

PAUL CHEN

463-210-6665 | chen3876@purdue.edu | [linkedin.com/in/paiweichen/](https://www.linkedin.com/in/paiweichen/) | <https://paul46186.github.io/portfolio/>

PROFILE AND SKILLS

Data enthusiast with proven technical qualities, analytical mindset, strong communication, and actionable insights

- **Technical Skills:** Deep Learning, NLP, Data Mining, Machine Learning, Python, SQL, Excel VBA, Tableau, R, Business Analytics, Integer Programming (Optimization), Data Visualization, Project Management, HTML
- **Power Skills:** Leadership, Active Listener, Change Management, Agile, Self-Learner, Empathy, Presentation
- **Language:** Mandarin Chinese (Native), English (Professional), Taiwanese (Native), German (Beginner)

CERTIFICATIONS AND AWARDS

- Data Camp Data Analyst with SQL Server Track
- Data Camp Data Scientist with Python Track
- Data Camp Data Scientist with R Track
- Codecademy Deep Learning Tensor Flow Path
- Google Data Analytics Certificate
- Google Digital Marketing Certificate
- 3rd place in 2021 KGSA Case Competition
- 3rd place in STAMINA Analytics Case Competition

PROJECTS

Industry Practicum – SIL Optimizing Seed Data for Machine Translation Jan 2022 – Present

- Established an architecture that optimizes of datasets for the training of machine translation models via **JoeyNMT**, and a clear strategy for sequencing a translation project that minimize the human translation cost and time. This method can be used for translating between different languages especially practical for rare language that requires human translation

Prescriptive Analytics – Selected by Pycon Conference 2022 Poster Session Dec 2021 – Present

- Integrated a trained predictive model using the python library **PyCaret** into an optimization model using **Pyomo** and provided end users decision confidence. This project is selected by **Pycon** poster session which is the biggest python conference in the US

Kaggle in-class competition - Credit Card Prediction Project Nov 2021 – Dec 2021

- Created and tuned an ensemble model that combined neural network, adaptive lasso and random forest through **SAS EM** to predict the likeliness of credit default and won 10th place in the Kaggle in-class competition

Improving the Craigslist User Experience - Video Gaming Classification Nov 2021 – Dec 2021

- Scrapped, tokenized, and lemmatized the Craigslist video game data via **beautifulsoup** and **nlTK**, then built 6 models with **keras** and **sklearn**. The best model has 88.45% accuracy which can help increase information gain by 34% that will improve the retention rate and customer experience on Craigslist website

Heart Disease Classification Project Aug 2021 – Aug 2021

- Created a decision-support system (DSS) application that performs two roles via R Shiny. One provides interactive exploratory data analysis by visualizing the dataset and showing the statistic results, while the other facilitates speedy risk factor identification by allowing users to input 14 body variables and forecast their prospective heart failure levels using random forest model

Special Topic Group Study Feb 2019 – Feb 2020

- Led a group to define the assembly line problem, aggregate and cleaned datasets with **Excel VBA**, and construct 3 models with **Python** and **Gurobi** that allow manager to find optimized workstation solutions within 1 minute
- This project allows manager can make optimized machine layout decision without relying on human

PROFESSIONAL EXPERIENCE

Shoetown Group Qingyuan, China Operation Intern Jul 2019 – Aug 2019

- Reduced workers' daily input time from 6 hours to 1 minute by using **Excel VBA**
- Built an **OEE** system to enable the company to capture their machine performance

EDUCATION

Purdue University, Krannert School of Management West Lafayette, IN Master of Science in Business Analytics and Information Management Jun 2021 – May 2022

- GPA: 3.59/4.0

National Cheng Kung University Tainan, Taiwan Bachelor of Management Science, Industrial and Information Management Sep 2016 – Jun 2020