XI CHEN

xichen612@gmail.com | +1-530-746-1479 | www.linkedin.com/in/xichen612

SUMMARY OF QUALIFICATIONS

- Self-motivated, creative, and cooperative person, quick learner and excellent team player.
- Deep love and passion on big data and data analytics.
- Demonstrated strong specialized skills and learning ability in machine learning and statistical analysis.
- Expertise in R, Python, Proficient in Unix Shell, Spark, SAS.

EDUCATION

Department of Statistics, **UNIVERSITY OF CALIFORNIA, DAVIS** — Davis, CA
Master of Science in Statistics — Sep.2014 – Mar.2016

School of Mathematics and Statistics**, SHANDONG UNIVERSITY** — Shandong, CHINA
Bachelor of Science in Applied Mathematics and Finance — Sep.2010 – Jun.2014

SELECTED RESEARCH EXPERIENCE

Projects, University of California, Davis — Davis, CA

Overdisperion on Longitudinal Data — Feb.2016 – Mar.2016

- Applied Poisson Mixed Model and Negative Binomial Model to address the over-dispersion issue on seizure counts data.
- Estimated variance components via Generalized Estimating Equations (GEE) and Monte-Carlo EM algorithm.
- Got the standard error of model parameters by Jackknife and Bootstrap method.
- Compared the results from our models with that from covariance matrix model proposed in Thall and Vail ‘s paper in 1990.

Crime Rate Level Prediction — May.2015 – Jun.2015

- Visualized the data of 76 neighborhoods in the East Coast City by unsupervised learning methods (PCA, clustering and MDS)
- Fitted the model by the following supervised learning methods: Lasso, Random Forest, kNN, LDA) and compared the misclassification rate of different models.

Mashable News Hotspot and Popularity Analysis — May.2015 – Jun.2015

- Cleaned the original database systematically using R and explored the daily shares of articles by time, day and channels.
- Plotted World Cloud graph to show the most frequent keywords in 2013 and 2014.
- Built Random Forest model to predict the daily shares of articles and presented the most important features in prediction.

Working with Big Data: Analysis of New York Taxi Operation — May.2015 – Jun.2014

- Performed ETL on New York Taxi operation data from website (116G)
- Fit linear regression to predict total amount less the tolls using trip time and surcharge as regressors.
- Implemented the analytical tasks by Python, R, C, Shell commands and Parallel processing and compared their efficiency.

RELATED WORK EXPERIENCE

Research Assistant, University of California, Davis — Davis, CA

Residential electricity usage analysis — Jan. 2016 – now

- Performed ETL to achieve hourly observed analytical data sets on 100 GB’s original database.
- Estimated the baseline mean of household/day-specific peak time energy usage using general Linear Mixed model.
- Uncovered the efficiency of energy saving program through measuring the model uncertainty.
- Developed Bayesian Sequential Learning on previous model and converging points on different data division.

Data Science Intern, Pennsylvania State University — State College, PA

Quantifying Soil Classification by Soil Chemical and Physical Properties — Jun.2015 – Aug.2015

- Explored the global and U.S soil property’s distribution and completed the descriptive statistics.
- Ran Tukey’s pairwise comparison test on the difference among soil horizons.
- Performed K-means clustering to visualize soil order/set on PCA dimensions.
- Developed machine learning models to predict the soil orders by Random Forest, SVM and ensembles of them.

SELECTED HONORS & AWARDS

- Second Place Award, National Mathematical Modeling Contest, CHINA — 2012
- Third Place Academic Scholarship, Shandong University — 2010 - 2012

ADDITIONAL INFORMATION

- Programming: R, Python, SAS, MATLAB; Database: SQL; Distribute Computing: Spark, Hadoop.