Kuan-Cheng(KC) Lai

Berkeley, CA, 94702 | (510) 646-5769 | kuanchenglai@gmail.com

EDUCATION

University of California, Berkeley, Berkeley, CA, USA 08/2014-12/2015

M.A. in Statistics

National Cheng Kung University (NCKU), Tainan, Taiwan 09/2008-06/2013

B.B.A. in Statistics and B.A. in Economics

INDUSTRY EXPERIENCE

Intern in Aiooki USA, Inc 08/2015-12/2015

- Collected demographic data from United Census Bureau's API
- Processed and inserted data into relational databases using Python and Java Hibernate
- Built statistical classification models and regression models
- Implemented Angular JS to front-end product

ACADEMIC EXPERIENCE

Identify Natural Images from Human Brain Activity 12/2015

- Built Lasso and Ridge regression on 10921 Gabor wavelets features from each natural image to predict fMRI voxels values
- Implemented AIC, AICc, BIC and CV to choose tuning parameter lambda
- Used Gradient Boosting to build a model and compare performance with other regression models

Organ Classification on Drosophila Embryos 11/2015

- Implemented exploratory data analysis and feature engineering to determine useful features
- Used KNN, weighted KNN and Random Forest to build classification models on embryo images to classify gut, yolk and tissue of embryos
- Implemented ROC to verify classification performance which resulted in 87% accuracy rate using KNN

Text Mining on Stack Overflow Posts 03/2015

- Processed and engineered features on more than 25,000 Stack Overflow posts with Python
- Trained multi-label Random Forest model and Boosting Tree in R to classify 5 different tags
- 97% prediction accuracy on validation set and achieved 4th place ranking in Kaggle competition

Text Mining on International News in Twitter 02/2015

- Used Twitter API to extract timeline data from popular news agencies
- Analyzed raw JSON data using Python package NLTK and Pandas
- Using visualization methods, found that agencies tend to have geographical 'preference pockets'

Using Statistical Methods to Help Community Sign Utility Contract 12/2012

- Worked with a team to analyze data of community utility usage data
- Implemented Bootstrap to estimate the variance of average of utility usage for each month and set up the contract
- Resulted in approximately 10% cost savings for the community in the coming years

SKILLS

Computer Skills: R, Python Java, PostgreSQL, Spark, Hadoop, Hibernate, Pandas, Angular JS