Shi Fan Jin
emily00718@gmail.com | 0979527738
https://github.com/shifanjin/my-project-job | https://www.linkedin.com/in/shi-fan-jin-6a1788117/

Profile

Data analyst with a strong background in analytics, statistics and math. Experience in data management, ETL, data modeling, visualization and optimization. Experience in building and analyzing machine learning models. Enthusiastic about translating analysis results into valuable information for both technical and non-technical audiences.

Education

University of California, Berkeley *Berkeley, CA*
Bachelor of Arts, Statistics *Aug. 2017 – May 2019*

De Anza College *Cupertino, CA*
Dean's List 2015 - 2017 *Aug. 2015 – June 2017*

Relevant Coursework
Data Analysis | Machine Learning | Principles of Data Science | Database Systems | Time Series | Data Structures | Computer Programs Structure | Sampling Survey | Business Analytics | Pricing

Analytics and Machine Learning Experiences

Spam/Ham Email Classification *April 2019*
- Created machine learning classifiers that can distinguish spam emails from ham (non-spam) emails
- Achieved extraordinary performance on our model while minimizing overfitting using SciKit-Learn library

Twitter Content Analysis *March 2019*
- Transformed and processed twitter data with Python and RegEx
- Built sentiment detection on tweet data using multiple analytical techniques with VADER lexicon

SF Food Safety Investigation *Feb. 2019*
- Identified anomalies in safety reports by exploring data at different levels of granularity using pandas
- Performed many data processing, data analysis, modeling and visualization using Seaborn package
- Predicted food safety scores by applying probability sampling techniques

London Stock Price Time Series Forecasting *Nov. 2018*
- Forecasted UK stock prices by performing data exploration and data transformation on time-series pricing behaviors and trends
- Incorporated seasonality factors with statistical techniques such as cross validation and ACF/PACF plots to achieve improvements in candidate models
- Increased prediction accuracy with suitable ARIMA/SARIMA models

Forest Fire Area Prediction *Nov. 2018*
- Increased accuracy by 7% using selected machine learning models for the prediction of burned area
- Developed and implemented linear regression, cross validation, regression tree, support vector machine and KNN to monitor models' outcomes and performance for the best prediction

Experiences

Intern, *Cathay United Bank, Taipei, Taiwan* *Aug. 2016 – Sept. 2016*
- Assisted customers with opening banking accounts, extracted and uploaded transaction records
- Organized real estate registration and loan files

Math Tutor, *De Anza College, Cupertino, CA* *Sept. 2016 – June 2017*
- Held training sessions and mini-lectures that contributed to an increase of 20% in student grades
- Provided 6 hours per week in tutor center to help students achieve understanding in calculus and statistics

Skills

Language: English, Chinese
Software and Programming Language: R, Python, Java, SQL, C++, Tableau, Excel VBA, Jupyter, Google Analytics
Data and Machine Learning Library:
- **Python:** NumPy, Pandas, Matplotlib, Seaborn, WordCloud, SciKit-Learn
- **R:** e1071, Caret, randomForest, glmnet, Survey, Mice, Naniar, ggplot2, dplyr