Yaoyao Cai

Address: Albany, CA 94706 | Telephone: 510-616-8911 | Email: cai.yao@northeastern.edu

SKILLS

Mathematics	Linear Algebra, Probability Theory, Statistics, Optimization
Microsoft Office	Word, Excel, PowerPoint
Programming Language	SQL, Python, R
Data Visualization	Matplotlib, Seaborn, Tableau
Machine Learning	Regression, Decision Tree, Random Forest, Cluster

EDUCATION

Northeastern University, San Jose, California — Sept. 2019 – Present
Master of Professional Studies in Analytics (GPA 4.0) — Expected Apr. 2021
DataCamp: Data Scientist with Python Certificate

Zhejiang Sci-Tech University, Hangzhou, China — Sept. 2011 – Jun. 2015
Bachelor of Engineering in Textile Engineering

PROJECT

Machine Learning - The Longevity of Non-Profit Organizations — Dec. 2019

Purpose: Using the data provided to predict the longevity of the NPOs in the United States. The database contains 7 Million records from 1988-2014.

- Collected data from NCCS and Census Bureau and merged them by customized function
- Transformed some variables and visualized the description results
- Used cluster analysis to divide data into three clusters and tried to figure out funding sources patterns in each cluster among poor and not poor zip codes
- Fitted data with decision tree model and random forest model, compared the actual and the prediction to get the confusion matrix (overall accuracy score: 0.83)
- Provided a new perspective to inspect NPOs funding pattern and their life

Australia Rental House Recommendation — Oct. 2019

Goal: Using data set to analyze the factors of rental price in New South Wales and then providing pricing recommendations.

- Downloaded dataset and cleaned it to tidy data
- Explored quantitative analysis on numerical variables for further feature engineering
- Stepwise python function: performed the universal testing and removed all variables with a p-value greater than 0.5
- Visualized the patterns between price and other features (e.g. square meters, bathrooms)
- Fitted different regression models according to areas and chosen the appropriate one
- Provided pricing recommendations to both the landlord and the tenant

EXPERIENCE

Shacun Community Center, Huzhou, China — Sept. 2017 – Sept. 2018

Dataset manager

- Built a basic data management system and maintained documents and other infrastructures for the community center
- Helped community center establish the rules for collecting data and building datasets
- Provided ideas that how to apply data science to improve the living conditions of residents
- Wrote monthly reports and designed infographic for the community