Yuki Kuwahara

(949) 449-3206 • ykuwahar@berkeley.edu • https://www.linkedin.com/in/yuki-kuwahara/ • Irvine, CA

EDUCATION

University of California, Berkeley *Expected May 2024*
Bachelor of Arts, Data Science GPA: 3.85

Coursework: Foundations of Data Science, Computational Structures in Data Science, Principles and Techniques of Data Science, Data Engineering, Introduction to Machine Learning and Data Analytics (In Progress), Natural Language Processing (In Progress), Data Structures and Algorithms, Concepts of Probability

EXPERIENCE

SwingVision *June 2023 - Present*
Data Labeling and Analysis Intern *Berkeley, CA*

- Utilized various Python libraries such as Numpy, Pandas, Matplotlib, and Seaborn to conduct Exploratory Data Analysis on multiple datasets of user data and wrote functions which created visualizations used to assist Data Scientists' model performance
- Created a visualization interface using Seaborn and Plotly involving the performance momentum of a tennis match to provide users information about their match performances
- Used CVAT Labeling Tool to label pickleball and tennis user videos for data which is used for model training
- Updated Data Labeling Guidelines with various edge cases, and maintained a labeling accuracy of greater than 95%

JOYA Scholars *January 2021 – June 2022*
Mathematics Tutor *Fullerton, CA*

- Tutored 5 High School students between 9-12th grade from low-income communities in the Fullerton Area in the topics of Mathematics, ranging from Algebra 1 to Pre-Calculus
- Made specific study guides and problem sets for each student to help them prepare for upcoming exams and quizzes, significantly improving their scores and grades

PROJECTS

Spam or Ham Email Classifier *April 2023*

- Utilized a Dataset consisting of 9,000 unique emails and conducted Exploratory Data Analysis using Numpy, Pandas, and Seaborn to see which features, such as word count, punctuation, and unique words would best classify if an email were spam or ham
- Created multiple visualizations such as line plots and ROC Curves which would aid and help identify which features and models would be the most ideal
- Trained, validated, and tested a Logistic Regression Classifier that classified the emails properly with 92% accuracy

Predicting Housing Prices in Cook County *March 2023*

- Utilized a dataset consisting of > 500,000 unique houses in Cook County Illinois and used Python libraries such as Seaborn, NumPy, and Sci-kit Learn in creating boxplots, joint plots, and feature engineering
- Trained, validated, and tested Multiple Linear Regression Models that would best predict the prices of houses, determined by a specific threshold of RMSE
- Evaluated if Cook County's Price Assessment on houses were done fairly and inspected potential confounding factors, such as race and income

SKILLS

Programming Languages: Python, R, SQL (PostgreSQL, MySQL), Java
Machine Learning: Classification, Regression, Clustering, Feature Engineering, Decision Trees, PCA, One-Hot Encoding
Technical Frameworks: Numpy, Pandas, Matplotlib, Seaborn, Plotly, PyTorch, Tableau, Sci-Kit Learn, MongoDB, HTML/CSS
Languages: Bilingual Fluency in English and Japanese