JARED YU

44357 View Point Circle
Fremont, CA 94539 http://www.yuqizheng.com/ www.linkedin.com/in/jaredvyu https://github.com/qzyu999 qzyu@jhu.edu (510) 366-6100

An adaptable statistician with interest and experience in statistics, data science, machine learning, and artificial intelligence.

Statistical Modeling and Analytics	Teamwork and Collaboration	Communication and Presentation
Good Applied Statistical Skills	Testing and Validation	Self-Motivated, Fast Learner, Innovative

EDUCATION

JOHNS HOPKINS UNIVERSITY, Baltimore, MD
Master of Science Degree in Data Science (online)**:** Expected May, 2021

UNIVERSITY OF CALIFORNIA, DAVIS, Davis, CA
Bachelor of Science Degree in Statistics: June, 2019 **GPA:** 3.56 **Dean's List:** One Quarter

CITY COLLEGE OF SAN FRANCISCO, San Francisco, CA
Associate in Science Degree with Honors: May, 2017 **GPA:** 3.36 **Dean's List:** Two Semesters

PROFESSIONAL SKILLS PROJECTS

UNIVERSITY OF CALIFORNIA, DAVIS, Davis, CA — Sep. 2017 – Jun. 2019

Completed eight projects noted below, utilizing relevant real-world datasets for statistical models mastered. Each project utilized RStudio or Python to introduce a dataset, explain the methods, analyze the data, interpret the results, and deliver conclusions in the form of a written report with tables and graphs. Received a grade $\geq$ 90% for each of the following projects:

Country Fertility and PPgdp Analytics; Diabetes Prediction Analytics; Image Recognition: Machine Learning/k-NN/Cross Validation; San Francisco Fire Department Response Time and Priority Level Prediction; Low Birthweight and Heart Disease Prediction Analytics; USA Spending Data Analytics for Business Owners; Time Series Analysis of Annual Temperature Anomalies; Nutrition Analysis

- **Teamwork and Collaboration:** Some projects required group discussions for how to best complete tasks. Others required independent work with later team collaboration.
- **Statistical Modeling and Analytics:** Utilized RStudio and Python where statistical models developed within packages were applied to various datasets. Used models from different packages or manual programming calculations to create various test statistics for later interpretation from the context of model assumptions and data.
- **Models used include:** linear regression, logistic regression, k-Nearest Neighbors, cross validation, Poisson regression, ARIMA, spectral analysis, Hotelling's T^2, Bonferroni's correction, principal component analysis, and linear discriminant analysis.
- **Testing and Validation:** Used transformation on variables to identify the ideal R squared for a dataset. Performed regression analysis and confidence interval testing. Incorporated techniques such as BoxCox to regress a variable on other data within the set. Created training and validation models to develop rigor within testing. Evaluated different criterions such as Akaike information criterion, Bayesian information criterion, and Mallow's Cp. Tested algorithms that performed forward stepwise procedure to obtain predictive models with various combinations of predictors.

Yu Qi Zheng's Blog — Sep. 2017 – Present

A personal website created with Word Press that evolved into a series of blog posts that teach Statistical lessons with connections to popular modern fields such as Data Science and Machine Learning, with >4,000-page views from relevant communities.

Teacher, Writer, Website Designer: Communicated lessons learned through university coursework and connected them with Data Science and Machine Learning concepts.

• **Communication and Presentation:** Analyzed various datasets to clean and rework data for proper modeling and reassessment within the project framework. Created reports for each project where detailed assessment was given for different data factors.

• **Self-motivated, Quick-Learner, Innovative:** Created a niche designed to assist community members with a generally weaker statistics background to further develop their knowledge for use in their related fields.

• **Good Applied Statistical Skills:** Extending on the notion that one learns best when teaching content, sought to re-learn (and sometimes reinterpret) my previous understanding of a concept so that it can be better understood in the form of a lesson for an interested learner. Reviewed / Reapplied past material to a lesson format to better develop personal subject matter understanding. Various articles have been written on topics such as: linear regression, logistic regression, analysis of variance, data science, and multivariate data analysis. Future plans include updating articles and creating new ones to be added to a GitHub page.

VOLUNTEERISM

Computer Science 4 Kids – Volunteer Programming Teacher: Taught Apple's Swift programming language to elementary school children near U.C. Davis. (2018)

HackDavis 2018 – Mentor: Participated as a mentor in U.C. Davis' annual Hackathon; used programming skills and experience to assist others in the development of their projects.

RELEVANT PROJECTS

Iris Dataset Classification (Programming Assignment 2): In the final programming assignment for the course, Algorithms for Data Science, from the Johns Hopkins University, the goal was for students to classify the famous Iris dataset using various algorithms. The programming was done completely in RStudio. A dummy version of the dataset was first cleaned by imputing values into for NA elements of observations one at a time. An additional two features were generated based upon the top two ranked features, where values were first randomized, rotated using a covariance matrix, and finally normalized. Using Wilk's Outlier Removal technique, individual observations were removed one at a time based on the distance of their features from the rest of the dataset. Feature ranking was done by creating class-wise density plots with ggplot. The accuracy of the features was then tested using the Mahalanobis distance to check their classification accuracy. Dimensionality reduction was done using Kernel PCA, where the biplot of the Kernel PCA matrix were shown to efficiently classify the data. Classification was done utilizing the following models: expectation-maximization algorithm (EM), Linear Discriminant Analysis (LDA), Feedforward Neural Network, and Support Vector Machines (SVM). Code for the above algorithms were written manually, based upon lecture notes or textbook information. The classification accuracy for all models were above 95%. The final grade for the project was 100%. (2019)

San Francisco Fire Department Response Time and Priority Level Analysis: Queried data through city government's web API. Scraped data using BeautifulSoup4 to increase number of features. Performed EDA using various plotting tools such as Seaborn and geospatial analysis. Analyzed response variables using Linear Regression, Logistic Regression, and Poisson Regression. Packages for this included statsmodels and scikit-learn. Delivered a 10-minute presentation using PowerPoint which covered the progress of the analysis. (2018)

SPECIALIZED SKILLS

Computer Skills: C++, Java, Python, R, RStudio, PostgreSQL, Tableau, Microsoft Power BI (beginner), Excel (beginner), STATA

Statistical Tools: linear regression, logistic regression, k-Nearest Neighbors, cross validation, Poisson regression, ARIMA, spectral analysis, Hotelling's T^2, Bonferroni's correction, principal component analysis (PCA), Kernel PCA, linear discriminant analysis (LDA), expectation-maximization algorithm (EM), support vector machines (SVM), feedforward neural networks (multilayer perceptron/ANN).