

Jiayu Fan

Fjy501@live.com | 508-7367887

EDUCATION

Temple University | Philadelphia, PA

Aug. 2022 – Present

Fox School of Business

PhD Student, Operations and Supply Chain Management

Awarded Presidential Fellowship

Research interest: Machine Learning, Forecasting, etc.

Clark University | Worcester, MA

Aug. 2019 – Jun. 2021

Master of Science, Business Analytics

Cumulative GPA: 3.96/4.00, Beta Gamma Sigma Lifetime Member, School of Management Scholarship

Core Courses: Analytics Programming (Python), Big Data Statistics (R), Business Intelligence (Tableau, KNIME), Database Management Systems (SQL, Oracle, Access), Advanced Big Data Computing & Programming (Hadoop, Spark)

Shandong University of Finance and Economics | Shandong, China

Aug. 2013 – Jun. 2017

Bachelor of Science, Statistics

Cumulative GPA: 3.24/4.00, Merit Scholarship, Merit Student

Core Courses: Mathematical Statistics, Advanced Mathematics, Advanced Algebra, Sampling Technology, Multivariate Statistical Analysis, Nonparametric Statistics, Time Series Analytics, Statistical Software (R, SPSS), Data Mining

RESEARCH & PUBLICATION

Conference

- **Jiayu Fan** (Presenter), Jie Bian, Yue Gao. (2021, October 27). Predictive Analysis of Greenness of U.S. Congress Members Using Machine Learning Techniques. INFORMS Annual Meeting
- Zejian Wu, Shuyu Zhang, **Jiayu Fan**, Chenhao You. (2020, November 13). Determinants of Car Accident Severity. INFORMS Annual Meeting

Publication

- **Jiayu Fan**. (2017). Analysis of Tourists' Attitude for Ancient Towns Based on Text Mining. Big Data Research, 3(6), 93-101. DOI: 10.11959/j.issn.2096-0271.2017064

PROJECTS

Predictive Analysis of Greenness of U.S. Congress Members

Mar. 2021 – Oct. 2021

- Through literature review, learned the potential factors that impacted senators' ideology toward pro-environmental bills. Collected demographic data and environmental relevant data at state level from United States Census Bureau, MIT Election Data & Science Lab, League of Conservation Voters, Energy Information Administration, etc. Merged data by state. Conducted descriptive analysis. Used LCV scores, as the target, to measure senators' propensity of supporting a pro-environmental bill. Selected features through correlation matrix. Transformed categorical variables into dummy variables. Standardized independent variables
- Applied supervised machine learning techniques to predict target—Nonlinear Regression, Random Forest, K-Nearest Neighbors and XGBoost. XGBoost yielded the smallest RMSE. For improving the interpretability of XGBoost model's results, ranked features' importance based on their absolute SHAP value, an approach based on Shapley value to explain outputs of machine learning. According to our results, senator's party affiliation is the most influential factor in the decision process of voting for pro-environmental bills. Especially, senators who affiliate to democratic party are more likely to support environmental conservation bills
- The study is valuable. For residents, they could choose an eco-friendly place to settle down by predicting the legislators' greenness. For manufacturers or enterprises, they could predict the trend of conservation regulations by predicting legislators' voting behavior, so that they can pre-respond to any change in the future

Determinants of Car Accident Severity | CU, MA

Jun. 2020 – Oct. 2020

- Collected data involved records of car accidents in 2018 from U.S. Department of Transportation. Reencoded and Unified missing values, then dropped the columns containing over 80% missing values. Dropped missing values. Re-encoded categorical variables into ordinal variable. The target variable was the number of injured people in an accident when built models at vehicle level. Another target variable was drivers' injury severity when built model at personal level. Both targets were scaled from zero to five. Clustered target variable from five classes into two classes by dividing them into injury and non-injury classes. Used oversampling technique, Smote, to boost minority classes
- Built classifier models from personal level and vehicle level to predict the driver's injury and injury condition for all occupants. Investigated the determinants of car accidents severity through building models with input features, including the information of vehicle involved in the accident, drivers' demographic characteristics, and surrounding environment such as road conditions. In this research, Random Forest outperformed other models. According to feature importance ranking, the

type of crash a vehicle involved in, number of occupants, etc. played an important role in determining whether an accident would cause an injury or not. This study would make contributions to car accident prediction and injury prevention

Tourists' Attitude Toward Ancient Towns / China

Jun. 2017

- Obtained text reviews about five ancient town from Mafengwo, a website parallel to Tripadvisor, which provided travel deals and travelers' reviews. Counted words frequency. Found out that admission fees to those scenic spots and words like "over commercialized" are frequently mentioned in reviews
- For investigating the potential reasons that caused most negative reviews, comments were resampled and divided into two groups according to whether comments included the key words about admission fees or commercialization. Conducted sentiment analysis to transform each comment into calculable number. Applied T-test to verify if two group of people processing obvious different satisfaction level. According to the result, comments mentioned admission fees or commercialization earned lower sentiment scores than those didn't. Therefore, improper tickets' price and overly commercialized scenery in ancient towns were the main reasons for negative reviews
- According to previous literatures, relevant demographic features were added in dataset in order to carry out a general sentiment score for the five ancient towns. New features such as the third industry indexes of different provinces were used to weight sentiment scores derived from various towns, instead of just simply averaging all scores. This study would provide insights for relevant departments with effective management of ancient town tourism

WORK EXPERIENCE

Teaching Assistant | Dr. Yue Gao (yugao@clarku.edu), Clark University, MA

Mar. 2021 – Jun.2021

- Worked as a teaching assistant for the course Machine Learning
- Helped other students with their doubts in homework and programming
- Graded homework and exams
- Held review session, and recorded programming introduction

Database Administrator | Guizhou Dongguan Technology Co., Ltd., Guiyang, China

Aug. 2018 – Nov. 2018

- Assigned to the Data Support Department of China Telecom, Guiyang Branch
- Managed Oracle database and imported daily generated data to Oracle database
- Composed SQL syntax to get the required tables and supported decision making

SKILLS

Programming: Python, R, SQL, SPSS

Software/Tools: Microsoft Office, Oracle Database, PL/SQL developer, Tableau, KNIME, Overleaf, Spark, Databricks

Skills: Data Cleaning, Data Visualization, Machine Learning, Relational Database Management

Operation System: Microsoft Windows, Apple macOS, Linux