

Yuan Gao

y1gao@ucsd.edu | www.github.com/Rabona17 | (858)-263-8426

EDUCATION

UNIVERSITY OF CALIFORNIA, SAN DIEGO 2018.9-PRESENT

– BACHELOR OF SCIENCE IN DATA SCIENCE

– MAJOR GPA:4.0/4.0

UNIVERSITY OF CALIFORNIA, BERKELEY 2017.6-2017.8

– GPA:3.37/4.0

SKILLS¹

TECH-SKILLS

- **PROGRAMMING LANGUAGES:** PYTHON, JAVA, HTML/CSS, PHP, MATLAB, JAVASCRIPT
- **DATABASE LANGUAGES:** MYSQL
- **MODULES:** NUMPY, MATPLOTLIB, SCIKIT-LEARN, PANDAS, BEAUTIFULSOUP, SELENIUM, TENSORFLOW
- **SOFTWARES:** TABLEAU, GOOGLESHEET, MICROSOFT OFFICE
- **WEB TECHNOLOGY:** JAVA SERVLET PAGES(JSP)

SOFT-SKILLS

- GOOD AT **QUICK LEARNING, KEEPING HUNGRY FOR NEW TECHNOLOGIES, WRITING AND COMMUNICATING**
- COMFORTABLE WITH **WORKING IN A TEAM AND SPEAKING IN FRONT OF CROWD**

PROJECTS²

DATA WRANGLING&CLEANING ON MASSIVE REAL WORLD DATA 2019.3

UC San Diego Data&GIS Lab's Data Wrangling Competition

- Cleaned a real-world messy weather dataset with over **10 million rows** of data and with **20% missing/unformatted data**
- Filled missing value using **Scikit-learn**, Pandas, and Matplotlib by implementing my own cleaner class and functions
- Utilized **Supervised Learning algorithms** (Linear regression & Support Vector Regression) to predict precipitation
- Minimized my code complexity by choosing different data structures
- **Won this competition as the only winner**

STATISTICALLY TESTING THE EXISTENCE OF SOCIAL MEDIA BOTS 2019.1

UC San Diego COGS9: Introduction to data science.

- Analyzed the posts of a celebrity suspected to use bots on internet to add popularity
- Web-scraped **more than 5000** posts from Weibo, the biggest social media with **the strongest anti-scraping technology**
- Employed **A/B test** to check whether the test statistic is statistically significant
- Utilized **NLP packages** in Python(NLTK and Jieba) to find evidence for the existence of social media bots

VISUALIZING FACEBOOK INFORMATION OF HDSI 2018.12

UC San Diego Halicioglu Data Science Institute Student Shoutout

- Web-scraped **all posts (more than a year)** from HDSI Facebook account
- Made a world cloud **out of thousands of words** to show HDSI's working emphasis
- Posted by HDSI on its **official website**³
- The visualized world-cloud is used by HDSI as a pattern on the department's shirt

DATA ANALYSIS ON PUBG PLAYERS THROUGH WEB-SCRAPING 2018.10

- Developed an algorithm to web-scrape **over 2000** game data of a twitch streamer 'Shroud' from PUBG website
- Makes analysis and visualization using Pandas and Matplotlib on the scraped file with **over 15 columns and 1600 rows**
- Summarized all his gameplay and found pattern indicating the change in his stream content

OBJECT DETECTION USING ML ALGORITHM 2018.7.20-2018.7.22

2018 HACK.INIT HACKATHON, Shanghai

- Modified and applied an existing Tensorflow Machine-learning algorithm for detecting human's mouth (Object-detection API)
- Implemented this feature to drinking fountain which helped it accurately eject water to human's mouth
- Scraped and trained **over 400** pictures

HONORS

WINNER OF UCSD DATA WRANGLING CONTEST 2019.3

UC San Diego Data&GIS Lab

UCSD PROVOST HONOR 2019.1

UC San Diego Eleanor Roosevelt college

EXPERIENCE

WEB DEVELOPMENT INTERNSHIP 2017.10-2018.1

Nanchang University, China

- Followed graduate IT students to develop a web application for a local hospital
- Implemented Java and Mysql code for JSP technology (JDBC involved)
- Learned basic syntax and application of HTML/CSS

¹ **Relative Coursework:** Principles of Data Science & Intro to OOP: Java & Data Structure and Algorithms for Python, Java and C/C++ & Linear Algebra & Vector Calculus.

² **More detailed description:** On my Github www.github.com/Rabona17

³ **HDSI Official Website:** http://datascience.ucsd.edu/news-and-events/news/Data-Science-Freshman_Word_Cloud.html