Ching-Chuan (Jamal) Chen

Data Scientist / Data Engineer

© CONTACT

C	+886-966-676-326	
\geq	zw12356@gmail.com	
in	www.linkedin.com/in/celestial0230	
0	github.com/ChingChuan-Chen	

SKILLS

R / MatLab	Master
Statistics	Advanced
Statistical Learning	Advanced
SQL / Python	High-Intermediate
LaTeX / Bash	Intermediate
C++ / C#	Basic
M LANGUAGES	
○ Chinese	Native speaker
● English	Advanced
⊘ Japanese	Intermediate (JLPT N3)

REFERENCES

Jeng-Min Chiou Research Fellow Institute of Statistical Science Academia Sinica +886-2-2783-5611 ext. 312

Sheng-Mao Chang Associate Professor Department of Statistics National Cheng Kung University +886-6-275-7575 ext. 53632 smchang@mail.ncku.edu.tw

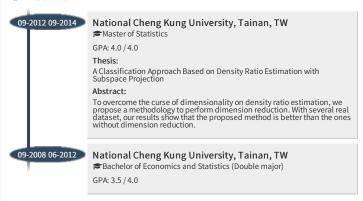
SUMMARY

I am a data scientist and sometimes work as data engineer. I would say I am enthusiastic, dedicated and analytic. I enjoy helping people to solve the difficulties they encountered.

am also

- ☑ a person enjoying sharing knowledge with people.
- \blacksquare a statistician works deeply with data visualization, statistical methodologies and statistical learnings.
- a engineer with creativity, critical observation and leadership.
- a experienced programmer skilled at R, Python, Shell, MATLAB, Scala and SQL.
- a skilled data engineer in data streaming / ETL, distributed computing and distributed database.

EDUCATION



JOURNALS

milr: Multiple-Instance Logistic Regression with Lasso Penalty

Ping-Yang Chen, Ching-Chuan Chen, Chun-Hao Yang, Sheng-Mao Chang and Kuo-Jung Lee *The R Journal* (2017) 9:1, pages 446-457.

https://journal.r-project.org/archive/2017/RJ-2017-013/index.html

III WORK EXPERIENCES

Trend Micro Inc., Taipei, Taiwan

Senior Data Scientist, Consumer, 01-2019 - Present

Objective

Along with a group of data engineers and domain experts, develop an IPS on network flows via statistical learning.

Projects

- 1. Network behavior analysis / data scientist
 - ★According to device, summarize the network flows to profile device behaviors.
 - ★Define a good score and threshold for device profiling with statistical sense.

Taiwan Semiconductor Manufacturing Company, Taichung, Taiwan Senior Data Scientist / Engineer CIM Department, 09-2018 - 01-2019

Senior Data Scientist / Engineer, CIM Department, 09-2018 - 01-201 Data Scientist / Engineer, CIM Department, 07-2016 - 08-2018

Objective

Develop automation systems on quality control during wafer processing from a big volume of data (3 billions per day).

Projects

- 1. WAT chart change detection / data engineer / data scientist
 - ★WAT is wafer acceptance test which is examined while finishing the process of a wafer. There is no detailed orders in data. It only contains date information.
 - $\bigstar Proposed \ an \ algorithm \ to \ detect \ the \ daily \ changes \ based \ on \ statistics. \ It \ is \ effective \ to \ detect \ the \ changes \ between \ upper \ control \ limit \ and \ lower \ control \ limit.$
 - ★Parallelly processed 3 billions records of data and output results of detections in R language and MPI.
- 2. Control chart change detection performance improvement / data engineer
 - $\bigstar \text{Proposed a new architecture powered by MPI to improve speed of detection algorithm.}$
 - $\bigstar \text{It}$ reduced the implementation time from 8 hours to 40 minutes in the new architecture.
 - $\bigstar Proposed$ an algorithm to dispatch the detection jobs with different running time which depends on data size.
- 3. Build up a development environment for the data scientists / data engineer
 - $\bigstar \textsc{Construct}$ a consistent and centralized controled development environment for data scientist.
 - $\bigstar \text{Writed}$ a customized RStudio server Dockerfile to ensure everyone get the same environment.
- 4. Data pipeline for processing history data and measurements data / organizer / data engineer
 - ★Propose a fast and reliable data pipeline powered by Spark in Scala and Python. (UDAF is written in Scala.)

- ★Any new ETL can be set flexibly and easily for users. This UI is done by R shiny.
- ★Well monitoring for job implementation and data quality.
- ★Stored 6 billions data into Hive with full automation and good data quality.
- 5. Propose, validate and construct a big data solution / organizer / data engineer
 - ★For the messy data query requirement for data analysts, I tested several solutions with SQL-like query language to test.
 - ★Tested several big data solution like Cassandra, Drill and Hive.
 - $\bigstar \text{Made}$ number of machine learning jobs can be done in 20 times shorter computing time than Oracle database.

Academia Sinica, Taipei, Taiwan Research Assistant, Institute of Statistical Science, 09-2015 - 06-2016

Objective

Complete at least one research in the field of functional data analysis.

Projects

- 1. Imputation of functional data / data scientist
 - ★Used functional clustering to impute missing values in traffic data with lower RMSE
- Create a data streaming for researches (data from Taiwan freeway bureau) / organizer / data engineer
 - ★Built a data pipeline to transform crawled data from XML to JSON and store them into a MongoDB.
 - ★Developed a platform to view the data with d3.js via R shiny.
- 3. Travel Time Estimation / data scientist
 - ★Studied journals about travel time estimation.
 - ★Realized the algorithms in journals and summarize the pros and cons.
- 4. Organize and refactor the source codes of previous researches / organizer
 - ★Studied the previous researches and learn how FPCA works.
 - ★Removed several redundant blocks and improve performance of key functions.

AWARDS

12-2017

TSMC Kaggle Competition for the Defect Recognition

Third Place

An internal competition in TSMC. There are over 100 teams to assist wafer factory decrease cost on the categorization of defects. There are only 3000 defect/reference images provided. The goal is do our best to get high accuracy rate on testing set (1200 images.), I used a 6-layer convolution neural network with two Xception modules and win third place in 91.2% accuracy rate.

08-2014

Competition for Data Analysis with R in Taiwan

PHonourable Mention

A national competition in Taiwan. There are over 30 teams to do a brainstorming on the data from a system to register the actual selling price of real estate. Each team have one day to come out a topic and apply R language to complete and demonstrate the results. Our team chose to predict the price of house from the messy data via LASSO approach.