Ye Tao (Janson)

(530) -761-3846 | yetao@ucdavis.edu | https://yetao7977.netlify.com/| https://github.com/YeTao1996

Actively Seeking both full-time and intern opportunities to leverage the power of data science

EDUCATION

University of California, Davis **Davis, CA**

B.S. Applied Statistics, **3.61/4.0;** *B.A. Economics* in Data Analysis **3.67/4.0** Expected June 2020

- Relevant Course: Applied Machine Learning, Unsupervised Learning, Statistical Data Science, Time Series Analysis, Applied Linear Algebra, Mathematics Statistics, Statistics Data Technologies, Advanced Statistical Computing, Financial Economics, Econometric,

PROGRAMMING SKILLS

Python (Pandas, Numpy, Scikit-learn, Tensorflow, PyTorch, Flask), R (Tidyverse, ggplot2), SQL(MySQL), Tableau, Jupyterlab,

WORKING EXPERIENCE

PricewaterhouseCoopers **Shanghai, China**

Data Science Intern *Jun 2019 – Sep 2019*

- Modelling with Python based on word embedding and transfer learning to analyze customer feedback sentiment; performed data tidying on **40K+** visa customer comment data; achieved **71%** in model accuracy.
- Build extractive text summarizer using NLP techniques, such as **TD-IDF, TextRank** in python.
- Conducted fraud detection for **300K+** credit card transactions by building machine learning models including Logistic Regression; increased accuracy by **23%** through optimization.
- Initialized reorganization and migration of client's data into **SQL** database; matched remained data in python by measuring string similarity (**Cosine & Jaccard similarity**); further performed features extraction and modelling with **naïve Bayes algorithm.**

Entropy Capital LTD **Shanghai, China**

Intern *Jun 2018 – Aug 2018*

- Forecasted future stock price by **ARIMA** model. Implemented **factor selection** through Backward Selection, Forward Selection with **AICc, BIC criteria**.
- Computed a trading strategy based on moving average and weighted moving average model to predict returns of stock in R.

University of California **Davis, CA**

Student Researcher *July 2017 – Aug 2017*

- Collaborated with professors and studied several topics including population, poverty; calculated Gini index for measuring wealth distribution in ancient Japanese society.
- Created **SQL** database of family trees, as well as writing comprehensive memos and updated reports.
- Analyzed the effect of wealth inequality in ancient Japanese society through **hypothesis testing** in R.

PROJECTS

Credit Card Fraud Detection *Oct 2019 – Dec 2019*

- Performed under-sampling and over-sampling on imbalanced credit card transactions data.
- Constructed classification models for fraud detection using various methods tree-based models (**XGBoost, Random Forest, Decision Tree**) and **SVM**.
- Presented model performance with data visualization techniques(**ggplot2**) to non-technique audiences.

Global Temperature Prediction *Jan 2018 – Feb 2018*

- Researched on dependent variable of Temperature, with related factors. Performed a complete analysis with linear, non-linear regression, box transformation, residual analysis etc.
- Model selection with forward inclusion and stepwise backward elimination with **BIC** and AIC criteria.

INTEREST AND EXTRACURRICUALR ACTIVITIES

- Languages: Chinese, English, Japanese
- Top50 Go semi-professional player with 4 Dan