Jiayin Guo

Contact

Phone: +1 510 850 7956
Email: guojiayin@berkeley.edu
LinkedIn:
linkedin.com/in/jiayin-guo-8b6880198/
GitHub:
github.com/vulcan480

Education

University of California, Berkeley
Aug. 2021 – Dec. 2022
Industrial Engineering & Operations Research (MEng)
GPA: 3.88/4.0

Tongji University
Sept. 2016 – July 2020
Environmental Engineering (BEng)
GPA: 3.72/4.0

Skills

[Program Languages]
- Python
- SQL
- R
- HTML/ JavaScript/ C#

[Machine Learning]
- Time Series and Neural Network
- Clustering
- Linear Methods and Trees

[Statistics]
- Hypothesis Testing
- Regularization
- Experiment Design

[Web Development]
- Database management
- Python Flask
- React/ Nodejs/ jQuery

Relevant Experience

- **Data Scientist Intern:** ***Applied Materials (Santa Clara, CA)***
  May 2022 – Aug. 2022
  - Queried data from the database, and built a forecast model for Non-standard parts (NSOs) based on clustering and time series analysis, which is used to predict the NSOs' BOM every quarter. More than 3000 parts were saved from wasting in the 2022Q4 supply chain planning from implementing this model.
  - Built a web application using Python Flask as the project front end, in which Data Tables and Tableau were used to display NSO forecast results, and WTForms were used to get user-provided information. The app enabled direct and automatic communication between the database and the users, slashing over 80% of emails regarding the project between developers and non-developers.
  - Used python and SQL to build the back end – Flask – front end (app) structure and its connection to the database in Pycharm, integrated separate code into one system, and saved up to 70% code running time.

- **Cost-benefit Analysis of Solar Panel Recycling** ***(Berkeley Capstone)***
  Sept. 2021 – May. 2022
  - Worked as the leader in a team of four, designed the project goal and schedule; divided tasks and checked on the team's working process, maintaining a good team working environment by constantly communicating with teammates and project advisor.
  - Collected and analyzed past data from California, and built a forecast model to predict future cost and carbon footprint of different solar panel recycling methods using Trees. Advice on selecting the most cost and environmentally friendly method is given based on the model.
  - Wrote the technical part of the final report; gave a final presentation to the project advisors.

- **Data Scientist Intern*: Tencent (Chengdu, China)***
  Feb. 2021 – May. 2021
  - In Treasure Sichuan Project, conducted A/B testing to decide what is the best time in a day to advertise snack manufacturers in WeChat's Moments; used scikit-learn to build a linear-regression model, drawing inference on the most important features relating to the impact of the online advertising posters of the project, eg., text size and background color.
  - The average monthly sales of the Treasure Sichuan project increased by 17% after adopting the best advertising time and poster format supported by the analysis.
  - Built a WeChat app using R's shiny app, in which users can interact with and see what's the most frequently used text format and images in the Treasure Sichuan posters; simply sentiment analyses were provided to show what the posters were conveying.

- **Data Scientist Intern*: China Southwest Architecture (Chengdu, China)***
  June. 2020 – August. 2020
  - Collected data and conducted exploratory data analysis (EDA) and PCA on the energy-saving rates of the existing green roof projects; wrote a report to summarize the findings. The results are used to assist the architects in preparing for the Sichuan 2019 Green Roof Competition.
  - Summarized the information of overseas projects in the past three years through Tableau; categorized the project characteristics in different continents, thus helping project managers to choose the location of future plans.

Relevant Course Projects

- **Predicting the Fraudulence of Job Posts with Binary Classifier** ***(STAT 154 @ UC Berkeley)***
  Spring 2022
  - Conducted data cleaning and feature engineering to create relevant text features, including summarization, PoS tagging, and vectoring of the words using Glo Be Dictionary; used multicollinearity analysis and ridge to do the feature selection.
  - Built three separate models using random forest, support vector machine, and XG Boost selected the best one based on the performance test. The best model (SVM) reached 0.984 accuracy, 0.999 specificity, and 0.688 sensitivity.

- **COVID Cases Analysis in California** ***(DATA 200 @ UC Berkeley)***
  Fall 2021
  - Conducted Exploratory Data Analysis, including heatmap, histograms, scatter/line plots, and PCA plots to visualize the relationship between covid cases and location, vaccination rate, and mask usage in California from January 2020 to September 2021 in California.
  - Based on the EDA results, propose hypotheses on the relationship between the number of covid cases per capita and local population density and mask usage. Two multiple linear regression models were built to test the hypothesis.