

PAUL CHEN

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

PROFILE

Data enthusiast with experience in the manufacturing industry having a passion for deep learning and optimization

- **Programming Languages:** Python, BigQuery, MySQL, Excel VBA, Tableau, R, SAS, Power BI, TensorFlow
- **Technical Skills:** Deep Learning, Natural Language Processing, Machine Learning, Model Validation, Time Series, Project Management, Data Visualization, Prescriptive Analytics (Mix-integer Programming), Predictive Analytics

EDUCATION

Purdue University, Krannert School of Management

West Lafayette, IN

Masters of Science in Business Analytics (STEM) - GPA(3.59/4.0)

Jun 2021 – May 2022

- **Academic Honors**

- **Relevant Courses:** Data Mining, Analyzing Text and Image, Computational Complexity, Database Management, Customer Analytics, Spreadsheet Modeling & Simulation, Prescriptive Analytics, Advance Business Analytics

National Cheng Kung University

Tainan, Taiwan

Bachelor in Industrial and Information Management

Sep 2016 – Jun 2020

PROFESSIONAL EXPERIENCE

SIL International – Capstone (Linguistic NGO Affiliated with United Nations)

West Lafayette, IN

Data Scientist Intern

Jan 2022 – Present

- Reduced monetary cost and training time for a translation model by scripting an algorithm to optimize training datasets and provide a clear strategy for selecting semantic domains (translation content) via Python
- This method is effective for translating rare languages that generally require human translation

Shoetown Group (Shoe Manufacturer for Nike)

Qingyuan, China

Data Analyst Intern

Feb 2019 – Feb 2020

- Automated an inventory system to reduce manual data input time by 3600% through Excel VBA
- Implemented a machine performance system to enable the company to monitor 3D printer running times

Research Team Lead for Assembly-Line Problem

- Led a group of students in defining an optimization problem by coordinating with stakeholders, assigning tasks, setting timelines for deliverables, and presenting to professors with an oral defense
- Provided Nike manufacturer management an optimized model for workstation layout computable within 1 minute to minimize assembly costs by using robust optimization models in Python and Gurobi

AWARDS

2022 Kaggle Crossroads Classic Datathon (Team Leader)

- Built/trained an AI with supervised learning techniques to detect phishing emails
- Integrated multiple data sources to create an ensemble model (SVC + XGBoost) with a 98.53% AUC score
- Won 3rd place from a pool of 47 teams in the Kaggle leaderboard

STAMINA Analytics Case Competition (Team Leader)

- Pitched an analysis for finding ideal piloting locations and potential revenue streams for a drone company
- Won 3rd place out of 50+ teams

SAS Optimization Challenge – Water Usage Contract (Team Leader)

- Constructed a nonlinear optimization model to determine which water contracts a corporation should adopt
- Won 1st place out of 16 graduate division teams

PROJECTS AND LEADERSHIP EXPERIENCE

Window Breakage – Selected by Pycon US Conference 2022 Poster Session

- Calculated window breakage rates and probabilities and integrated model values into an optimization program to compute optimal variable values for cutting speed, window size, batch size, and other relevant metrics

Video Gaming Classification - Improving Craigslist's Key Word Searching

- Scrapped, tokenized, and lemmatized Craigslist video game data by building Python models (Random Forest, Neural Network, Naive Bayes, Logistic, RNN, SVM) utilizing the BeautifulSoup, Keras, sklearn, and nltk packages and selecting the best performer to increase information gain by 34% with an 88.45% accuracy rate

Graduate Teacher Assistant – Python Programming and Management Information System

- Collaborated with professors to assess student assignments and exams and answered questions for 80+ students

Heart Disease Classification Project

- Created a Rshiny app to provide an interactive dashboard to visualize patient data and predict the likelihood of heart failure by allowing users to input 14 body variables to help doctors facilitate risk factor identification