

Roy Ho

San Francisco, CA | 415-741-8955 | royho346@gmail.com | [linkedin.com/in/royho1/](https://www.linkedin.com/in/royho1/) | github.com/royho1/

EDUCATION

UNIVERSITY OF CALIFORNIA, DAVIS

Bachelor of Science

Sep 2022 - March 2026

Major in Statistical Data Science and Minor in Computer Science

Relevant Coursework: Statistical Learning, Regression, Probability Theory, Advanced Computing, Algorithm Design & Analysis, Data Structures

Certifications: NVIDIA – Fundamentals of Deep Learning

October 2025

EXTRACURRICULARS

TechSprint Innovators

Head of Data Engineering

March 2024 – September 2025

- Built a multi-factor stock screening model using fundamental, technical, and NLP-based sentiment features (FinBERT).
- Developed and evaluated a supervised classification model in scikit-learn to predict price appreciation, performing feature selection, model tuning, and performance validation on historical market data.
- Engineered and automated a daily ETL (Extract, Transform, Load) data pipeline (Python, yfinance, Alpaca API) running on a Raspberry Pi to filter equities, generate structured CSV outputs, and deliver real-time investment signals via Discord webhook.

AI Student Collective (AISC)

General Member

September 2024 – April 2025

- Built machine learning models for stroke risk prediction and real-time drowsy driver detection within structured project cycles.
- Performed data preprocessing, feature engineering, model development, evaluation, and project presentations.
- Worked within quarter-long sprint cycles with defined milestones, code reviews, and final project demos, following structured machine learning development workflows from ideation to deployment.

EXPERIENCE

Journal of Artificial Intelligence and Knowledge Engineering (JAIKE)

Artificial Intelligence Researcher

January 2025 – Present

- Conducted research on LLM-based automation, focusing on API-driven system design, agent orchestration frameworks, and productivity applications across research, coding, and enterprise workflows.
- Authored a 25-page research paper on LLM-based automation and agent architectures, synthesizing peer-reviewed and industry research on API integration, architectural design patterns, system limitations, and responsible deployment; submitted for journal publication.
- Served as a peer reviewer for JAIKE, evaluating research on retrieval methods in large language models, reasoning performance in extended tasks, and large-scale model architectures for methodological rigor and evaluation quality.

SIGNIFICANT PROJECTS

Job Market Analytics Dashboard

September 2025 – December 2025

- Built an end-to-end job market dashboard as part of a team, enabling students to analyze in-demand skills, salary trends, and resume alignment across 9,000+ statistics-related job postings in CA, NY, and TX.
- Processed and cleaned large-scale scraped datasets and applied NLP (spaCy) to extract and standardize technical skills from job descriptions.
- Designed a resume-based job recommendation model using TF-IDF (Term Frequency-Inverse Document Frequency), improving model accuracy from 30% to 60%, and deployed the solution as an interactive dashboard to present insights to faculty and peers.

Drowsy Driver Detection System

January 2025 – March 2025

- Developed a real-time drowsy driver detection system, powered by a Convolutional Neural Network (CNN) trained on the MRL Drowsiness Dataset (4,000+ images) to improve road safety.
- Led model development and preprocessing, implementing image augmentation, normalization, and CNN training in TensorFlow with hyperparameter tuning, improving validation accuracy from 76% to 89% and reducing false positive detections.
- Integrated the trained model into a live webcam system using OpenCV and Dlib facial landmark detection to compute Eye Aspect Ratio (EAR), delivering a fully functional alert system; project awarded Best Execution for the AISC Winter Cycle.

Heart Stroke Risk Prediction

September 2024 – October 2024

- Built a supervised machine learning pipeline predicting stroke risk, designed to allow individuals to input health and lifestyle factors (e.g., age, BMI, lifestyle indicators) and assess heart stroke risk.
- Led data preprocessing and feature engineering, constructing predictors to improve model performance.
- Implemented cross-validation and model comparison (Logistic Regression, Random Forest), selecting the final model using ROC-AUC and F1 metrics.

SKILLS

Programming: Python, R, SQL, JavaScript, HTML, CSS

Data Analysis & Machine Learning: Excel (advanced formulas, pivot tables, data visualization), Regression (Lasso, Ridge, Logistic), Random Forest, Gradient Boosting, CNNs, Feature Engineering, Cross-Validation & Hyperparameter Tuning, Model Evaluation (ROC-AUC, Precision/Recall), TF-IDF, NLP (spaCy)

Libraries & Frameworks: Scikit-Learn, Pandas, PyTorch, NumPy, BeautifulSoup, Flask, SQLAlchemy, Selenium, PySpark, Matplotlib