Huy Ho

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

EDUCATION

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

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

SKILLS & TECHNOLOGICAL TOOLS

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

EXPERIENCES

Data Science Intern – *IDEXX Laboratories* May 2022 - Present
- Clinical Informatics Data Science Team

Undergraduate Research Intern - *UC Berkeley* Sep 2021 - May 2022
- Conducted research in Machine Learning and Computer Vision (DE-TR).
- Investigated the ethical implications of facial recognition.

Data Science Intern - *Nozomi Networks* May 2021 - Aug 2021
- Researched and experimented with statistical, supervised, and unsupervised machine learning techniques for time-series data using the statsmodel and scikit-learn libraries.
- Developed neural network models to forecast time-series data using Keras and Tensorflow and an ensemble classifier to classify anomalies with a 95% accuracy.
- The model was put into production and introduced to the company's pipeline.

Project Manager and Logistics Lead - *DataGood @ Berkeley* Jan 2021 - Present
- Led a team of 8 students to help external organizations find insights and solutions using computer science, data science, and machine learning techniques.
- Served as a point of communication to organizations and managed the club's logistics.
- Mentored intending data and computer science major students and taught fundamental skills such as Git, SQL, Tableau, and relevant libraries.

PROJECTS

Stock Market Forecasting Analysis - *Python, Pandas, Keras, Tensorflow, StatsModels*
- Used Yahoo Finance API to investigate and analyzed stock data such as P/E ratio, dividends, etc.
- Applied machine learning techniques for forecasting, clustering, and analysis. Implemented LSTM, CNN, XGBoost, and ARIMA models to test different approaches toward time-series forecasting and used them to predict other stocks.

Spam and Ham Email Classifier - *Python, Pandas, Scikit-learn*
- Performed data cleaning, created a data pipeline to process the information, and built a model to predict whether an email was spam or not with a 90% training accuracy on the test set.
- The following methods were used: feature selection, one-hot encoding, decision trees, natural language processing, and logistic regression.
- The final deliverable was a finalized model and visualizations using Tableau.

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 produced visualizations.

Simple Git - *Java*
- Implemented a clone version of Git with command line commands such as init, add, commit, etc.
- Designed the internal structures using serialization, object-oriented programming, and data structures.