林子軒 Lin Sam · ✉ samlin266118@gmail.com · ☎ 0972724528 · **in** sam lin · ⊙ linsamtw ·

Summary

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

Work Experience

Tripresso - Data Engineer   Oct. 2018 - Now

- Analysis travel data and build a machine learning model. Estimating increase 3% 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.

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.

Projects

FinMind python package   **280 stars on github.**

- Open source of financial data, more than 40 data, more than 100 million stock trade data.

- Develop Api with python and flask.

- Automatic update daily by distributed queue system, rabbitmq and celery (8 cloud machines).

- Develop web by django and visualize data by dash (http://finmind.servebeer.com/).

TaiwanTrainVerificationCode2text python package   **60 stars on github.**

- The verification code consists of 26 English and 10 digits, which may be 5 or 6 yards.

- Simulate generating 100,000 images and building CNN models to fit and predict.

CIFAR-10 (Kaggle)   **87% classification accuracy.**

- Using pre-trained VGG and ResNet models classify objects.

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.

- The goal is to predict which parts will fail quality control.

Grupo Bimbo Inventory Demand (Kaggle)   **Post-competition analysis, top 8% rank.**
Rossmann Store Sales (Kaggle)   **Post-competition analysis, top 10% rank.**

- Time series problem, eighty millions data size. Building models predict inventory demand after 2 weeks and predict sales after 48 days.

Skills

- Experience with **data mining** and **feature engineering** by Python and R.

- Experience with **machine learning** by Python (**xgboost-gpu**) and R (**xgboost, svm, random forest, knn**).

- Experience with **deep learning** by Python (**kears-CNN, LSTM, GRU**).

- Experience with **distributed queue system** web crawling by Python (Cloud **Linode**).

- Experience with **API** development..

Education

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