Wenzhe(Evelyn) Xu

217-417-9270 | wxu23@illinois.edu | Address: 100 24 Ave #3 ,San Mateo, CA 94403

CORE QUALIFICATIONS

Overall Highlight: Solid data science background with data engineer skill sets applied in industry

Statistical Skill:

- Data manipulation: ETL (extract-transform-loading) data technique with large dataset (Impala 1.x)
- Modeling: machine learning (regression, classification, clustering), categorical data analysis, time series analysis, sampling, ANOVA analysis, A/B testing, dimension reduction, model selection

Computer Skill:

- Programming: expertise in R (4+yr,dplyr,ggplot2,shiny) , Python(1+yr, Scipy, numpy, pandas, scikit-learn), SQL(3+yr)
- Data engineering: Apache Impala 1.x, Apache Hadoop 2.x, Cloudera CDH 5.x, Apache Spark 1.x
- Software and system: SAS (Advanced Certified), SPSS, Looker, Linux (CentOS6.5), Microsoft Office, Map Reduce

Communication and Presentation:

- Meeting with business colleagues and transfer the commercial goal into data-driven objective
- Quick response to the later-added requirement from business side
- Presentation of the modeling result and data insight by PPT and building interactive dashboard through Shiny

Project Management: Able to track each stage according to scheduled timeline for an independent project

EDUCATION BACKGROUND

University of Illinois at Urbana-Champaign   8/2013-5/2015

Master of Science in Statistics-Analytics | Major GPA:3.96/4.0

Tianjin University of Finance and Economics, *Tianjin, China*

9/2009-6/2013

Bachelor of Science in Statistics | Overall GPA: 3.75/4.0 | Major GPA :4.0/4.0

PROFESSIONAL EXPERIENCE

Data Scientist Intern, MasterCard Inc. *San Carlos, CA*   6/2015 - 12/2015(expected)

Independently designed the automatic anomalies detection system for business metrics time series data

- Researched on anomaly detection algorithm and finally choose the S-H-ESD test from Twitter as basic logic
- Read package source code of R and re-organized the logic and input to fit the business metrics data application scenario, with integrating special requirement from business and engineer staff's execution
- Used Impala/SQL for ETL the data from production, complied the R code into mapper and reducer function in Python with each group id as mapper key, and set up the map-reduce data pipeline job on Oozie
- Build PPT for presentation and interactive dashboard from Shiny for collecting feedback of the algorithm to team and quick response to additional need from business and engineer colleagues

Collaborative tasks from other projects

- Helped to write mapper and reducer function on word count for merchant classification based on transection records
- Build Hadoop cluster of 12 nodes including tuning critical parameters, and was responsible for the maintenance on memory usage and related package installation under CentOS6.5 Linux environment
- Participated in the Cloudera training on Apache Spark

Tech-Sale Analytics Intern, Anheuser-Busch Inbev, *Champaign, IL*   5/2014 -5/2015

Customer profiling with social media data source and internal wholesaler database

- Learned the original algorithm and created advanced string cleaning step for Foursquare/Yelp database
- Improved the whole workflow by adding lat/lon information as well as innovative filters to increase mapping accuracy
- Wrote and integrated all steps in pipelined R code for future usage and cooperated with 3rd party to build dashboard
- Overcame the challenges of handling Asian languages and popularized the customer profiling project to global market
- Mentored new interns for taking over the customer profiling work flow

Quantitative exploration experiment projects (Ad Hoc)

- Predicted volume for each POC through logistic models and learning algorithms, identified the influencing POCs
- Used regression tree to predict the number of pipe needed by the target volume for Belgium on-premise market
- Transfer the data analysis result to commercial insight and presented to business managers on a regular basis