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, Computer Networking

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

- Programming Languages: Python, SQL, R, C, JavaScript, HTML, CSS
- Tools: Tensorflow, Spark, Hadoop MapReduce, Nginx, Redis, Scrapy, Linux, Git

PROFESSIONAL EXPERIENCE

Orimuse   Xi'an, China
Co-Founder, Database Design, System Design   March - May 2014
- Developed a T-shirt customizing application, Orimuse, which has been thrived for 5 years
- Designed a database schema, split data into three shards: user, product and order
- Built MySQL Master/Slave replication, copied data on slaves, achieved Read/Write splitting
- Utilized Nginx as a load balancer, split system horizontally to run multi-backend

PROJECTS AND ACADEMIC PUBLICATIONS

Sentiment Analysis of GitHub Commit Comments   Boston, MA
Individual Project, Data Manipulation, Sentiment Analysis using VADER   Feb. - March 2019
- Measured emotions expressed in different projects' commit comments by using lexical sentiment analysis
- Revealed the relationship between the expressed emotions and different factors such as used programming language

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

Adult Income Prediction Using Classification Techniques   Boston, MA
Course Project, Modeling and Data Mining   Nov. 2018
- Implemented Classification tree, Boosting tree, Random forests and Neural network by Python 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

Raw Sockets *HTTPGET* Application   Boston, MA
Course Project, TCP/IP Packets Implement   Oct. 2018
- Wrote a program that takes a URL on the command line and downloads the associated file and saves it to the current directory
- Rebuilt the system's TCP/IP stack in socket, implemented all features of IP/TCP packets

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