Hao Li

857-891-9461 Leo.t5@outlook.com vida42.github.io

EDUCATION

Northeastern University Boston, MA
- Candidate for Master of Science in Data Analytics Engineering May 2020
- Relevant Courses: Supervised Machine Learning, Data Mining

Alibaba Research Center for Complexity Science Hangzhou, China
- Master of Science in Computer Science June 2018
- Focused Areas of Study: Network Science, Linked Prediction

Xidian University Xi'an, China
- Bachelor of Science in Mathematics and Applied Mathematics July 2015

SKILLS

- Languages: Python (pandas/numpy/scipy/matplotlib), SQL, R, JavaScript, C
- Tools: Sklearn, Tensorflow, A/B Testing, Spark, MapReduce, Linux, Git, Tableau

PROFESSIONAL EXPERIENCE

Meiao Biotechnology Co., Ltd. Hangzhou, China
Business Analysis Intern Jan. - Feb. 2017
- Utilized PEST Analysis and SWOT Analysis to offer advice on selling Dendrobium
- Produced a business report and helped the company increase their sales by 9% in 2 months

PROJECTS AND ACADEMIC PUBLICATIONS

Content-based Movie Recommender System Boston, MA
Individual Project, Modeling Feb. 2019
- Predicted ratings for movies with Tensorflow and recommended movies to users

Adult Income Prediction Using Classification Techniques Boston, MA
Course Project, Modeling and Data Mining Nov. 2018
- Implemented 3 versions of tree-based models and neural network to achieve our goal in classifying whether an individual's annual salary is more than $50,000
- Estimated imbalanced data with SMOTE technique and improved accuracy of all models to 86% and higher

Measuring Diversity of Music Tastes in Online Musical Society Hangzhou, China
First Author, Data Science Workflow Oct. - Dec. 2017
- Designed a true diversity measurement to better capture diversity of users' musical tastes
- Discovered factors (levels of economic development etc.) that greatly impact users' music tastes
- Full Paper accepted by International Journal of Modern Physics C

Distributed WebCrawler and User Behavior Analytics for Xiami Music Hangzhou, China
Individual Project, Data Collection, Manipulation and Analysis Sept. - Dec. 2017
- Created Python-based web crawler engine and collected favorite music lists and users' information of more than three million users
- Designed MySQL database to manage 2GB data, generated six relational tables under 3NF
- Provided Ximai user profiles with fifteen features such as socioeconomics and demographics
- Evaluated Influence of users by using H-index method

The Weighted DHC Theorem and its Application on the World Trade Web Hangzhou, China
First Author, Data Science Workflow May - Sept. 2017
- Employed a weighted DHC theorem and analyzed twenty years' global trade data
- Identified influential countries in the World Trade Web and their evolution over time
- Proved DHC is a better method of quantifying node influence in directed weighted networks
- Abstract accepted by NetSciX2018