Yichen Li

Phone (Mobile): +1 7348818107 **E-mail:** liyichen@umich.edu
Address: 1537 McIntyre St., Ann Arbor, Michigan 48105
LinkedIn: https://www.linkedin.com/in/yichen-li-7178041b0
GitHub: https://github.com/Patronus-Yichen-Li

OBJECTIVES

- ✧ Seeking **DATA SCIENTIST/DATA ENGINEER/DATA ANALYST**, can start from **03/01/2023**
- ✧ Programming Languages: (Proficient) **Python, C++, SQL**; (Experienced) Java, Julia, C, C#, Matlab
- ✧ Solid foundation of Machine Learning theories and strong practical skills of popular ML packages (**TensorFlow**, **Keras**, **PyTorch**, **Sklearn**) and Data Analysis/Statistical tools (**Numpy**, **Pandas**, Excel, Matlab, Tableau). Deep understanding and practical experience of database (**SQL**, **MongoDB**), cloud computing system (**AWS**), big data system (**Hadoop**, **Spark**, CDP/CDH), and Linux development environment.

EDUCATION

University of Michigan-Ann Arbor (Umich), Ann Arbor, U.S 01/2021~12/2022
School of Electrical Engineering and Computer Science **Master.** *Data Science and Machine Learning* GPA: 3.8/4.0
Main Courses: Probability and Random Process, Matrix Method for Data Science and Machine Learning, Advanced Machine Learning, Artificial Intelligence Foundation, Optimization of Machine Learning, Lab Design for Data Science and Machine Learning, Program Management Consulting
Michigan Institute of Data Science **Graduate Certificate.** *Data Science* GPA: 4.0/4.0
Main Courses: Programming for Scientist and Engineer (C++), Database Management System, Statistical Inference

Huazhong University of Science and Technology (HUST), Wuhan, China (Ranked Top 9 in China) 09/2016~06/2020
School of AI and Automation **Bachelor.** *Electrical and Computer Engineering* GPA: 3.4/4.0

PROFESSIONAL EXPERIENCES

The Linux Foundation San Francisco, CA (Remote)
AI/ML Data Scientist Intern 07/2022 - Present

- ➢ Implemented a software system log analysis pipeline using Python, including data ETL, data feature extraction based on log event type and variables vectorization, and a log event time-series sequence prediction method based on the multivariant-Bi-LSTM algorithm. The algorithm can maintain precision of more than 96% of diagnosing abnormal log messages. It will be integrated into NFV system, an important layer in communication system, to prevent critical software failures and help developers determine workflow
- ➢ Collaborating on developing and training GAN generator for log data generation, in order to compensate for the lack of existing data.
- ➢ Working on using K-anonymity and Differential Privacy methods to build anonymization tool to keep privacy of existing log data

Intel Corporation Shanghai, China
Machine Learning Engineer/Big Data Analyst Intern 03/2021 - 08/2021

- ➢ Implemented an end-to-end Machine Learning pipeline, using datasets with size more than 10TB, using the recommendation model, *Deep Interest Evolution Network* (DIEN), using NumPy, Pandas, TensorFlow, and Modin, making the performance of data ETL and feature extraction parts 10 times better in single-threaded processing, and deployable in distributed systems
- ➢ Collaborated with the full-aspects analysis and optimization (application, framework, and cluster) of machine learning and big data projects deployed on Hadoop, Hive and Spark for different workloads, including SQL queries, MapReduce and multiple machine learning tasks, according to the performance statistics from benchmark tools TPCx-BB & TPCx-DS, providing critical suggestions on optimization tasks to the team, and rewritten the company's guide book of Hadoop and Spark configuration for multiple tasks

GrapeCity, Inc. Xi'an, China
Software Development Engineer Intern 08/2020 - 12/2020

- ➢ Implemented APIs for *Document for Excel*, one of GrapeCity's flagship software products using C# and Java, with functions like cellpadding, cross-file references, JSON IO with options, etc. No major bug reports after each release
- ➢ Collaborated on developing basic package function Calculation Engine, implementing data manipulation logics

PROJECT EXPERIENCES

Stock Price Prediction based on Concept-Oriented Shared Information 01/2022 - 05/2022

- ➢ Developed and implemented the new neural network, taking the relationship between stocks into account, constantly update relationships based on time changes, combining time series information, to better predict the future trend of stocks. Reaching a goal of Information Coefficient (IC) around 0.0167 and a precision rate over 0.50 **[Poster]** **[Report]**
- ➢ Published a demo application, implementing a simplified version of this project, on AWS (ECR) and Heroku platform, performing a Grayscale correlation analysis to visualize the results **[Web Application]** **[Report]**

A data-based SOFC performance reasoning and optimization method 08/2019 - 06/2020
Patent: Yichen, L. 2020. Performance reasoning and optimizing method for solid oxide fuel cell. Publication as **CN112131775A**

- ➢ Applied the Particle Swarm Optimization (PSO) algorithm to find the most reasoned coefficients with known model expression in state space using MATLAB and transferred the model into a finite state space for better convenience of further analysis
- ➢ Performed Granger Causality Test on each variable (inputs, measures, and outputs) in the actual system operating data to obtain the relationship between them, to further improve and modify the existing model and coefficients