

Wei-Hua Hsu

✉ wafer110@gmail.com ☎ +1 (484) 744-8853 📍 Washington, DC 🐙 github.com/wafer110 🌐 linkedin.com/in/waferhsu

PROFESSIONAL SKILLS

- **Data Science:** EDA, Tidyverse, Statistical Methods, Regression, Visualization, and Interpretation.
- **Machine Learning:** SVM, PCA, kNN, Logistic Reg., LDA, QDA, Decision Trees, and Random Forests.
- **Tools:** Python, RStudio, Weka, and SPSS.
- **Database:** MySQL, Oracle, Relational DB, and basic Neo4j.
- **NLP:** BoW Analysis, Feature Engineering, Word Embedding, Text Mining, Sentiment Analysis, and Applied ML in NLP.
- **A.I.:** Data Mining, Deep Learning, and Neural Networks.
- **Languages:** Mandarin, English, Taiwanese, and Cantonese.

EDUCATION

Master of Science in Data Science

American University, Washington, DC, USA

Graduated: December 2020 **GPA:** 4.0/4.0

- Quant Supports & Stats Tutor, AU Academic Support & Access Center (Mar. – Dec. 2020).
- Technical Leader & Table Co-host, AU Catholic Student Organization (Jan. – Dec. 2020).
- Spring Cultural Festival Volunteer, International Student & Scholar Services (Apr. 2019).

Bachelor of Business Administration in Communications Management

Shih Hsin University, Taipei, Taiwan

Graduated: July 2016 **GPA:** 3.9/4.0

- Winner (2015) & Student Leader Coordinator (2016) of International Creativity, Innovation, and Entrepreneurial Charrette.
- Excellent Work Award in National TBSA Innovative Planning and Marketing Competition (2015).

RELEVANT PROJECTS

Utilized Machine Learning to Conduct Triple-Negative Breast Cancer Prognosis (Python, R, & Weka)

- Applied Feature Selections and Association Rules to define features that predict two-year survivability.
- Used Clustering and Classification methods to perform accurate TNBC prognosis.

Applied Natural Language Processing on Reddit Comments About Politics and Sports (Python)

- Identified trending issues in US politics, using Topic Modeling and Clustering.
- Trained Classification Model, resulting in remarkable accuracy in classifying similar themed subreddits.

How to Succeed in E-Sports: A Data Analysis on PlayerUnknown's Battlegrounds (R)

- Applied data preprocessing and data analytics using Tidyverse package to provide data analytics results.
- Visualized data correlations between variables including kill rate, damage rate, weapon/vehicle usage, survival time, moving distance, and standing positions on the game map utilizing ggplot2 package.

U.S. High School Track & Field and Cross-Country Data Analysis (R)

- Cross-analyzed track and field datasets with US statewide income and climate data to identify lower income athletes with better performance to recruit with Federal Work Study funds.
- Applied web scraping, data cleaning, and data analytics and visualized the statewide results on Shinyapps.

EXPERIENCE

Data Engineer Intern

AU Optronics Corporation, Taichung, Taiwan

June 2020 – August 2020

- Boosted the accuracy of anomaly detection from 95% to 98% in manufacturing TFT-LCD (display array).
- Applied NLP using Feature Engineering (Clustering and Topic Modeling), Keyword Analysis, AUO self-building dictionary, Word Embedding (TF-IDF, Bag of Words, and Bert), and Word Cloud on shift notes.
- Produced daily visualization reports alerting management to potential manufacturing line problems.
- Acted as Intern Leader, liaised between management and interns, and managed entertaining residential life.

Visual Content Specialist & TV Production Assistant

Sanlih E-Television Co. Ltd, Taipei, Taiwan

October 2018 – January 2019

- Broadcasted live streaming data, sports data, song lyrics, and graphics in a real-time production truck.
- Assisted the Production Director behind the screen to coordinate with live TV production flows.

Visual Content Producer & Software Trainer

Voxel Vision Technology Co. Ltd, Taipei, Taiwan

September 2017 – December 2018

- Gathered, cleaned, and organized data through Microsoft Excel and/or FTP Systems for multi-media productions.
- Produced data/visual content on CG Graphic Systems, streaming out on TV channels and live stream platforms.