34834 Wabash River Place
Fremont, CA, 94555
https://www.linkedin.com/in/netjimmy

Chun-Yi Yang

(929) 332-6086
jimmy.yang@nyu.edu
https://github.com/Netjimmy

OBJECTIVE

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

EMPLOYMENT

Software 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 5 **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
- 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

REST APIs by Flask and Python *[REST API, Flask, Heroku, Postgres]* **Fall / 2018**

- Implemented **RESTful APIs** endpoints by Python to make HTTP requests for uploading, remote endpoint access
- Worked with **Flask** framework, **Flask-SQLAlchemy ORM** to manage backend and SQL-like database
- Auto-deployed web service on **Heroku** machine, stored data in **Postgres** database

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

- Programmed a web application that categorizers users' social media feeds and converts text feeds to steaming audio clips
- Ingested data with **Kafka** to stream process Twitter API and 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 s**cikit-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, GPA: 3.3
- 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, GPA: 3.87
- Core Course: Data Mining Application, Computational Intelligence in Engineering, Numerical Analysis

COMPUTER SKILLS

- **Programming Languages:** Python (NumPy, SciPy, scikit-learn, Matplotlib), Java, Scala, SQL, Linux Shell
- **Framework:** Hadoop, Spark, Kafka, Hive, Pig, Sqoop, MapReduce, REST, Flask, SQLAlchemy, Docker
- **Data Analysis:** Machine Learning, Regression, Bayesian, Statistics, Data Visualization, Tableau