Jishen Yin

(984)-245-5596 • jishen.yin@duke.edu
Durham, NC 27705

CAREER GOAL

To work on real-life challenges and develop solutions that have a deep impact on how businesses and people thrive with statistical tools.

EXPERTISE

Programming Language: Python, SQL, R
Software: SAS, MATLAB
Skill: Statistical Model, Machine Learning Algorithm, Hierarchical Model

EDUCATION

Duke University NC, USA
Department of Statistical Science, MSS Program 09/2019 – Current
Master of Science expected in June, 2021, GPA: 3.8/4.0
Related Course: Predictive modeling, Statistical Programming, Machine Learning, Modern Bayesian Statistics

Beijing Institute of Technology (BIT) Beijing, China
School of Mathematics and Statistics, Mathematics Elite class 09/2015 – 06/2019
Bachelor of Statistics and Economics, GPA: 3.9/4.0

INTERNSHIP EXPERIENCE

Clarivate Analytics Company Beijing, China
Products and Solution Intern (*Python&Excel&SQL*) 10/2018 – 11/2018

- Collected information of subsidiaries of 94 State-owned enterprises in China using Python Crawler from *qichacha.com*
- Searched for patent data of certain enterprises using SQL from database *Derwent Innovation*, double-checked it with subsidiaries information using Excel and filtered out patent data wrongly labeled/without a label.
- Modified label on *Derwent Innovation* using SQL.

PROJECT & ACADEMIC EXPERIENCE

Final Project of Duke course "Predictive Modeling" 12/2019
Team project on Predicting the price of artwork exhibited on Paris Painting Exhibition(*R*)
https://github.com/YinJishen/Paris-Painting-Project

- Cleaned original datasets with 59 variables and 3,393 data points through deleting variables containing useless and overlapping information, filling in missing value with Bayesian model and creating new categorical variable representing the fame of the author.
- Built a two-stage predictive pipeline including BIC model selection and Random Forest algorithm and got top 15 significant predictors.
- Combined with cross-validation to set up parameters in the model, made predictions on test data and gave a group presentation with Q&A session.
- Got the lowest MSE and highest coverage rate on predicting test data among all 13 groups.

Contest of Data Challenge at UNC 10/2019
Team project on Analyzing Crime Data of Chapel Hill (*Python&Tableau*)
https://github.com/YinJishen/CDC2019

- Applied kernel-based clustering algorithm to locate the cluster centers and distribution of crime incidents across geographical coordinates, interacting with Chapel Hill map data.
- Visualized results through Heatmap, Contour plot and Scatter plot with Tableau, and gave a group presentation with Q&A session.