TIANYUE (ANDY) MAO

Open to relocation | Permanent Resident | (314)983-8018 | tianyuemao@gmail.com | tianyueandymao.com

EDUCATION

University of Illinois at Urbana-Champaign Champaign, IL

Master of Science in Statistics, GPA: 3.8/4.0 Expected Graduation Date: May 2018

Relevant Coursework: Applied Regression and Design, Statistical Learning, Foundations of Data Science, Statistics Programming Methods, Statistical Data Management, Advanced Data Analysis**,** Mathematical Statistics, Applied Multivariate Analysis, Databases, Data Mining, Data Visualization

Bachelor of Science in Statistics Graduation Date: May 2017

PROFESSIONAL EXPERIENCE

Dow AgroSciences, *Business Analyst Intern* May 2017-Present

- Determine expected duration of global product registration progresses by creating features and fitting predictive models in R
- Save the company $2 million a year by implementing Tableau visualizations and filters to protect data integrity
- Accelerate workflow by developing data entry forms and designing SQL database framework for registration tracking
- Demonstrated value of accessing historical data to apply machine learning to provide insights for future registrations

Axis Capital, *Statistical Consultant* Aug 2016-Dec 2016

- Conduct research to build reinsurance models that would best forecast future rates using historical loss data
- Assist clients in understanding findings through building visuals such as loss triangles, forecasting graphs, and ratio tables

PROJECT EXPERIENCE

Analysis of Amazon Reviews to Predict Users Ratings, *Project Lead* Jan 2018-Present

- Use text mining techniques such as regular expressions and bag of words to clean 3 million Amazon reviews
- Create machine learning models on newly built features to test if predictability of user review scores is improved
- Utilize dimension reduction techniques such as PCA, SIR, and LASSO to eliminate unimportant features to the model

Text Mining and Classification on Twitter Dataset, ***Analyst*** Mar 2017-May 2017

- Conduct text cleaning on Twitter dataset to create new features/predictors, such as bag of words, to increase predictions
- Utilize machine learning models such as RF, GBM, RDA, NB, and KNN to classify user gender and bot detection
- Achieve best accuracy of 66% for gender classification and 87% for bot detection using Random Forest w/ensemble methods

Data Cleaning Application to Improve Accessibility, *Project Lead* Jun 2016-Jan 2017

- Combine data cleaning, management, and visualization techniques into a single RShiny interface
- Apply regular expressions to clean the data while using bootstrapping and EM algorithm to impute missing values

EXTRACURRICULAR

Illini Statistics Club, *President/Graduate Advisor* May 2016-Present

- Collaborate with employers and other on-campus organizations to provide a wide scope of opportunities for club members
- Successfully hosted the inaugural club data analytics competition and provided the opportunity for over 60 passionate students
- Lead and manage other executive board members to create a collaborative and productive working environment

Data Analytics/Machine Learning Competitions

- Saint Mary's Data Analytics Competition, *Top 5 (25 Teams)* Mar 2018
- Statistical Learning Kaggle Competition, *Top 10 (70 Individual Competitors)* May 2017
- Midwest Undergraduate Data Analytics Competition, *Top 5 (37 Teams)* Apr 2017

SKILLS

Statistical Software: R, SAS, Tableau, WEKA, RShiny

Programing Skills: Python, SQL, Regex, Natural Language Processing, Java, SharePoint, Hadoop, MapReduce, Unix

Statistical Knowledge: Regression, Variable Selections, PCA/SIR, Factor Analysis, Discriminant Analysis, Clustering Analysis, Hypothesis Testing, Categorical Data, Dummy Variables, Data Imputation, Resampling, Bootstrapping, etc.

Machine Learning: RF, KNN, Logistic, Naïve Bayes, SVM, Ensemble methods, Cross-Validation, Test/Train Split, etc.

Languages: English (native), Mandarin (native), French (hobby)