SHIQI YU

Email: yushiqi1999@163.com
Phone: (336)4734719
Add: 113 Valentine PL, Ithaca, NY

EDUCATION BACKGROUND

Cornell University **Ithaca, NY**

Major: Operations Research and Information Engineering; Concentration: Data Analytics 08/2021 - 05/2022

- Core courses: *Optimization I, Operation Research Tools for Financial Engineering, Data Mining and Machine Learning, Data-driven Marketing, Case Studies, Big Data Technologies, Machine Learning Applications in Business*
- Technical Skills: RStudio, RMarkdown, SQL, Python, JAVA, Excel, Word, PowerPoint

Wake Forest University **Winston-Salem, NC**

Major: Mathematical Statistics; Minor: Psychology 08/2017 - 12/2020

- Overall GPA: 3.92/4.0
- Core courses: *Time Series and Forecasting, Probability, Statistical Inference, Linear Model, Introduction to Computer Science, Introductory Real Analysis I, Multivariate Statistics, Design and Sampling*
- Honors & Awards: Dean's List Scholar (Fall2017, Spring2018, Fall2018, Spring2019, Fall2019, Spring2020); Wake Forest Research Fellowship (Mar 2020);
- Job on campus: TA for Multivariate Statistics Course (01/2020 – 12/2020)

INTERNSHIP EXPERIENCES

ByteDance Ltd. **Beijing, China**

Business Analyst, Business Analysis Department 03/2021 - 05/2021

- Collected and compiled personal information sheet from each base
- Created pivot table, wrote analysis, and optimized the computations of employee's turnover rate and man-hour utilization rate

ZHICHENG Technology Group Limited **Beijing, China**

Data Analyst Intern, Data Technology Department 07/2019 - 08/2019

- Collected and cleaned monthly sales data and revenue data
- Compared the average monthly sales of last year, analyzed the internal (regional differences) and external (market competition, policy, market capacity) reasons for the sales data fluctuations of Beijing branch and put forward reasonable suggestions

RESEARCH EXPERIENCES

Senior Thesis Project in Text Data Mining-Modeling Hate Speech **Winston-Salem, NC**

Team leader, Advisor: Dr. Nicole Dalzell 08/2020 – 12/2020

- Cleaned and processed a large text data set (about 160,000 observations) into tidy version
- Used Logistic Regression and Classification Tree to classify comments into different levels of toxicity
- Tested accuracy of the final model by comparing the classification error rates in test data and train data
- Finished a comprehensive write-up and presented the project

Wake Forest University Research Fellowship-Curve Registration in Regression Model **Winston-Salem, NC**

Research Assistant, Advisor: Dr. Sneha Jadhav 06/2020 - 09/2020

- Used B-spline basis to smoothed practice data
- Learned Basis Expansion and applied it to Warping Function and Beta Function
- Compared and contrasted different misalignment conditions through comparison plots and calculated their mean curves
- Wrote functions to build Functional Regression Model, get estimated coefficients, and calculate mean standard error in R
- Used k fold cross validation to test how well our model behaved on new data points

SELECTED COURSEWORK

CSC111 Introduction to Computer Science Final Project **Winston-Salem, NC**

Team Member, Advisor: Dr. Minghan Chen 04/2020 - 05/2020

- Self-learned programming in JavaScript; created a Snake Game in JavaScript

STA279 Categorical Data & Multilevel Model Final Project **Winston-Salem, NC**

Team Member, Advisor: Dr. Nicole Dalzell 03/2020 - 05/2020

"What Factors of Employee in IBM Company Have an Impact on the Level of Their Job Satisfaction"

- Cleaned the data and did exploratory data analysis on potential predictors
- Fitted the Multinomial Regression and run ANOVA test of nested models to pick the best one
- Calculated the final model's classification error rate and discussed the future scope

MST256 Statistical Model Final Project **Winston-Salem, NC**

Team Member, Advisor: Dr. Nicole Dalzell 09/2017 - 12/2017

"What Factors of Potatoes Influence the Flavor"

- Created boxplots to choose possible explanatory variables
- Used backward elimination and Nested F test to select significant predictors, then checked conditions for inference