

Liuyi Bao

Belmont, MA|+1-5083082601| Liuyib861@gmail.com|www.linkedin.com/in/liuyi-bao

SUMMARY

Programming: MySQL, Python (NumPy, Pandas, scikit-learn), R (tidyverse), Excel (VLookup, Pivot table)

Data Visualization: Tableau, Matplotlib (Python), ggplot (R), JMP, MS Excel, MS SQL Server

Machine Learning: Linear/Logistic Regressions, KNN, Clustering, Decision Tree, Random Forest, etc

Techniques: A/B Testing and Experimentation, Data Reduction, Data Visualization, Data Warehouse, Data Wrangling, Data Mining, Database Management, Mathematical and Statistical Modeling.

EDUCATION

University of California, Davis

Davis, California

Sept 2020 - Dec 2021

- **Master of Science, Statistics (STEM)**
- **Certification:** Advanced Google Analytics
- **Courses:** Big Data & High-Performance Statistical Computing, Computational Statistics, Data & Web Technologies for Data Analysis, Optimization for Big Data Analytics, Probability Theory, Statistical Data Science

Iowa State University

Ames, Iowa

Aug 2017-May 2020

- **Bachelor of Science, Statistics (STEM)** (GPA:3.76/4.0)
- **Rewards:** Deans' list, Marian Daniells Scholarship
- **Courses:** Calculus, Computer Processing Data, Acquisition and Exploratory Data Analysis, Theory of Probability Statistics

WORK EXPERIENCE

Shanghai Fumasoft Co., Ltd.

Shanghai, China

Project Manager Internship

Jul 2021-Sept 2021

- Built 20+ dashboards with the Sensorsdata platform (including business models and features similar to Tableau) for customers' usage of the company's Customer Relationship Management platform.
- Utilized SQL (subqueries, window function) for data verification and correction, supervised the data tracking process, discussed and managed the timeline for the project, and communicated with data engineers.

Zhejiang Baiyite Plastic Co., Ltd.

Taizhou, China

Data Analyst Internship

Sept 2019- Dec 2019

- Identified, analyzed and executed seasonal profitable plastic products among 600+ products with Excel and R. Created multiple Excel inventory planning and analysis sheets to improve supervision efficiency.
- Utilized statistical models to reduce the cost and optimize the profit for several products.

RESEARCH EXPERIENCE

Machine Learning of House Prices Prediction

Davis, CA

Course: Big Data & High-Performance Statistical Computing

Mar 2021-May 2021

- Conducted data cleaning for 1000+ Housing data in Python (NumPy, Pandas), including dealing with missing values, wrong types of data, and data anomalies.
- Utilized Exploratory data analysis, feature engineering and variable Selection for 80+ variables to establish K-Nearest Neighbor Model, Random Forest and Gradient Boosting Decision Trees.
- Decreased model errors after parameter tuning with Python (sklearn), found the best model GBDT with the lowest RMSE and visualized top features for housing prices.

Research on the Prediction of Liver Disease

Davis, CA

Course: Fundamentals of Statistical Data Science

Sept 2021-Dec 2021

- Performed data cleaning and explanatory data analysis for 500+ liver disease data in R (tidyverse, ggplot), including data transformation and validation to prepare for Logistic Regression, Linear Discriminant Analysis and K-Nearest Neighbor Model.
- Built the graphs of ROC curve and confusion matrix, compared Accuracy, F1-score, MSE, Log-loss, and found K-Nearest Neighbor Model (K=11) as the best model.