Huy Ho

huyho@berkeley.edu | huy-ho.github.io | linked.com/in/huydhho | github.com/huy-ho

EDUCATION

University of California, Berkeley | Berkeley, CA — Expected Graduation: Dec 2022
B.A. - Data Science and Statistics — Cumulative GPA: 3.0

San Diego Miramar College | San Diego, CA — Aug 2018 - May 2020
IGETC Transfer Certificate - Statistics — Cumulative GPA: 3.87

Coursework: Principles and Techniques of Data Science, Data Structures, Algorithms, Databases, Multi-Variable Calculus, Linear Algebra, Statistics, Probability, Macro/Micro Economics, Econometrics, Advanced Businesses Analytics, Web Design Decal

SKILLS & TECHNOLOGICAL TOOLS

Languages: Python, Java, SQL/SQLite, R, HTML, CSS, JavaScript
Technologies: Numpy, Pandas, Tensorflow, Keras, Scikit-Learn, Matplotlib, Seaborn, Geopandas, Flask, SciPy
Others: Jupyter Notebook, Anaconda, Git, MS Office, G Suite, Adobe Suite, LaTeX, Regex, DE-TR

EXPERIENCES

Research Intern| United States Army Research Laboratory — Sept 2021- Present

- Research in Machine Learning, Image Processing, and Computer Vision (DE-TR)

Data Engineer Intern | Nozomi Networks — June 2021 - Aug 2021

- Researched and experimented with supervised and unsupervised machine learning techniques for time-series data.
- Used Keras and Tensorflow to implement neural networks models and an ensemble method for anomaly detection.
- The model was put into production and introduced to the company's pipeline.

Curriculum Development Team & Logistics Lead | DataGood @ Berkeley — Jan 2021 - Present

- Created workshops for club technical meetings and manage club's logistics.
- Teach fundamental CS/DS skills and techniques such as Git, SQL, and relevant libraries.

Finance Intern | UC Berkeley Vietnamese Student Association — Aug 2020 - Dec 2020

- Created info-graphics to promote the club and upcoming events.
- Along with another intern, hosted a fundraiser that raised over $2000.

PROJECTS

Stock Market Forecasting Analysis| *Python, Pandas, Keras, Tensorflow, StatsModel*

- Used Yahoo Finance time-series data to investigate major stocks (AMD, NVDA, INTC). I applied machine learning techniques for forecasting, clustering, and analysis. I implemented LSTM, CNN, XGBoost, ARIMA and other models to test different approaches towards time-series forecasting.

Washington DC and Chicago Crime Analysis| *Python, Pandas*

- I analyzed data using Pandas, regression, geospatial analysis and produced data visualization. The final deliverable was an analysis of crimes in Chicago and DC to explore trends across factors such as neighborhoods and time that included GeoPandas maps.

Spam and Ham Email Classifier| *Python, Pandas, Scikit-learn*

- I Created a data pipeline to process the data and built a model to predict whether an email was spam or ham with a 85% training accuracy on the test set. The following methods were used: feature selection, one-hot encoding, NLP, and logistic regression.

Lahman Baseball Database| *SQL, SQLite3*

- Wrote SQL queries to extract information from the Lahman's Baseball Database. Investigated trends ranging from saber-metrics, school-attended, salaries, and others.

Spotify Automation| *Python*

- Using the Spotify API, I created a program that automate and save a new playlist for myself weekly. Using this data, I then perform data analysis on these playlists.