Yunzhe **Li**

Statistics · Data Science

137 Chestnut St. Apt 212, Santa Cruz, CA, 95060

310-795-4718 | yli566@ucsc.edu, madli@ucdavis.edu | Jacksoften

Education

University of California, Santa Cruz — *Santa Cruz, CA*

M.S. in Statistics, Bayesian Statistics — *Sep. 2018 - EXP Jun. 2020*

University of California, Davis — *Davis, CA*

B.S. in Statistics, Computational Statistics — *Sep. 2016 - Jun. 2018*

Work Experience

DSI Davis

Intern — *Mar. 2017 - Aug. 2018*

- Participated in the Development of a R package, ReadPDF. Using text mining and XML pointers to build a dictionary of diagnostic tests in Zoonotic fields. Using the Tesseract OCR, an OCR engine, for wine catalogs and extracting wine prices in R.
- Helping a English group visualize the patterns of their coding data from audios and videos of William Shakespeare with ggplot2 package
- Cleaning attendance information provided by keycard machine, analyzing the pattern of employees attendance corresponding to day of the week

Projects

Algorithm

- Finding the shortest paths between cities in a Map by using Dijkstra's algorithm and Minimum spanning tree algorithm in C++.

Database

- Building a relational database system for a virtual college, based on students' IDs, students' emails, courses, grades, instructors, etc with postgresql.

Statistical Analysis

- Predicting the popularity of an online video game, League of Legends, using ARIMA model and spectral analysis with google trends data.
- Analyzing the happiness of people in country level with their economy, people's trust in government, people's generosity, and people's health and the contribution of their family, using linear and nonlinear regression methods, ANOVA, model selection and PCA.

Machine Learning

- Collecting users' data from the Spotify's API, and classify songs based on users' comment with NLP methods.
- Building a recommender system for Netflex's users in python, using packages, sklearn, scipy, and suprise.
- Recognizing if there are humans in a image using various objects using SVM and CNN algorithms with SVM from sklearn package and CNN from keras package.

Relevant Courses

Statistics

- Probability Theory and Mathematical Statistics, Time Series, Regression and ANOVA, Multivariate Data Analysis , Bayesian Data Analysis, Big Data and High Performance Statistical Computing, Practice in Statistical Data Science

Computer Science

- OOP and Software Development, Data Structure, Scripting Languages (R and Python), Algorithm Design and Analysis, Database System

Skills

Programming

- R, Python, Matlab, C, C++, SQL

Other Techniques

- LaTeX, Microsoft Office, Linux, MacOS, Github