

# Tianye Song

clairesong24@gmail.com | 412-933-9177 | <https://www.linkedin.com/in/tianye-song/> | open to relocation

## EDUCATION

**Carnegie Mellon University** - Master of Science in **Information Systems Management**, GPA: 3.75 Expected 12/2022  
**Xiamen University** - Bachelor of Economics in **Finance**, GPA: 3.8 06/2020  
• **University of Oxford** - Winter Exchange Student, GPA: 4.0

## TECHNICAL SKILLS

*Programming Languages:* Python, SQL, Java, R, HTML

*Tools:* PyTorch, TensorFlow, scikit-learn, Spark, Hive, Hadoop, Dask, MySQL, MongoDB, Docker, AWS, Tableau, Kafka, pandas, Excel

## WORK EXPERIENCE

**Data Scientist (Full Time)** - *Meritco Services Technology Consulting Firm* Shanghai, China | 07/2020–07/2021

- Orchestrated multiple promotion opportunities for companies by analyzing big data with **Python** and **SQL**; increased conversion rate by 5%; facilitated the customer relationship management process with **statistics model** and increased ROI by 7%
- Simplified sku name matching tools with **XGBoost** and accelerated the generation time by **120 times**
- Designed and engineered **interactive tool-chains** for manufacturing planning including integrating data from **multiple sources**, building a sales prediction model and visualizing indexes through dashboard; lifted prediction accuracy by **50%**
- Boosted the key function of **Key Opinion Leader product** by recognizing cosmetic product names and features automatically with **NLP algorithms**; achieved 87% accuracy and simultaneously saved **80%** of time

**Data Engineer Intern** - *Walmart* Dallas, TX | 05/2022–08/2022

- Implemented accelerated computing by transferring data pipelines and modeling processes from CPU, which originally used **pandas** and **Spark**, to **GPU distributed computing** with RAPIDS cuDF and **Dask** on **web-scale data**; increased the speed by **30 times**
- Collaborated with NVIDIA to contribute to RAPIDS, an open-source suite of GPU accelerated libraries in Python, by refactoring and generalizing data analysis, data pipeline and modeling functions

**Co-Founder** - *Trace* Shanghai, China | 12/2021 - 09/2022

- Constructed a digital identity network in the form of an **IOS application** for Z generation with stories to help them discover, record and share different aspects of themselves and connect with similar people
- Qualified for **seed round funding** of Y Combinator China, a technology startup accelerator that has been used to launch more than 3,000 companies including Stripe and Airbnb (top **2%** of 1000 companies)

**Data Scientist Intern** - *Meritco Services Technology Consulting Firm* Shanghai, China | 07/2019–09/2019

- Explored feasible algorithms from papers and blogs related to association rules and formalized Restricted Boltzmann Machine, Hidden Markov Model and PrefixSpan on 60 million data points with **Python**
- Enhanced the online purchase recommendation system by PrefixSpan algorithm and increased conversion rate by 10%

## PROJECT EXPERIENCE

**Production of Movie Recommendation System** 01/2022–05/2022

- Designed and Engineered a **recommendation system** by collecting data from **Kafka** and **API** with 1 million customers and 27k movies, storing data in **MongoDB**, building a **collaborative filtering** model and deploying the inference service with **flask** in Python
- Constructed **Docker** to containerize the serving infrastructure for **A/B testing** of different versions of model, automated the model retraining process, and implemented **Jenkins** for continuous integration
- Created metrics to evaluate and monitor data quality, fairness, and system health with **Prometheus** and **Grafana**

**Speech Recognition System** 04/2022–05/2020

- Implemented framework of wav2vec 2.0 algorithm, k-means clustering and GAN for automatic unrecognized speech recognition
- Built a combination of RNN and Attention to design end to end recognition system for speech-to-text transcription

**Application of Machine Learning in Online Clothes Products Sales Prediction** 05/2020–06/2020

- Formalized machine learning models (Random Forest and XGBoost) in Python for online sales prediction for new clothes products

## LEADERSHIP EXPERIENCE

Toastmasters International Speech Club, Secretary, Xiamen University 01/2019–01/2020

AIIESEC, International Volunteer, Sri Lanka 07/2017–09/2017