

# Sun Weiran | 孙蔚然

## Data Scientist

Mobile: +65 97913684 | Email: [sunwrn@gmail.com](mailto:sunwrn@gmail.com)

Address: 1 Moreton Close 556476

Chinese Citizen, Singapore Permanent Resident

GitHub: <https://github.com/ianforme>

Blog: <https://medium.com/@sunwrn>

LinkedIn: <https://www.linkedin.com/in/sun-weiran/>

## EDUCATION

### National University of Singapore

Aug 14 – Jun 18

- Major in Business Analytics (BSc, Honours Distinction)
- Dean's Lister
- Certificate of Distinction in Applied Analytics Knowledge Area
- Certificate of Distinction in Data Science Analytics Knowledge Area

## WORK EXPERIENCE

### Data Scientist – Advanced Analytics Team, NCS Group

Aug 18 – Now

- Delivered project for Changi Airport to develop a real time predictive platform for taxi operators
  - Designed the overall solution architecture
  - Integrated data from different systems including flight control, taxi and passenger sensors
  - Engineered end-to-end data pipeline from data extraction to near-real time prediction
  - Built the most fundamental models (random forest) to predict for future taxi demands for each terminal. These models have 80% accuracy
- Delivered project for Changi Airport to predict for number of passengers on incoming flights
  - Analyzed 500k+ records from more than 1,000 flights in past 2.5 years
  - Built customized machine learning models (xgboost) for different types of airplanes. The model improved accuracy from the existing solution by 20%
  - Designed and built Tableau dashboards for airport staff to facilitate their operation planning
  - Proposed improvements for the next phase to operationalize this project
- Delivered project for Land Transport Authority to predict for traffic speed and volume
  - Engineered and maintained data pipelines for data extraction, model training and prediction
  - Analyze ERP traffic data and built models (linear regression, random forest) for different roads, vehicles and date types
  - Designed and built Tableau dashboards for tariff planners
- Volunteered to join the recruitment committee to enhance the hiring process of the team

### Data Engineer Intern – Global Data Innovation Centre, Dentsu Aegis Network

Jan 17 – Jun 17

- Scrapped advertisement slogans using Selenium and trained LSTM models for generating advertisement slogans from user input keywords
- Built interactive chatbot prototype using Python. Input questions are processed (text normalization, part-of-speech tagging, name entity recognition) and the most relevant answers are searched from a knowledge graph and ranked by relevancy
- Built predictive models (random forest and logistic regression) to predict customer conversion rate based on their online browsing behaviors
- Improved web application computation efficiency by 600% by implementing multi-threading

### Data Analyst Intern – ORT Solutions

May 16 – Jul 16

- Performed ETL on large datasets in terabytes using Python and Google Big Query
- Analyzed client campaign data collected from WiFi devices and social network. Conducted network analysis over thousands of participants using Gephi
- Initiated customer profile analysis using WiFi data for shopping malls. Different groups of customers, i.e. frequent vs non-frequent, lunch time vs mid night shoppers were identified from their visiting patterns
- Enhanced operation efficiency by automating routine paperwork using Python which reduced report generation time from 15 minutes to 2 minutes

## PERSONAL PROJECTS

---

### Contributing Writer – Towards Data Science

Aug 19 – Now

- “Hyper Parameter Tuning with Randomised Grid Search” (curated by editors)
- “Why is Boosting Fitting Residual”

### Codimage - Python Library for Code Encryption

Aug 19

<https://pypi.org/project/codimage/>

- Codimage is a public Python library available through pip installation, that encrypts text files, i.e. Python scripts, into a picture's pixels without changing its original look and protects sensitive information with passwords

### WeChatBot - Q&A Assistant Chatbot on WeChat

May 19 – Jul 19

<https://github.com/ianforme/wechatrobot/>

- WeChatBot is built on wxpy framework and listens to WeChat conversations. This chatbot can be hosted on any WeChat accounts. Once certain keywords or text patterns are detected, pre-built functions will be triggered to respond to users
- The chatbot is able to
  - retrieve weather forecast, news and bus arrival time from various publicly available APIs
  - scrape job openings from JobStreet.com using Selenium, process raw text into structured format, create PDF report and email to user

## Certificates

---

### Big Data Specialization – MicroMaster, University of California San Diego

Oct 19

- Introduction to Big Data
- Big Data Modeling and Management Systems
- Big Data Integration and Processing
- Machine Learning with Big Data
- Graph Analytics for Big Data
- Capstone Project

## Skills

---

- Programming Languages: Python, R, Java
- Databases: Oracle SQL, Neo4j
- Visualization: Tableau, HTML, CSS
- Version Control: Git, SVN
- Toolkits & Packages: Numpy, Pandas, Scikit-learn, Statsmodels, Xgboost, Keras, Selenium, Matplotlib, Seaborn, Yellowbrick

## Languages

---

- Chinese: native, fluent in both spoken and writing
- English: fluent in both spoken and writing
- Japanese: beginner

## Interests

---

- Chinese debate: former team captain of NUS Chinese Debate Team. Won multiple international championships during college time
- Game: huge fan of Dota 2 and JRPG games
- Coach: coaching Anderson Junior College Chinese debate team