Jing Li

(424)343-4236 • Jingli940731@gmail.com

2215 Rogene Drive, Apt. T3, Baltimore, MD, 21209

Profile

First-year master of science (ScM) candidate in Biostatistics interested in statistical genomics and drug development optimization with undergraduate majors in Statistics and Biochemistry. Over 2 years' experience in R and obtained both SAS advanced and base certificate. Strong analytical backgrounds with experience of Kaggle competition using Random Forest and Boosting. Fluent in Mandarin and conversant in Japanese.

Education

Master of Science (ScM) in Biostatistics, GPA: 4.00/4.00 — September 2016-present
Johns Hopkins University Bloomberg School of Public Health — Baltimore, MD
Relevant Coursework: Epidemiologic Inference in Public Health, Introduction to Clinical Trials, Essentials of Probability and Statistical Inference, Methods in Biostatistics, Computational Genomics: Data Analysis, Analysis of Longitudinal Data

Bachelor of Science in Statistics and Biochemistry, GPA: 3.63/4.00 — September 2012-June 2016
University of California Los Angeles — Los Angeles, CA
Honors: Alpha Lambda Delta and Phi Eta Sigma, Scholar-Athlete Award

Certificate

SAS Advanced Programmer Certificate (AP016723v9) — March 2016
SAS Base Programmer Certificate (95%, Certificate No: BP056272v9) — January 2016
Certificate of Japanese-Language Proficiency Test N3 (N3A116471A) — February 2015

Projects and Activities

Kaggle Competition (Top7) — Los Angeles, CA
Participant — May, 2016
Kaggle is a platform for establishing data science competitions involving data mining, modeling and predicting

- Predicted animals' outcomes (euthanasia, adoption and etc.) for shelters in the Great Los Angeles area (both training and test data sets had over 110000 observations)
- Used data mining (wrote over 500 lines in R) to clean and reshape variables, and combined external data as potential predictors
- Selected predictors using cross validation and conducted multiple prediction models using Random Forest, Boosting, Multinomial Regression, Support Vector Machine, KNN as well as Neutral Networks
- Compared the prediction outcomes of all models based on log loss and established final ensemble classification models using Random Forest and Boosting

Stats C183 Portfolio Project — Los Angeles, CA
Participant — May, 2016
Stats C183 is about statistical models in finance and this project mainly concerned with optimal portfolio construction and portfolio performance

- Constructed portfolio with biotechnology and life insurance stocks
- Used single index model, constant correlation model and multi-group model to find the composition of optimum portfolios

• Analyzed and compared expected returns of each models corresponding to the market (S&P 500) using approaches such as Sharp Ratio, Differential Return as well as Treyor Measure

ASA Datafest Los Angeles, CA

Participant April 2016

ASA Datafest is a highly programming based data competition sponsored by American Statistical Association

• Implemented ticket pricing optimization strategy concerning with locations and events popularity based on data from Ticketmaster

• Quantified the popularity of events and established a classifier for events based on ticket values, number of tickets being sold as well as the popularity of events using R

• Conducted ANOVA test for the inter-city popularity difference and measured the significance ratio of the difference by TukeyHSD test as an index in the final pricing optimization model

Experience

InciteData Chengdu, China

Analyst Intern July 2016-August 2016

InciteData is a technical company focusing on establishing automated big data analysis platform

• Cooperated with the team in conducting vehicle plate recognition algorithm using convolutional neural networks on TensorFlow

• Tested run both softmax regression model and CNN model on Linux to check for precision of prediction

Sellmark Corporation Los Angeles, CA

Business Analyst Intern September 2015-Feburary 2016

Sellmark Corporation is a company carries hunting optics products

• Worked remotely to conduct statistical consulting based on orders and actual sales data using R

• Established a hybrid time series model combining ARIMA and multivariate regression for forecasting sales amount using R and Excel

• Predicted the potentiality of new products using the hybrid time series model

Zhang's Dental Clinic San Jose, CA

Assistant Intern August 2014-September 2014

• Assisted dentists with scheduling appointments, taking X-ray, filling, cleaning and sterilizing instruments

• Managed over 5000 patients' profiles and helped file dental claims to insurance carriers

C-elegans' Research Lab Los Angeles, CA

Student Researcher April 2014-June 2014

C-elegans' research lab is supervised by Prof. Arisaka from UCLA to monitor behaviors of c-elegans under physical stimulation

• Researched relevant scientific literature and monitored C-elegans' behavior based on electrical stimulus

• Maintained C-elegans' growth and culturing by preparing agar as growth media and E.coli as food source

Skills

Computer: R (2 years+ experience)/SAS/Python/Tensorflow/SQL/Stata/SPSS/QGIS/Pymol/Lyx/Excel/Tableau
Language: Fluent in Mandarin and conversant in Japanese