Yu-Wen Lai

ywl273@nyu.edu | (929) 401-8800 | New York City, NY

Work Experience

Acer Inc. New Taipei, Taiwan
Data Engineer, Value Lab *Sep. 2016 - Jul. 2019*

- **Cancer Immunotherapy**: A research project on neoantigen prediction
  - Conducted study on peptide identification using LC-MS/MS spectra
  - Re-implemented Whole Exome Sequencing data analysis pipeline resulting in 10X enhancement in efficiency
- **Big Data Platform for FarEasTone Telecommunications**: A platform based on Hortonworks (IBM)
  - Collaborated with IBM to create a prototype of real-time analysis of location based data and successfully won the bid (≈ $645K)
- **Smart Taxi Operating Platform for Taiwan Taxi**: A demand prediction APP for taxi drivers
  - Developed web crawlers for information about concerts and exhibitions by Python
  - Implemented a near-real-time application using Spark Streaming to analyze the status of taxis (≈ 100K records/min)
  - Designed and built data pipeline to integrate multiple data sources into internal HDFS for training

SAS Institute Taiwan Ltd. Taipei, Taiwan
Consultant Trainee (Intern), Professional Services Division *Apr. 2015 - Feb. 2016*

- **Pattern Recognition**: Researched and implemented algorithms to detect defects on semiconductor wafers
- **Data Analysis**: Assisted customers in building data analysis pipeline with SAS Enterprise Miner

Education

New York University New York City, US
Master of Science in Information Systems *Sep. 2019 - May. 2021 (expected)*

National Taiwan University Taipei, Taiwan
Master of Science in Physics *Sep. 2012 - Jun. 2016*

Relevant Coursework: Machine Learning, Data Visualization, Web Retrieval and Mining, Fundamental Object Oriented Programming, Data Structures and Algorithms, Algorithm Design and Analysis, Systems Programming, Computer Architecture, GPU Programming

National Taiwan University Taipei, Taiwan
Bachelor of Science in Physics *Sep. 2008 - Jun. 2012*

Relevant Coursework: Numerical Analysis and Programming

Projects

- **Chinese Characters Recognition (Machine Learning Course)**:
  - Classified 32 possible labels of written Chinese characters
  - Attained approximated 70% precision by using KNN, random forest, and SVM
  - Implemented data pre-processing, training, and testing by Python

Achievements

- Translated Romeo Kienzler's book, ***Mastering Apache Spark 2.x***, to Chinese
- Won *2nd Place* out of 200+ teams with $3,200 grant at **Big Data Analytics for Semiconductor Manufacturing 2015**
- Won *Bronze Medal* (placed 4th out of 2000+ teams from all over the world) with $20,000 grant at **Acer Incredible Green Contest 2012-2013**

Skills (order by fluency)

- **Languages**: Python, C/C++, Java, SQL, Bash, R, SAS
- **Libraries**: Spark, Scrapy, Flask, Numpy, Pandas, Scikit-Learn, TensorFlow, CUDA
- **Tools**: MySQL, Impala, Hadoop, Kafka, Git, Docker, Ansible, Prometheus, Grafana