Zhaolong Lu (Luke)

(312)-888 6529, lzhaolon@gmail.com

SUMMARY

Advanced data analyst with quantitative modeling skills. Having working experience on both machine learning and statistical modeling using SAS, R, SPSS and Python in Marketing and Medical Fields; Solid skills in Database Programming with PL/SQL; Hand on experience in Monte Carlo simulation and Time Series Modeling; Well-trained knowledge in, Hadoop ecosystem, Hive, Pig and Mahout, SparkR; Applied skills on data visualization using R(shiny apps, ggplot), Tableau and Gephi.

EDUCATION

DePaul University, Computing and Digital Media, Chicago, IL **Expected November, 2016**

Master of Science in Predictive Analytics

DePaul University, Kellstadt Graduate School of Business, Chicago, IL **June, 2014**

Master of Science in Marketing Analysis

PROFESSIONAL EXPERIENCE

INSTITUTE FOR HOUSING STUDIES-DePaul University , Chicago, United States **September, 2015 - Present**

Data analyst intern ------ R, Python, Tableau

- Analyzed the margin of error and missing values of data gathering from sources internal and external
- Housing Segmentation study on 7 counties in Chicago area using clustering techniques such as K-means, K-medoids, Pam
- Anova test and boxplot plots to confirm analysis' conclusions of clusters together with geographical map
- Visualized results with geographical heat maps at census tract level using Rstudio and Tableau
- Developed Shinny app in Rstudio enabling interactive heatmap, geomap, boxplot and text input for better visualization

LOYOLA MEDICAL CENTER, Chicago, United States **March, 2015 – August, 2015**

Data Scientist Intern at OnetoMap Analytics LOYOLA SURGERY ------ R,Python

- Communicated with medical students for gaining medical knowledge on dataset of 13,000 patients between years 2008-2013
- Used paired t-test to identify the differences of patients' day-night shifts given medical treatment
- Executed Chi-squared test to testify the significance of the access frequency of patients' information database
- Setting bench mark using Tree model in different criteria such as C4.5, CART and CHAID
- Compared logistic regression and tree model in classifying patients' mortality status
- Responsible for the findings that intensive health care lead by significant day-night shifts and frequent database accessing helped to decrease mortality rate

NATIONAL LEVEL DATA MINING CASE COMPETITION, Chicago,United State **April, 2016 - Present**

Rang Technology and KVRA ------ R

- Goal is presenting the best binary classification from real-world data from retail industry concentrating on accuracy
- Manipulating data by permutation and removing missing value using VIM package
- Feature Engineering using random forest (BORUTA), Logistic regression with penalty (Lasso,glmnet)
- Comparing different binary classification method based on selected features

FC CONSULTANT GROUP, Chicago, United States **January, 2015 – April,2015**

Analytics of FC Consultant Group ------ SPSS

- Understand client's business expectation, identified gaps, presented quantitative analytics solutions to help clients make decisions
- Built marketing strategies based on PEST, Five Forces model, SWOT analysis, distribution analysis and competitor analysis
- Collected and aggregated raw data from different resources, manipulated data into efficient and informative format for modeling
- Developed and implemented quantitative methods to discover key factors affecting customers' attitude of purchasing
- Targeted customers by analyzing customer segmentations using k-means and hierarchical clustering methods

PROJECT EXPERIENCE

MINING BIG DATA **Spring 2015-2016**

Analyzing Chicago Crime data using HIVE and PIG------HADOOP

- Pre-setups for building multiple clusters within Hadoop environment in Linux environment
- Using PIG and HIVE queries extract data into practical analysis
- Comparing time used for single and multiple nodes under HIVE and PIG queries

TIME SERIES ANALYSIS **Autumn 2014-2015**

Microsoft Stock Price Prediction------R

- Examining the Autocorrelation and partial autocorrelation analysis of the data
- Fitting data using ARMA and GARCH model with the different distributions for Residual Analysis
- Establishing IGARCH Model to Predict the Stock Price of Microsoft

PROGRAMMING DATA MINING APPLICATION **Autumn 2015-2016**

Collaborative Recommender System for Music website last.fm ------Python

- Compared standard similarity and SVD similarity from both speed and accuracy aspects
- For each similarity methods using similarity functions such as Euclidean, Cosine, Pearson Correlation to select best model
- Selected SVD method with Euclidean similarity measurement to build the recommender system

CERTIFICATE

- SAS Certified: Advanced Programmer for SAS 9; SAS Certified Base Programmer for SAS 9
- Coursera: R-The Data Scientist's Toolbox, R Programming; Python- An introduction to interactive Programming in Python; Bigdata and NoSQL database: Introduction to Big data