value" Jan 2022 - Mar 2022

- Carried out data cleaning and Exploratory Data Analysis on publicly available soccer datasets for thousands of players
- Completed permutation feature importance with Python to determine relevant performance metrics to use in analysis
- Engineered linear regression model with various hyperparameters and segmented analysis for 5 different percentiles

SKILLS

Programming Languages: Python, R, Java, SQL
Spoken Languages: Arabic, German