

Michael (Hau-an) Shieh

(206) 502-9668 | mikehash8382@gmail.com | Seattle, WA | LinkedIn: [mikes13](#) | Portfolio & GitHub

Advanced data modeling and analysis skills for solving problems with solid understanding of machine learning

EDUCATION

University of Washington, Seattle, WA, United States

Master of Science, Industrial Engineering | Sep. 2018 - Aug. 2020

- Research Topic: Security Update Decision Making Research
- Relevant Coursework: Data Analytics, Data Programming, Inferential Data Analysis, Statistical Learning, Design of Experiments

National Cheng Kung University, Tainan, Taiwan

Bachelor of Science, Management Science | Sep. 2012 - Jun. 2016

- Relevant Coursework: Applied Statistics, Database Management, Data Structure, Computer Programming, Operations Research

EXPERIENCE

Golden Financial Exploration Association, Bellevue, WA, United States

Data Analyst | Sep. 2020 - Present

- Built and managed a centralized MySQL database to combine data from multiple spreadsheets for easier access using queries
- Mentor and teach other colleagues SQL and how to use the database with a data pipeline I built with Python to import and export data
- Utilized data to select and invite potential participants to return to our new events, which resulted in 60% increase in number of participants
- Provided statistical consultation for analyzing, modeling, and predicting trends using the feedback collected from event participants

Data Analyst Intern | Jun. 2020 - Aug. 2020

- Analyzed and visualized data weekly to present insights with internal stakeholders to understand our audiences by events and over time
- Modified survey questions that resulted in responses with better quality to simplify the subsequent data processing

University of Washington, Seattle, WA, United States

Teaching Assistant | Jan. 2019 - Mar. 2020

- Probability and Statistics for Engineers (1 quarter), Manufacturing Scheduling and Inventory (1 quarter), Simulation (2 quarters)
- Improved students' performances by initiating conversations with students to collect feedbacks regularly to help instructors adjust materials, pace in classes, etc.

Institute of Information Science, Academia Sinica, Taipei, Taiwan

Research Assistant | Jan. 2018 - July 2018

- Developed investment strategies and yielded 50% more return compared to historical average return annually
- Processed trust fund dataset (1M rows) to identify which funds perform better regardless of market performances using regression models to apply Capital Asset Pricing Model (CAPM), S&P 500 dataset to predict stock market performances
- Collaborated with fellow researchers by preparing data for research projects, reviewing progress, and exchanging ideas

DATA SCIENCE PROJECTS

Security Update Decision Making Research | R, SQL, PHP, HTML, JavaScript, AWS

- Improved existing online experiment software to identify key features that affect humans' learning behavior on making decisions
- Collected and stored data in SQL database on SSH server for the subsequent analysis using regression models using RStudio
- Found positive reinforcement have little effect on improving human behaviors

Analyzing the Relationships between Airbnb Listing Prices and New York City Reported Crime | Python

- Processed and visualized reported crimes over the past 20 years by categories and by boroughs to analyze crime rate trends
- Developed Regression models and found average listing prices and occupancy rates are higher in boroughs with lower crime rate
- Applied random forest algorithm to produce a ML model to suggest listing prices for new listings

Predicting Growth of Gross Domestic Product per Capita | R

- Intended to predict GDP growth using human development indexes from the World Bank HDI dataset using RStudio
- Cross-validated different machine learning models to select the best approach by comparing learning error
- Implemented LASSO algorithm to select 44 independent variables from over 1600 variables to avoid overfitting
- Raised correlation by 33% and reduced learning error by 10% after clustering the countries by economic indicators

SKILLS AND CERTIFICATES

- **Proficient** with Python, R, and SQL and **Familiar** with PHP, HTML, C++
- Effective in Hypothesis Testing, Regression models, Clustering, Dimensionality Reduction, Causal Inference, and Design of Experiments
- Data Science Libraries and Tools: Pandas, NumPy, SciPy, Matplotlib, scikit-learn, TensorFlow, Keras, XGBoost, PySpark, Jupyter notebook, Databricks, Spark, Docker
- Udemy Certificates in Spark and Python for Big Data with PySpark, DubsTech Data Science Certificate