



Hao Li

857-891-9461 | Boston, MA | li.hao6@northeastern.edu | [mugen20.com](mailto:li.hao6@northeastern.edu)  

EDUCATION

Northeastern University, Master of Science in Data Analytics Engineering, Boston, MA Dec 2020
Relevant Courses: Database, Data Mining, Programming Design Paradigm, Big-Data System Engineering
Hangzhou Normal University, Master of Science in **Computer Science**, Hangzhou, China Jun 2018
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, PostgreSQL), Non-relational databases (MongoDB)
Data & Tools: Git, Django, GCP, Scrapy, Spark, Scikit-learn, TensorFlow, Tableau

PROFESSIONAL EXPERIENCE

Epion Health North Andover, MA
Data Engineer Intern Jan 2020 - Jun 2020

- Built data pipeline using Python API to communicate with Hubspot and Salesforce to consolidate over 2k+ data
- Created Python scripts to research and provide sales and marketing teams 20-30 potential accounts weekly from Athenahealth and Eclinicalworks within 4 weeks
- Built report automation process by transforming manual workflow into code, reducing the processing time from 1 week to 1 hour
- Standardized file management by setting up GCP storage for Django app for file upload

Lenovo Beijing, China
Software Engineer Intern 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
Research Intern 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
Software Engineer Intern 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

PUBLICATION

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