



Hao Li

857-891-9461 | Boston, MA | li.hao6@husky.neu.edu | leot5.com  

EDUCATION

Northeastern University , Master of Science in Data Analytics Engineering, Boston, MA	Dec 2020
Relevant Courses: Data Mining, Machine Learning, Big-Data System Engineering Using Scala	
Hangzhou Normal University , Master of Science in Computer Science , Hangzhou, China	Jun 2018
Relevant Courses: Database, Operating Systems, Artificial Intelligence and Pattern Recognition	
Xidian University , Bachelor of Science in Mathematics and Applied Mathematics, Xi'an, China	Jul 2015

SKILLS

Programming Languages: Python, SQL, R, Java, Scala, JavaScript, HTML, CSS, MATLAB

Databases: Relational databases (MySQL), Non-relational databases (MongoDB, Redis)

Data & Tools: Spark, Git, Scrapy, Scikit-learn, TensorFlow, Tableau

PROFESSIONAL EXPERIENCE

Lenovo	Beijing, China
<i>Software Engineer Intern</i>	May 2019 - Aug 2019

- Built one front-end BI tool to monitor the Lenovo smart education system
- Connected the client and the **MySQL** database with **JDBC** to provide data for visualization dashboard
- Implemented **Java** scheduled tasks regularly based on **Spring** to update visualization panels of BI tool
- Shed light on students' and teachers' behavior information and provided guidance on improving teaching quality

Alibaba, Research Center for Complexity Sciences	Hangzhou, China
<i>Research Intern</i>	May 2017 - Dec 2017

- Implemented **Python-based web crawler** and collected 8,338,301 records of users' favorite music data
- Designed **MySQL database** to manage 585,801 users' profile and favorite music data
- Proposed an improved Hill-type true diversity metric that fairly captures the diversity of users' music tastes and discovered factors that greatly impact users' music tastes, paper accepted by IJMPC as first author

Orimuse	Xi'an, China
<i>Software Engineer Intern</i>	Mar 2014 - May 2014

- Developed a T-shirt customizing mobile application which has been thriving for 5 years
- Designed **database schema** by partitioning data vertically into three databases to improve search capabilities
- Achieved **Read/Write splitting** by building **MySQL Master/Slave replication** to raise database performance

PROJECTS

E-commerce Recommender System	Sep 2019 - Present
--------------------------------------	--------------------

- Implemented matrix factorization based recommender service by using **ALS** algorithm and training latent factor model in **Spark**
- Implemented statistics based recommender service by using **Spark SQL** to solve cold-start problem
- Built main database with **MongoDB** to manage data and provide recommendation results to front-end server
- Developed an interactive web application using **Play** framework to recommend products based on users' input

Doge Invasion Game	Mar 2019 - May 2019
---------------------------	---------------------

- Developed a 2D shooting game by using completely **Object-oriented** approach and Pygame

Adult Income Prediction with Classification Techniques	Nov 2018 - Dec 2018
---	---------------------

- Handled imbalanced data with **SMOTE** using Python, improved sensitivity of all models up to 88%
- Implemented **Classification Tree**, **Boosting Tree**, **Random Forests** and **Neural Network** by Python to classify whether an individual's annual salary is more than \$50,000

PUBLICATION

- Li, Hao, et al. "Measuring diversity of music tastes in online musical society." International Journal of Modern Physics C 29.05 (2018): 1840006.