



Skills

- **Python** (Pandas, NumPy, SciPy, sklearn, PySpark, Beautiful Soup, Requests, NLTK, gensim)
- **Excel** (Conditional Formatting, Data Simulation, Pivot Tables)
- **R** programming (R Markdown, ggplot2, dplyr, glmnet)
- Microsoft Azure (DataBricks), **MySQL**, PHP
- Jupyter Notebook, Google Colab, Git

Projects

[NATURE LANGUAGE PROCESSING FOR FINTECH LITERATURE](#) – Personal Project – Nottingham, UK July-Sep 2023

- Construct a fintech-related scientific articles database and extract plain texts from PDFs with **Python**
- Perform and compare several word embedding methods to fintech literature with **Python** NLTK toolkit

[IMBALANCING DATA PREPROCESSING IN BIGDATA APPROACHES](#) – Group Project – Nottingham, UK April 2023

- Utilized **PySpark** and **DataBricks** to fix incorrect formatting, remove incomplete data and create secondary features
- Implemented nearest neighbor based algorithms (**SMOTE** and **ENN**) to balance data quantities in a **PySpark** approach
- Evaluated the preprocess with prediction accuracy metrics of several benchmark classification models

[TRAFFIC INCIDENT MANAGEMENT SYSTEM](#) – Personal Project – Nottingham, UK Dec 2022

- Created a **web UI** to register, edit and query for traffic incidence, create officer accounts and manage audit trails
- Utilized **PHP** to access to **MySQL** database to add, edit and query traffic incidence records

CALCIFICATION CORONARY ARTERIES PREDICTION – collaboration project – Taiwan Nov 2021 - July 2022

- Performed missing data imputation, text data embedding and data standardization with **Excel** and **Python**
- Evaluated models and analyzed the performance in prediction accuracies, model complexity and time-consuming
- Collaborated with Changhua Christian Hospital by meeting once a month

[LIVER CANCER RECURRENCE AND SURVIVAL PREDICTION](#) – collaboration project – Taiwan Jul 2019 – Jun 2021

- Data preprocessing, model training and performance evaluation as the above project
- Extracted features from medical images with **MATLAB**
- Published a journal paper and collaborated with Taipei Veterans General Hospital by meeting once a month

Work Experience

DATA ENGINEERING RESEARCH ASSOCIATE – Intelligent Computational Lab, NCTU – Taiwan 2019-2022

- Constructed prediction model for medical prognosis with numerical or image data and **communicated with leading hospitals**
- Imported domain knowledge and applied **cohort study**, **hypothesis test** and independent test to diverse data
- Built **automatic data processing** and report generation scripts which reduced manual processing time by 90%
- Practically applied the Liver Cancer Treatment Decision Support System to the gastrointestinal surgery clinic in VGH Taipei

Publication

Lee, I., Huang, J. Y., **Chen, T. C.**, Yen, C. H., Chiu, N. C., Hwang, H. E., ... & Huang, Y. H. (2021). [Evolutionary learning-derived clinical-radiomic models for predicting early recurrence of hepatocellular carcinoma after resection](#). Liver Cancer, 10(6), 572-582.

Education

MASTER OF SCIENCE IN DATA SCIENCE – University of Nottingham –Nottingham, UK September 2023

Majors: Data Analysis, Natural Language Processing and Big Data Processing

MASTER OF SCIENCE IN BIOINFORMATICS – National Yang Ming Chiao Tung University – Taiwan November 2021

Majors: Machine Learning, Model Optimization and Data Mining