Ryan Yee

408-896-7161 | aryyee88@gmail.com | LinkedIn/aryyee88/ | github.com/koitaku2323

Education

University of California, Santa Barbara — Santa Barbara, CA
Bachelor of Science, Statistics and Data Science — *June 2024*

Sponsored Projects

CalCOFI AI/ML Products: Whale Call Detection | *Scripps Institution of Oceanography* — Jan - June 2024

- Collaborated with ML PhDs and Marine Scientists to rapidly prototype data preprocessing pipelines, transforming 15 years of acoustic/sonobuoy data into spectrograms for real-time marine mammal detection to assist oceanic research using AI.
- Implemented a comprehensive image processing pipeline that addressed challenges in noise reduction through Gaussian filter, bilateral filter, median filter, morphology, and column-wise matrix subtraction.
- Conducted Principle Component Analysis (PCA) on spectrograms, resulting in a 45% dimensional reduction while retaining 90% of the total variance. Reduced R-CNN ResNet50 model training time from 10 to 5 minutes per 15 epochs using HPC.
- Removed 15 types of vertical artifacts and background noise while preserving essential signals, improving model performance by 30% and data usability by 50%.

CalCOFI Data Products: Marine Mammal Visualization Tool | *Scripps Institution of Oceanography* — Jan - June 2024

- Developed and deployed an *R Shiny App* on the official CalCOFI website to provide scientists and non-technical end users with an interactive tool to visualize and query 20 years of marine mammal data collected onboard CalCOFI.
- Integrated multiple data streams using SQL and Python, highlighting over 1,000,000 rows of marine mammal visual sightings, eDNA detection, and acoustic detection data through time and space.
- Presented innovative AI/ML solutions and collaborative initiatives through exhibit booths, fostering sustainable ocean observing practices at the *Ocean Observing in California joint conference* with over 500 attendees.
- Winner of the Best Poster Presentation Award at the *Ocean Observing in California Joint Conference* hosted by CalCOFI, SCCOOS, and CeNCOOS. Competed against 70 teams from academia and industry backgrounds across the nation.

Professional Experience

Business Intelligence Analyst Intern — June - September 2023
Huntington Learning Center — *Cupertino, CA*

- Conducted regression analysis and financial modeling on sales data using R and Excel, optimizing profit margin by 15%.
- Worked with senior management to define financial and operational KPIs and strategize areas for improvement.
- Compiled 20+ reports in Tableau and Excel to provide key stakeholders with actionable information for decision-making.
- Collaborated with marketing teams to track search engine results page (SERP) and conduct search engine optimization (SEO) using Google Lighthouse, increased organic web search traffic by 200%, inquiry rate by 50%, and sales by 30%.

Research Assistant - Data Storage Solutions — June - September 2022
Zerras Inc. — *Santa Clara, CA*

- Analyzed existing data storage systems, architectures, and protocols to identify areas for improvement and innovation.
- Co-authored 10+ blogs and newsletters on optical storage archives, digital preservation, and cold data storage solutions.
- Conduct market research on the supply & demand of data storage products based on industry, region, and storage medium.

Academic Projects

Driving Behavior Classification | *R, dplyr, tidyverse, yardstick, keras, tensorflow* — September - December 2023

- Implemented ML solutions on detecting and classifying driving behaviors using motion sensor data for risk management.
- Fine-tuned machine learning algorithms including Logistic/Multinomial Regression, KNN, Linear/Quadratic Discriminant Analysis, Elastic Net, and Random Forest using regularization/shrinkage methods, grid search, and cross-validation.
- Trained deep learning models such as CNN-LSTM and ResNet using CUDA, performed hyperparameter tuning and statistical analysis using the loss function and precision-recall as a metric.

Time Series Project: U.S. Real GDP Forecast | *R, tsdl, astsa, qpcR, forecast* — March - June 2023

- Created forecasts for U.S. real GDP using quarterly data from 2002 to 2023. Analyzed trends and patterns in U.S. real GDP using time series techniques, including ARIMA and SARIMA, along with data transformation methods.
- Conducted model diagnostics using Box-Pierce, Ljung-Box, Shapiro-Wilk, Yule-Walker, and other statistical tests.

Vehicle Valuation and Automotive Configuration Web App | *Java, Java.net, Java.io* — March - June 2022

- Developed a Java-based web application for vehicle valuation and automotive configuration. Allow clients to request and configure car models through a client-server architecture using the Java.net API, Java Servlets, and Apache Tomcat.

Search Engine for Music Lyrics | *Java, CRUD, JUnit, UML Diagrams* — April - May 2022

- Developed a Search Engine from scratch using Inverted-Index and Hash Tables for CRUD operations on music entries.

Technical Skills

Python (Pandas, NumPy, Scikit-learn, Keras, Tensorflow, PyTorch, spaCy, Matplotlib, Librosa, OpenSoundscape, Pillow)
R, SQL, SAS, JavaScript, Tableau, Shiny, Git, Apache Spark, Kafka, Hadoop, Docker, Kubernetes, Google Cloud Platform