Hao (Philo) Wu

Bachelor of Commerce / Bachelor of Adv. Studies (Honours)
in Business Analytics and Finance
University of Sydney

+61-432 190 188 | +86-150 1093 2868
hawu9812@uni.sydney.edu.au
philo_wu@outlook.com
GitHub: CodeBoyPhilo
LinkedIn: Hao Wu

Education

Degree	Institute	Score	Year
B.Comm / B. Adv.Studies (Hons)	University of Sydney	WAM 85	2021-Present
High School	Knox Grammar School	ATAR 96.95%	2018-2020

Work Experience

- **PingAn Securities Co., Ltd.,** ***December 2023 - February 2024***
  Machine Learning Engineer (Intern) **Shenzhen, China**
  - Participated in the Smart Stock AI Project, **refined previously built models**, collaborated with cross-functional teams to deploy the models.
  - **Researched and examined popular LLMs** including ChatGLM3-6B, Mistral-7B, etc., facilitated in the **supervised fine-tuning stage of LLMs** to specialise in downstream tasks under supervisions.
  - Individually **developed a censoring model** to detect inappropriate prompts and prompt injection attacks. Overcame the problem of data insufficiency, formulated innovative solution with the use of generative AI and sophisticated prompt engineering techniques. Fine-tuned a mDeBERTa-V3-base model to include in the system, achieved a final **predictive accuracy over 99%** on test set. The system is deployed and connected to all internal LLMs of the company, and received **0 misdetection reports over 5,000 requests** sent from the users to the agents.
- **PingAn Securities Co., Ltd.,** ***June 2023 - September 2023***
  Machine Learning Engineer (Intern) **Shenzhen, China**
  - Participated in the Smart Stock AI Project and overseeing technical aspects and guiding the formulation of effective modelling solutions, collaborated with cross-functional teams, facilitating seamless information exchange and project management.
  - Conducted appropriate data cleaning and feature engineering on the provided stock data, processed the financial and trading data of stocks into meaningful indicators, finally balanced models prediction accuracy and interpretability based on a combination of experimental results and literatures findings to determine the most appropriate modelling solution for the project.
  - Independently **developed machine learning algorithms** to best suit the projects needs, results include: 1. **A semi-supervised model** to precisely detect low-risk stocks in the A-Shares market. After using stock data from different time periods to validate, results show a **97% accuracy** on average. 2. **An ensemble classification model**, achieved **5% increase** in accuracy in validation set compared to traditional tuned StackingClassifier and VotingClassifier. Both models were highly recognised by manager, who especially praised the latter for its innovation, feasibility, and ease in future deployment and maintenance.

Research Experience

- **Developing Machine Learning solution for Big W** ***August 2023 - November 2023***
  Capstone Project
  - Collaborated with the client *Big W*, a subsidiary of Woolworths Group (top 3 retail company in Australia)
  - Individually **conducted Exloratory Data Analysis** to uncover unique patterns in Big W's sales data, composed findings in a 15 pages long report. The reviewer especially recognised the variety of data visualisation techniques in the report, and highlighted the business messages conveyed from statistical analysis. The report achieved a score of **86 (HD Level)**.
  - Led a team of 5 to develop unique machine learning solution and formulate advise for Big W with the aim to boost sales. Conducted comprehensive research to include additional predictors, **lowered RMSE by 12%** compared to use the original predictors. Individually **proposed a two-stage ensemble model named Body or Tail Ensemble Learning' (BoTEL)**, **lowered RMSE by over $1000** using Linear Regression as the base model solely. Based on the novelty of the model, the team became **one of the only two teams** in the cohort that had the opportunity to present their solution to the managements of Big W.
- **Predictive Model Development for Airbnb in Sydney Rental Market** **March 2023 - April 2023**
  Advisor: Dr. Marcel Scharth, University of Sydney
  - Led a team of three to conduct data mining and modelling on dataset provided by Airbnb, developed predictive models based on the listing information of the property and hosts, reported the determining factors that affects the listings price.
  - Conducted proficient **feature engineering** with **dimension-reduction techniques** such as PCA, conducted attribution analysis for the differences in various model classes performance, **compared the performance** of Tree-based models (e.g. Decision Tree, Random Forest, GBDT), linear models (e.g. Linear Regression, GAM), Neural Networks (MLP), and Ensemble Models on regression-type problem
- **Predicting Total Passenger Arrivals for Public Transport in Queensland** **May 2022 - June 2022**
  Advisor: Dr. Pablo Montero-Manso, University of Sydney
  - Analysed **total passenger arrivals** at seven stations over the past three years and independently developed time series models, **formulated a methodology** for predicting total passenger arrivals in 2021.

– Conducted time series decomposition and feature engineering on the data, designed a Python algorithm for **automatic cross-validation** to identify the optimal time series model among candidates, and assessed 18 variants across three popular time series model families, including ARIMA, Holt-Winters, and more.
– Concluded that AR, MA, ARIMA, and Seasonal ARIMA models generally demonstrate superior performance in time series analysis compared to the Holt-Winters model family. The algorithm selected the ARIMA model family as the best predictive model in over 71% of cases, with an average difference in MAE between the two model families reaching 36%. The report projected a 23.1% increase in total passenger arrivals in 2021 (the actual increase was 28.9%).

- **U.S. Housing Price Predictions** **March 2022 - April 2022**
  Advisor: Dr. Pablo Montero-Manso, University of Sydney
  – Analysed a dataset comprising over 80 variables and 14,000 observations, independently developed predictive models to pinpoint the key determinants of housing prices.
  – Conducted exploratory data analysis on the dataset, optimizing features, hyperparameter k, and penalty weight through cross-validation and feature selection methodologies.
  – Identified the optimal predictive model among Linear Regression, KNN, and LASSO Regression model families, considering bias-variance trade-off and utilising Mean Absolute Error as the evaluation metric.
  – Concluded that LASSO Regression's superior performance over both Linear Regression and KNN model families in the test set in high-dimensional data, achieving a MAE of only $20,030.

- **A Statistical Analysis of The Determinants of Fuel Economy** **September 2021 - October 2021**
  Advisor: Professor Artem Prokhorov, University of Sydney
  – Led a team of three members to investigate key factors impacting fuel economy, taking into account variables such as start-stop technology,' vehicle class,' displacement,' fuel type,' and others, while formulating data-driven recommendations for improving fuel efficiency.
  – Utilised Python data analysis libraries, such as Dataprep, Scipy, StatsModel, Sklearn, and Seaborn, for data preprocessing, encoding, and visualization, independently constructed 12 Linear Regression models based on variable understanding, feature selection methods, and bias-variance trade-off to determine the optimal predictive model.
  – Achieved an adjusted R2 of 0.777 with the top-performing model after implemented dimensionality reduction for dummy variables, demonstrated a 9% enhancement in predictive and generalisation performance compared to similar Linear Regression models.

Independent Projects

- **ChatWithYou**
  Independent Project **Github: CodeBoyPhilo/ChatWithYou**
  – Instruction on how to DIY a personalised WeChat bot that talks like the user.
  – The bot is powered by a fine-tuned ChatGLM3-6B model using WeChat chat history data.

- **Phi's GRE Prep Toolkit**
  Independent Project **Github: CodeBoyPhilo/Phi-s-GRE-Prep-Toolkit**
  – A continuously updated repository containing self-developed Python based tools to facilitate the preparation of GRE test and any English exams for non-English speaking users.
  – Including two features, VocabCloud (computes a vocabulary cloud graph given any vocabulary bank) and QuestionOverfit (a CLI and Web based app that allows an efficient, immersive, and easy-to-use question practising and drilling experience). Both are integrated into a Python packaged called **toolkit**. Pre-release download is available on GitHub.

Volunteer

- **Sydney University Chinese Student Association (SUCSA)** **April 2022 - April 2023**
  Staff of the Department of Academic and Career Planning **Sydney, Australia**
  – Participated in planning and organizing **two New Student Orientation events** and **On-Campus Career Fair**, fostered the partnership between SUCSA and numerous renowned companies for recruitment, ultimately offering employment-related opportunities for students.
  – Managed event promotion on social media, in charge of new member recruiting each semester, and conducting interviews for Mentoring Program.

Certifications

- **Former Member of Dalyell Scholar program** University of Sydney
- **Life time member of Beta Gamma Sigma** Beta Gamma Sigma
- **2022 Dean's List of Excellence in Academic Performance** University of Sydney
- **2022 Academic Merit Prize** University of Sydney

Skills & Interests

- **Programming Languages**: Python, PostgreSQL, Cypher, Shell Script
- **Tools and Frameworks**: Jupyter, PyCharm, Neo4j, Docker, Vim, Git
- **Operating Systems**: Windows, Ubuntu, MacOS
- **Language**: Mandarin (Native), English (Fluent), Japanese (Beginner)