

Hao-Chun (Oscar) Shih

+1(734)-780-4256 / haochuns@umich.edu / Michigan, Ann Arbor / www.linkedin.com/in/oscar-hao-chun

ACADEMICS

University of Michigan, Ann Arbor

Aug 2024 - present

Master of Data Science

Highlight Coursework: Data Mining, Database Management System

National Chung Hsing University

Sept 2019 – June 2023

Bachelor of Science in Applied Mathematics

GPA: 3.92 Ranked 1st

Highlight Coursework:

Mathematical Statistics, Discrete Mathematics, Regression Analysis, Data Structure & Algorithm

Award: 2023 Graduating Class Representative Academic Excellence Award, 2022/2023 Dean's List

COMPUTER SKILLS

Python(PyTorch, TensorFlow, Scikit-learn)/ C++/ SQL/ PL SQL/ MySQL/ R/ Linux/ Latex/ Microsoft SSIS/ IBM DataStage/ Cognos

WORK EXPERIENCE

Research Assistant | (Linux, Python, Latex)

Oct 2023 – June 2024

Academia Sinica, Chinese Knowledge and Information Processing (CKIP) Lab

- Investigated memory capabilities on Llama2-7b with plug-in module Parameter Efficient Fine-Tuning (PEFT)
- Conducted 60+ experiments on number of parameters, pretraining checkpoints, and data paraphrasing effects
- Enhanced retrieval performance by 10% through the integration of PEFT and GAR (Generation-Augmented Retrieval), applied to optimize service systems at E-Sun Bank, boosting enterprise productivity by 4%

BI Intern | (SQL, PL/SQL, MySQL, SSIS, DataStage, Cognos)

July 2022 – Sept 2022

WPG Holdings, Department of Business Intelligence

- Streamlined cross-departmental workflows by enhancing 27 electronic forms using IBM Cognos, resulting in a 12% increase in processing speed and reducing manual input errors by 8%
- Migrating 54 DataStage databases to SSIS, by implementing optimized ETL processes and constructing new databases for industrial decisions, which reduced data processing time by 13% and improved system stability
- Delivered operational improvements that boosted 15% in data accuracy and an estimated annual savings of \$50,000 in maintenance and error correction costs

Teaching Assistant | (R, Python)

Feb 2022 – June 2022

National Chung Hsing University, Course of Statistics

- Led discussion sections for large lecture courses, led to a 15% increase in participation and overall exam scores
- Utilized statistical software, R, to demonstrate practical applications, enhancing students' data analysis skills

DATA SCIENCE PROJECTS

Decisive Factors and Prediction of Health Insurance Premium - Regression Analysis

- Improved insurance premium prediction, achieving R-squared of 0.84 (Polynomial Regression) and 0.76 (Multiple Regression) through cross-validation.
- Enhanced precision using ANOVA, VIF analysis, variable selection, and Box-Cox to address multicollinearity.
- Developed Random Forest models, reducing RMSE by 70% in test datasets.

Data Processing and Display through Web Scraping- Air Quality Monitoring and Visualization

- Developed a web scraper to automate the collection of PM2.5 data, saving 5 hours/week of manual effort
- Performed data preprocessing, including missing value imputation processing and outlier removal

Model Selection, Ridge Regression, and Hyperparameter Tuning - Optimizing Diabetes Prediction

- Developed a predictive model for diabetes risk assessment to support early diagnosis
- Conducted Ridge Regression modeling using K-Fold Cross-Validation to split data
- Applied RidgeCV to optimize the hyperparameter and utilized GridSearchCV for further refinement.

PUBLICATIONS

Optimizing Knowledge Updates with PEFT in Neural Language Models [pdf]

Hao-Chun Shih, Wei-Yun Ma 2024

- Surveyed different Parameter Efficient architecture, investigating the features of FFN-Integrated and Attention-Integrated module
- Revised evaluation methods on memorization ability of model, and quantifying knowledge in LLMs
- Showcased research results at the internationally recognized International Conference on EMNLP