Lihua (Neo) Pei

• Email: lihua.peidata@gmail.com • Phone: +1 847-848-5805 • Portfolio of Projects: https://lihuapeineo.github.io

Objective

Data Engineer: I am looking for Data Engineer and Data Scientist opportunities.

Legal Worker: I am an US Permanent Resident.

Education

The George Washington University — Washington, DC
Master Degree in Data Analytics. — *Expected May 2020*

Stony Brook University — Stony Brook, NY
Bachelor Degree in Applied Mathematic & Statistics. GPA: 3.5/4.0 — *December 2017*

Skills

- Languages: Python, SQL, R, Java, HTML/CSS and JavaScript.
- Systems: MySQL, MongDB, Linux, Excel, and Latex
- Others: Familiar Machine Learning, Data Visualization, Database Management and Nature Language Processing.

Industry Experience

Fresh Air DC Project - uRADMonitor, Inc. — Washington DC
Data Scientist Intern — *May 2019 - Present*

- **ETL**: Designed and Built new ETL pipelines to transform data into analytics friendly schema by Python (Pandas and Numpy) resulting in a 90% drop in ETL and 50% efficiency improvement on this project.
- **Database**: Designed and managed the remote MySQL database (PyMySQL) to collect 6 uRAD monitors' data updating every 30 seconds, storing at the George Washington University's cloud servers.
- **Comparative Analysis**: Conduct the Time Series Analysis and the Correlation Analysis to test new uRAD monitors' performance on millions of observations between the test groups and control groups.
- **XFN Work**: Presented reports with the business insights and improvement suggestions to the engineering department.

Fire Pillar Studio — Hong Kong, China
Database Developer — *June 2017 – May 2018*

- **Database**: Designed and managed the local MySQL databases for the project which stored over 20,000 original user updated pictures and processed data.
- **Deep-Learning**: Participated in developing a deep-learning program to set up an artificial neural network that can recognize the cartoon characters in pictures by using TensorFlow.
- **Unity**: Developed a Unity program with the team to make characters have an action like breathing.

Research Experience

Optimal Topologies Searching Research Project — Stony Brook University
Research Assistant Guided by Prof. Yuefan Deng — *August 2017 - March 2018*

- **Database**: Designed and managed the local database to store over 10 millions generated topologies for research purposes.
- **Algorithm**: Designed and implemented a genetic algorithm to search the optimal topologies from 10 millions generated topologies on High-performance Computer.
- **Data Analysis**: Collected the performance data of new topologies and visualized experimental results. The optimal topology achieved 150% to 320% better efficiency than commonly used ring topologies.

Data Analytics Projects

Stocks Trend Prediction System — The George Washington University
SVM (Support Vector Machine) — *September 2018 – December 2018*

- Collected data of top 500 stocks with 18 significant features crawling (Beautiful Soup) from Yahoo Finance websites.
- Created a SVM-based machine learning model to predict the market trend and stocks' price. Achieved average 78% accuracy.

Donald Trump's Twitters Analysis — The George Washington University
Topic Modeling — *May. 2018 – October 2018*

- Applied regular expression and NLTK to construct the twitters' context cleaning functions to transform 12,000 twitters of realDonaldTrump to analytics friendly words bag.
- Constructed an NLP model applied the LSI and LDA methods. The model showed good performance to extract the topic for each given Trump's twitter.