# Wei-Han (Zoe) Hsu

zoe70416@gmail.com | (332) 249-9798 | Website | GitHub | https://www.linkedin.com/in/weihanz-hsu

### **WORK EXPERIENCE**

# New York University IT - Anomaly Detection System for Network Security(GitHub)

New York

Machine Learning Engineer/Data Scientist

Feb 2023 - May 2024

- Performed data mining, data analysis (EDA), preprocessing, and feature engineering on a 100GB dataset.
- Developed scripts for the ETL process, ingesting unstructured network data through the ML pipeline.
- Led the team to develop **TensorFlow Autoencoder-based** anomaly detection, achieving a 73% F1 score on the NYU research network in identifying network anomalies, surpassing traditional firewalls.
- Implemented MLOps best practices using MLflow, Docker, Kubernetes, and CI/CD,

# Medidata Solutions, a Dassault Systèmes company - Medidata AI Consultant Department

New York

Data Scientist/ Commercial Analytics Intern

May 2023 - Aug 2023

- Leveraged A/B testing methodologies to evaluate and enhance the performance of sales teams.
- Developed and implemented a **Random Forest model** for revenue prediction, conducting thorough **feature importance analysis** to identify the most effective promotion channels (phone, in-person, or email).
- Allocate marketing budget across media channels, reducing expenditure on less effective channels.
- Anticipated a 3% improvement in targeted customer engagement through data-driven optimization.
- Collaborated cross-functionally to decode intricate data and extract comprehensive insights.
- Employed Snowflake, Pyspark, and AWS tools for ad-hoc data analysis and data processing.
- Created impactful data visualizations via Tableau to simplify complex information for stakeholders.

#### aetherAI- Medical device startup

Taiwan

Medical Technician/ Medical Data Management

Jul 2018 – Jun 2022

- Employed **SQL** to assess **data quality** for labeled cell images, achieving 0.8 accuracy in cell classification.
- Led development of Taiwan's first BM cell classification device, overseeing clinical trial data management.
- Collaborated cross-functionally to facilitate clinical validation processes for TFDA approval (Publication)

### Sleep Center of Shuang-Ho Hospital

Taiwan

Data Science Research Assistant

Aug 2020 - Mar 2021

- Undertook the preprocessing and statistical analysis of Real World Data (RWD). (Python and SPSS)
- Optimized and evaluated various ML models (Random Forest, Naive Bayes, LR, SVM, KNN) for apnea subtype classification (over 80% accuracy) using cross-validation and hyperparameter tuning in sci-kit-learn.
- Improved internal clinical decision-making using the model and co-authored an article in Diagnostics, 2021.

### **PROJECTS**

#### Chest Radiology Report Generation | PyTorch, BigQuery, LLM, Huggineface, CV, NLP, ViT, HPC (GitHub)

- Preprocessed large amounts (567GB) of X-ray images and electric health records from MIMIC-CXR.
- Applied **transfer learning** with BLIP-2 (ViLT and LLM-Vicuna) to automate Chest X-ray report generation.
- Enhanced report quality, achieving a BLEU-1 score of 0.34 through **find-tuning (LoRa)** and **prompt engineering**.

#### Credit Card Transaction Fraud Detection | Python, sklearn, AWS (S3, Sagemaker) (GitHub)

- Conducted EDA and applied undersampling/upsampling to address class imbalance.
- Trained classifiers (Logistic Regression, KNN, SVM, Decision Tree) using AWS S3 and SageMaker.
- Achieved a 0.98 ROC score in detecting fraudulent transactions.

#### **EDUCATION**

**New York University** 

New York, NY

M.S in Data Science (GPA: 3.7/4.0)

Sep 2022 - May 2024

**Taipei Medical University** 

Taiwan

B.S in Medical Laboratory Science and Biotechnology (GPA: 3.9/4.0)

Sep 2016 – Jun 2020

Core Coursework: Probability and Statistics, Linear Algebra, Machine Learning, Big Data, Deep Learning SKILLS

**Libraries & ML Frameworks:** Numpy, Pandas, Matplotlib, Pytorch, Tensorflow, sci-kit learn, Langchain **Programming Languages:** Python, SQL, R, SPSS, SAS, C++

**Tools:** AWS, Azure, BigQuery, Spark, Hadoop, Snowflake, Databricks, Docker, Kubernetes, GitLab, Git, Tableau **Techniques:** Statistical Analysis, Classification Analysis, Machine Learning, A/B Testing, Cloud Computing