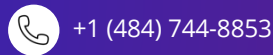


Wei-Hua Hsu



PROFESSIONAL SKILLS

- **Machine Learning:** algorithms in Python (Scikit-learn, t-SNE, Umap), Weka, and Programming R.
- **Data Science:** EDA, Tidyverse, Statistical Methods, Regression, Visualization, and Interpretation.
- **Natural Language Processing:** BoW Analysis, Feature Engineering, Word Embedding, and Text Mining.
- **Database Management:** SQL queries, and Relational DB.
- **A.I.:** Data Mining, Deep Learning, and Neural Networks.

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 (Jan. – Dec. 2020)
- Spring Cultural Festival Volunteer, ISSS (Apr. 2019)

Bachelor of Business Administration in Communications Management

Shih Hsin University, Taipei, Taiwan

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

- Winner (2015) and 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 which respond to 2 years survivability.
- Used Clustering and Classification methods to perform accurate TNBC prognosis and prediction.

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

- Successfully identified trending issues in US politics using Topic Modeling and Clustering.
- Trained Classification Model, resulting in remarkable accuracy in classifying similar themed subreddits.
- Utilized Multiple Regression models to conduct Sentiment Analysis – found weak correlation.

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

- Applied data preprocessing and data analytics using Tidyverse package and provided 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 athletes track & field datasets with US statewide income and climate data to identify possible lower income athletes with better performance – targeted FWS sponsorship students.
- Applied web scraping, data cleaning, data analytics and visualized the state-wide result on Shinyapp.com.

EXPERIENCE

DATA ENGINEER INTERN UTILIZING NATURAL LANGUAGE PROCESSING

AU Optronics Corporation, Taipei, Taiwan

June 2020 – August 2020

Demonstrated increased accuracy anomaly detection from 95% to 98% in manufacturing TFT-LCD (display array).

- Applied NLP using Feature Engineering (Clustering and Topic Modeling), Keywords 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.

Visual Contents SPECIALIST & TV PRODUCTION ASSISTANT

Sanlih E-Television Co. Ltd, Taipei, Taiwan

October 2018 – January 2019

The biggest cable television network in Taiwan, producing news, sports, dramas, and entertainment programs.

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