

34834 Wabash River Place, Fremont, CA, 94555

+1 (929) 332-6086

jimmy.yang@nyu.edu

Chun-Yi Yang

<https://netjimmy.github.io/>

<https://github.com/netjimmy>

<https://www.linkedin.com/in/netjimmy>

OBJECTIVE

Graduated student from NYU, specialize in big data framework such as Hadoop/Spark, web programming and machine learning. Regular attendee of data profession meetups. Seeks for fulltime software/data engineer position starting Spring/2019

COMPUTER SKILLS

- **Programming Languages:** Python, Java, Scala, CSS, HTML, JavaScript, SQL, NoSQL, Linux Shell
- **Framework:** Hadoop, Spark, Spark Streaming, Kafka, Flink, Hive, Pig, Sqoop, Node.js, React, Redux, Flask
- **Miscellaneous:** AWS (EC2, DynamoDB, S3, EMR), Git, Docker, Tableau, REST, Data Analysis, Machine Learning

EMPLOYMENT

Software Engineer Intern **Toii Virtual Gaming Lab, NY** **09/2018 – 12/2018**

- Implemented **RESTful APIs** endpoints to make HTTP requests for a A.R. location based mobile game
- Worked with **Node.js** framework, **Express.js** to manage backend and migrated data to **MongoDB**
- Implemented authentication with Google OAuth function and handling payments by Stripe function on App

Data Engineer Intern **Intumit, Taiwan** **06/2017 – 08/2017**

- Built a recommendation system for job seekers to map key words to clusters by python **scikit-learn library**
- Implemented **web crawlers** and bypassed anti-scraping mechanism from job board websites by python script
- Batch/stream processed data with crontab on **AWS** instances and store context into **MongoDB** database (10 GB)
- Extracted job entities and feature attributes from job boards and transformed text to vector form for **NLP** analysis

Project Manager **QNAP System, Taiwan** **07/2014 – 06/2015**

- **Product Strategy Design:** to define customer needs, and to deliver product blueprint to engineering and marketing team
- **Business Development:** help to exploring new market with senior managers and increase selling by **39%**

PROJECTS

Audible Social Media [*Kafka, Spark Streaming, AWS (EC2, S3, DynamoDB, EMR)*] **Spring / 2018**

- Programmed a web application that categorizers users' social media feeds and converts text feeds to streaming audio clips
- Ingested Twitter data with **Kafka** and stream processed by **Spark Streaming** to stored data on **AWS DynamoDB**
- Deployed service on **AWS EC2** machine, batch processed user tweets to audio and stored on **AWS S3**

Online Music Service Churn Prediction [*Python, NumPy, SciPy, scikit-learn*] **Fall / 2017**

- Built a predictive model on whether users will renew service by **Python** and **scikit-learn**
- Applied **feature engineering** to generate 30 new features and improved accuracy from 86.1% to 95.7%
- Designed **data schema** optimized space and speed, transformed raw data to training data to save data space by 95%

Stock Price Trend from Financial News [*Scala, Spark, Spark ML, Tableau*] **Spring / 2017**

- **Automated data pipeline** and Extract-Transform-Load (ETL) from 6 data sources in periodic intervals
- Utilize **Spark ML** to implement Logistic Regression, Random Forest, and Gradient Boosting (best precision: 86%) and visualized the comparing effectiveness by Heat Map via **Tableau**

EDUCATION

New York University, New York, NY, **09/2016 – 06/2018**

- M.S. in Computer Science
- Core Courses: Cloud Computing, Big Data Application and Development, Machine Learning, Predictive Analysis

National Taiwan University of Science and Technology, Taipei, Taiwan **09/2012 – 06/2014**

- M.S. in Industrial Management
- Core Course: Data Mining Application, Computational Intelligence in Engineering, Numerical Analysis