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

Coursework: Principles and Techniques of Data Science, Data Structures, Algorithms, Databases, Intro to Artificial Intelligence, Intro to Machine Learning, Data Inferences & Decisions, Calculus, Linear Algebra, Statistics, Probability, Macro/Micro Economics, Econometrics, Businesses Analytics, Web Design

SKILLS & TECHNOLOGICAL TOOLS

Languages: Python, Java, SQL, 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 Sep 2021- Present
- Conducted research in Machine Learning and Computer Vision (DE-TR).
- Investigated about the ethical implications of facial recognition.

Data Science 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 to classify anomaly.
- The model was put into production and introduced to the company's pipeline.

External Project Manager & Logistics Lead | DataGood @ Berkeley Jan 2021 - Present
- Lead a team of 6 students to help external organizations find insights and solutions using computer science, data science, and machine learning techniques.
- Serve as a point of contact to outside organizations and manage club's logistics.
- Teach fundamental CS/DS skills and techniques such as Git, SQL, and relevant libraries.

PROJECTS

Stock Market Forecasting Analysis| *Python, Pandas, Keras, Tensorflow, StatsModel*
- Used Yahoo Finance API to investigate major stocks. 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 the 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 an 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*
- Wrote SQL queries to extract information from the Lahman's Baseball Database. Investigated trends ranging from saber-metrics, school-attended, salaries, and others.

Twitter Bot| *Python*
- Using the Twitter API, I created a bot autonomously tweets and retweets. The bot posts daily optimistic quotes to Twitter and extract the stock market news. I then performed analysis on these tweets and build a classifier using NLP and regression.