ZINIU YU

Irvine, California 92612 | (949) 992-5907 | ziniuy1@uci.edu| linkedin.com/in/ziniu-yu-2396291a6 | kaggle.com/unizy22

SUMMARY

- Current Master of Data Science student with internship experiences as an analyst and data scientist.
- Strong skills in using Web Crawlers to collect data, applying SQL to manage data, and building Statistical or ML models to solve problems.

TECHNICAL SKILLS

Core Domain Expertise: Data Analysis, Statistical Modeling, Machine Learning, Data Mining
Tools: Python3, Web Crawler, Hive SQL, Postgresql, Jupyter, Pytorch, Pandas, Numpy, Keras, Keras Tuner

EDUCATION

University of California, Irvine, California
Master of Data Science, GPA: 3.90/4.00 Expected Dec. 2022

Harbin Engineering University, Harbin, China
B.S., Mathematics and Applied Mathematics, GPA: 3.62/4.00 June. 2020

EXPERIENCE

XIAOMI TECHNOLOGY. Wuhan, China
Data Scientist Intern, Big Data Department Oct. 2019 – Jan. 2020

- Built a **word2vec word vector** NLP model to calculate the similarity between finance words. Expanded 5,000 words, and enhanced the recall of the tagging process by 3%.
- Deployed a **Web Crawler** to collect music entities, added more than 10,000 entities, and 87% of them were utilized for tagging users to make advertising more precise.
- Operated **hive SQL** to query music data, analyzed the popularity of them with **Pandas**, and helped increased DAU by 5000.
- Coded in **Scala**, helped improve the data mining logic, and improved the accuracy of tagging process to 90%.

WUHAN BUREAU OF STATISTICS. Wuhan, China
Analyst Intern Aug. 2019 – Sept. 2019

- Estimated the total factor productivity (TFP) leveraging Cobb–Douglas production function. Helped researchers understand the impact of technological innovation on Wuhan's economy from a quantitative perspective.
- Applied the Time Series Regression to predict the potential economic growth of Wuhan, and contributed to a theoretical paper about the influence of technology development on the economy.

PROJECTS

Real-time gesture recognition, ● Python3 ● Pytorch ● OpenCV ● Numpy ● Pandas Nov. 2021 – Dec. 2021
Real-time gesture-classifier using neural networks(SSD & Resnet) on Pytorch. (https://youtu.be/Kx9p1sGUAGg)

- Used EgoHands and COCO-Hands datasets to train the **SSD300** model, connected it to the web camera by **OpenCV**, and developed a hand detector with 18 FPS.
- Performed transfer training on the ResNet-18 model with accuracy of 98%. Connected it to the hand detector and constructed a real-time gesture classifier with 12 FPS.

Jane Street Market Prediction, ● Python3 ● Keras ● Keras Tuner ● Numpy ● Pandas Dec. 2019 – Feb. 2021
Kaggle competition that uses attributes to decide whether a trade should be made.

- Cleaned the data that has little impact on decision-making, filtered the features bias with **Autoencoder**, adjusted the hyperparameters with **Keras Tuner**, and boosted the score by 3000.
- Constructed an MLP model by **Keras** to determine whether the transaction should proceed, ranked top 17% on the public leaderboard, and defeated 3176 teams on the private leaderboard.

Traffic Mode Recognition, ● Python3 ● Sklearn ● Numpy ● Pandas Jan. 2019 –Feb. 2019
A classifier based on the random forest model to distinguish traffic mode.

- Used the genetic algorithm on Matlab to calculate the approximate signal function, found a suitable function.
- Constructed a random forest model with Sklearn, and got a classifier of traffic mode whose accuracy and recall are both over 90%.