Yiting Lan

E-mail | **Phone** lan.106@osu.edu | 6148041272 **GitHub** www.github.com/yitingyiting

Personal Home Page http://yitingyiting.github.io **LinkedIn** www.linkedin.com/in/yiting-lan

About 5 years of research experience in data science. Self-motivated researcher with solid data analysis, manipulation, visualization, modeling, and writing skills. Responsible, creative, organized, good interpersonal skills, can work both independently as well as in a team. Seeking full-time position starting Mar 2021.

Skills

Tools: SQL, Python, R, Tableau, STATA, SPSS, SAS, Excel

Languages: Python(Sklearn, Numpy, Pandas, Matplotlib, PySpark), R(ggplot, dplyr), SQL

Skills: Data Cleaning, A/B Testing, Hypothesis Testing, Predictive and Statistical Modeling, PCA, Linear/Logistic Regression, Random Forest, SVM, KNN, Neural Network, Gradient Boosting, XGBoost

Education

Aug 2017 - May 2021 **The Ohio State University, Ph.D., Consumer Economics, Columbus, Ohio**

- **Graduate minors: Computer Science and Engineering, Statistics**
- Selected courses: Machine Learning, Database System, Data Mining, AI, Statistics, Econometrics, Multivariate Experimental Design

Aug 2015 - Apr 2017 **The Ohio State University, M.S., Consumer Sciences Specialized in Applied Economics, Columbus, Ohio**

- College of Education and Human Ecology Scholarship, Research Graduate Student Travel Award
- Selected courses: Exploratory Data Analysis, Econometrics, Consumer Behavior, Applied Nonparametric Statistics

Experience

Aug 2017 - present **Graduate Research & Teaching Associate**

The Ohio State University, Columbus, Ohio

- Designed empirical methodologies in applied econometrics, causal inference and machine learning for quantitative research in area of consumption, social policy and health.
- Conducted literature review, data collection, cleaning, exploratory data analysis and data visualization, prepared reports, draft results independently and collaboratively with other researchers for more than 15 projects.
- Shared responsibility of designing, lecturing and grading undergraduate-level statistics (Excel) and consumer behavior courses, corresponded with about 100 students on inter-campus communication system.

May 2017 - Aug 2017 **Data Analyst Intern**

Fuyoulonghui Genetic Disease Clinic, Beijing, China

- Conducted data analysis and data visualization on patient visits and re-examination, pharmaceutical product revenue, and performance of sales representatives via Excel.
- Collaborated with managers to design the needed database which reduces the burden of data input and storage.
- Assisted with conducting an internal training survey for over 30 sales representatives.

May 2016 - Aug 2016 **Web Analyst Intern**

New Oriental Education & Technology Group, Beijing, China

- Assisted the analyst with data collection, summary, and visualization for pageviews on different topics in the English Study area.
- Prepared report on performance to support topics selection and other decision making.

Projects

1. Predicting Food Delivery Time [Python, Modeling]

- Conducted data cleaning, data preprocessing, data visualization, data transformation such as PCA via **Python(Scikit-learn, Numpy, Pandas, Seaborn, etc)** using one of the food delivery datasets from Doordash, which includes over 4 million food delivery records.
- Explored performances for multiple models, such as **deep neural network** and **XGB Regressor**, and achieve root mean square error as about 600 seconds.

2. Predict Alcohol Consumption [Python, Machine Learning]

- Conducted data cleaning, data preprocessing, and data visualization via **Python(Scikit-learn,Numpy, Pandas, Seaborn**, etc) using consumer expenditure datasets to predict the consumption of alcohol.
- Explored performances for multiple models including **random forest**, **gradient boosting** and **XGBoost** with AUC 0.78 and achieved 77.43% accuracy, created feature importance, and found out the expenditure in food-away from home is the most important feature in three tree models.

3. Redemption of Benefit Predicting Analysis [Python, SQL, Machine Learning]

- Pulled out about 600k participants' data with over 40 million records via **MySQL**, conducted data cleaning, exploratory data analysis with **Python(Pandas, Numpy, Seaborn**) designed data dashboard with **Tableau** and presented to partners in Ohio Department of Health.
- Trained several models with **Python (Scikit-learn)**, including **logistic regression**, **decision tree**, **neural network**, **support vector machines** and **random forests** to predict redemption.
- Achieved 88.62% accuracy (f1 score as 0.88) predicting whether consumers would redeem their benefits.