HUIHUI DUAN

✉ duanhuihui2009@gmail.com 1-608-556-3368 New York City, NY

PROFESSIONAL EXPERIENCES

American International Group (AIG, Via Cognizant) **May/2015 – Present**
Data Scientist **New York City, NY**

- Built text mining model to predict which claims can be recovered through subrogation;

Farmers Insurance Group **Jul/2013 – May/2015**
Senior Commercial Product Analyst **Los Angeles, CA**

- Built a Logistic regression model to identify the odds ratios to become bad debt for industry, agent type, billing plan, and premium band;
- Built classification models to predict customers' behavior using logistic regression, decision tree and random forest with an accuracy of 0.88;
- Identified a fraud pattern for abusing cross sell discount, fixed the bug and recovered $10,000 losses;
- Provided cross sell report to support marketing strategy, increased cross sell rate by 3%;
- Automated 3 reports and reduced generating time from 2-3 day to 1 hour by connecting R with SQL;

SPHERE Institute **Jan/2013 – Jul/2013**
Data and Policy Analyst **San Francisco, CA**

- Implemented a regression model to suggest Centers for Medicare and Medicaid Services (CMS) to adjust hospitals reimbursement;
- Designed and implemented algorithm and reduced a report produce time from 4 hour to 40 minute;

University of Wisconsin – Madison **Aug/2009 – Aug/2012**
Research Assistant and Statistical Consultant **Madison, WI**

- Built 5 Bayesian models and 2 Non-Bayesian models to compare accuracies of predictions;
- Applied distributed system to solve advanced computing problems in Bayesian regression;
- Assisted 6 professors in data analysis, model improvement and implementation;

DATA MINING AND MACHINE LEARNING PROJECTS

Prediction

- Implemented multiple linear regression in Octave, applied a simple data set on this implementation and compared the results from package in R and module in python; **Apr/2014**
- Clustered Medicare beneficiaries in 2008 and built cluster-specific prediction model to predict 2009 medical costs using medical claims data and decreased the root mean square error from 2.1 without clusters to 1.8 with clusters; **May/2013**

Classification

- Predicted if earning is over $50K and improved testing data accuracy from base of 0.76 to 0.86 for logistic regression, 0.85 for decision tree and 0.85 for random forest; **Apr/2014**
- Classified the letter of A, B, P or R using decision tree and random forest and improved testing accuracy from 0.26 (base) to 0.88 (decision tree) and 0.98 (random forest); **Mar/2014**
- Implemented decision tree algorithm with categorical variables in python; **Dec/2013**
- Implemented support vector machine to build a spam classifier in Octave; **May/2013**
- Implemented logistic regression in Octave, applied a simple data set on this implementation and compared the results in R and in python; **Apr/2013**

Clustering

- Clustered 3,430 news articles published on Daily Kos, an American political blog, to 7 groups for searching results grouping using hierarchical clustering and k-means clustering and compared the differences among groups; **Apr/2014**
- Clustered 3,999 airlines members of frequent flyer program to 5 groups for market segmentation using hierarchical and k-means clustering and compared the differences among groups; **Mar/2014**
- Implemented the K-means clustering algorithm and applied it to compress an image; **Jun/2013**

HUIHUI DUAN

✉ duanhuihui2009@gmail.com 1-608-556-3368 New York City, NY

Text Mining

- Detected Vandalism on Wikipedia using decision tree with added and removed bag of textual words, and non-textual data, improved testing accuracy from 0.53(base) to 0.72; **Apr/2014**
- Classified whether a paper is a clinical trial testing a drug therapy for cancer using decision tree with textual paper title and abstract, improved accuracy for testing data from 0.56(base) to 0.76; **Apr/2014**

Data Visualization

- Visualized the U.S. election prediction in 2012 by state using logistic regression; **May/2014**
- Visualized a friend network, drew the vertices size by friends number and colored vertices by school using igraph package in R; **May/2014**
- Visualized text data from tweets using word clouds and displayed the words size and color by words frequency; **May/2014**

Computing Investing

- Built trading algorithm from stock price event study. At every event, buy 100 shares of the equity, sell them 5 trading days later and the compound annual return is 14%; **Apr/2013**
- Built trading algorithm from Bollinger Bands event study. At every event, buy 100 shares of the equity, sell them 5 trading days later and the compound annual return is 15%; **Mar/2013**

EDUCATION

Georgia Institute of Technology **Jan/2015 – Present**
Online Master Degree in Computer Science *Area: Interactive Intelligence & Machine Learning*

University of Wisconsin – Madison **Sep/2011 – Aug/2012**
Master Degree in Statistics *Area: Data Mining and Machine Learning*

University of Wisconsin – Madison **Sep/2009 – Dec/2012**
Master Degree about Quantitative Genetics *Area: Bayesian Statistics*

Sichuan Agricultural University **Sep/2006 – Jun/2009**
Bachelor Degree about Quantitative Genetics *Area: Statistical Analysis*

- Awarded Meritorious Winner in the 2008 International Mathematical Contest in Modeling (IMCM);
- Finished the undergraduate school within three years and awarded full scholarship to UM-Madison;

TRAINING

Johns Hopkins University at Coursera **Apr/2014 – Jul/2014**
Specialization Courses about Data Scientist *Area: Data Science*

- Finished 9 courses and course projects (The Data Scientist's Toolbox, R Programming, Getting and Cleaning Data, Exploratory Data Analysis, Reproducible Research, Statistical Inference, Regression Models, Practical Machine Learning, and Developing Data Products);

PROGRAMMING TOOLS

SQL/R/PYTHON/SAS/JAVA/OFFICE/OCTAVE/MATLAB/LINUX/HADOOP/PERL/WEKA/GIT/ANDROID/LATEX