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

EDUCATION

- **University of California, Berkeley** *Expected Graduation: December 2022*
  Bachelor of Arts: Data Science and Statistics
  Relevant Coursework:
  Principles and Techniques of Data Science, Data Structures, Algorithms, Databases, Artificial Intelligence, Machine Learning, Data Engineering, Data Inferences and Decisions, Statistics, Probability, Econometrics, Macro/Microeconomics, Business Analytics, Linear Algebra, Web Design

SKILLS AND TECHNOLOGICAL TOOLS

- Languages: Python, Java, R, SQL, HTML, CSS, JavaScript
- Libraries: NumPy, Pandas, Tensorflow, Keras, Scikit-Learn, SparklyR, Flask, SciPy, Matplotlib
- Others: Apache Spark, AWS, H2O AI, Hadoop, Anaconda, Tableau, Git, Regex, Jenkins

EXPERIENCES

- **DataGood at Berkeley** – *Project Manager, Logistics Lead, and Marketing Lead* *Sep 2020 – Present*
  - Led a team of 8 students to help external organizations to find insights and solutions using computer science, data science, and machine learning techniques.
  - Mentored freshmen, sophomore students and taught fundamental computer and data science skills such as Git, SQL, Pandas, and relevant libraries.
  - Served as a point of contact, managed the club's logistics, and created infographics for marketing purposes.
- **IDEXX Laboratories** – *Data Science Intern* *May 2022 – Aug 2022*
  - Investigated a potential data leakage issue with the company's machine learning pipeline using Spark and SQL.
  - Built a classifier with an accuracy of 92% to classify transaction items using feature engineering, natural language processing, H2O AI, and deployed it into the company's latest ontology version.
  - Wrote R and Spark scripts to help wrangle data, automate processes, and improve general workflows.
- **UC Berkeley Lab** – *Research Intern* *Sep 2021 – May 2022*
  - Conducted research in machine learning (DE-TR) in collaboration with US Army Research Laboratory.
  - Researched the ethical implications of facial recognition and created weekly reports.
- **Nozomi Networks** – *Data Engineer Intern* *May 2021 – Aug 2021*
  - Researched and experimented with statistical, supervised, and unsupervised machine learning techniques for time-series data using statsmodels and scikit-learn libraries.
  - Developed neural network models to forecast time-series data using Keras, TensorFlow, and an ensemble classifier to detect anomalies with a 95% accuracy.

PROJECTS

- **Stock Market Forecasting Analysis** – *Python, Pandas, Keras, TensorFlow, Statsmodels*
  - Used Yahoo Finance API to analyzed stock data such as P/E ratio, dividends, market capitalization, etc.
  - Implemented custom LSTM, CNN, 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 and built a model to predict whether an email was a spam or not with a 90% accuracy.
  - The following methods were used: feature selection, one-hot encoding, decision trees, and regressions.
- **Git Clone** – *Java*
  - Implemented a clone version of Git with command line features such as init, add, commit, etc.
  - Designed the internal structures using serialization, object-oriented programming, and various data structures.