



linsam



data engineer 、 data scientist • 0972724528 • 台灣
• samlin266118@gmail.com

Experience with data mining, machine learning, distributed system, and web crawling. Hopes to focus more on data science and data engineer in future career.

Work Experience

- **SinoPac Holdings - Data Engineer · Nov. 2019 年 11 月 - now**
 - Develop Api for stock/option/future place orde.
- **Tripresso - Data Engineer · Oct. 2018 年 10 月 - Nov. 2019 年 10月**
 - Analysis travel data and build a machine learning model. Estimating increase 4% orders (revenue).
 - Maintain and develop an ETL distributed queuing system with 20 machines.
 - Optimize the ETL system reduced more than 50% execution time.
 - Develop new product crawler let product volume increase 1.5%.
 - Making analysis charts provide for other departments.
- **Mandatory Military Service · Oct. 2017 - Oct. 2018**
- **NDHU - RA · Mar. 2016 - Aug. 2017**

Analysing G7 financial data. Model validation and parameter estimation by regression models (SUR, MLE, Bootstrapping). And comparing single equation estimators and confidence interval with system equation.
- **NDHU - TA · Sep. 2015 - Jul. 2017**

Calculus, Linear Algebra, Statistics.

Projects

FinMind Open data Api

Open source financial data, more than 50 dataset, provide Api. Automatic update daily by docker swarm, distributed queue system rabbitmq and celery (8 cloud machines).
800 stars on github.

Bosch Production Line Performance - Kaggle Post-competition analysis, top 6% rank.

Highly imbalance data, ratio is 1000 : 1, 10 GB dataset size. And the data is 50% missing value. More than 4000 variables, but I build models by only 50 features.

Rossmann Store Sales - Kaggle

Post-competition analysis, top 10% rank.
Time series problem. Building models predict sales after 48 days.

Grupo Bimbo Inventory Demand - Kaggle

Post-competition analysis, top 8% rank.
Time series problem, eighty millions data size. Building models predict inventory demand after 2 weeks.

Instacart Market Basket Analysis - Kaggle

Real competition, top 25% rank.
Predicting which products will an consumer purchase again.

Verification code to text

Create python package of Taiwan Train Verification Code to text.
The model is made by keras-CNN.

Skills

Data Mining

Python - numpy, pandas, sklearn, multiprocessing, joblib.
R - parallel, dplyr, data.table, mice.

Machine Learning

Python - xgboost-gpu.
R - xgboost, svm, random forest, knn.

Deep Learning

Python - keras-CNN.

Statistical Model

R - GLM, GLMNET, NLS, SUR, MLE.

Web Crawling

1. Python - request, BeautifulSoup, lxml, selenium.

Rabbitmq & Celery

1. Build works on 8 Linodes (Cloud) Distributed queue system for Web Crawling.
2. install worker documents
3. supervisor
4. Flower

Create Python Package

1. [FinMind](#) 800 stars

WEB and Visualization

<https://finmindtrade.com/>
(by django, flask, nginx, uwsgi, redis)

Api

1. Load FinMind data by api.
2. Using python and fastapi develop api.
3. Monitor user request log.

2. [TaiwanTrainVerificationCode2tex](#)
t 60 stars

Docker

1. Using docker swarm manage
8 cloud machines for
distributed queue system \` api.

Education

- **National Dong Hwa University, Master of Science, Sep. 2017.**
Major : Mathematics and Statistics.
- **Tamkang University. Bachelor of Science, Sep. 2015.**
Major : Mathematics

Languages

R, Python. Basic in English and proficient in Chinese.

Powered by **CakeResume**