Lihua (Neo) Pei

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

EDUCATION

The George Washington University — WASHINGTON, DC
Master Degree in Data Analytics. GPA: 3.70/4.00 — *May 2020*

Stony Brook University — STONY BROOK, NY
Bachelor Degree in Applied Mathematic & Statistics. GPA: 3.55/4.00 — *December 2017*

SKILLS

- **Languages:** Python, SQL, R, Java, HTML/CSS and Javascript.
- **Systems:** MySQL, MongoDB, Linux, AWS, Spark, Excel, and Latex
- **Others:** Familiar with Machine Learning, Data Visualization, Database Management, Deep Learning, Nature Language Processing and Blockchain Technology.

INDUSTRY EXPERIENCE

Student Veterans of America — WASHINGTON DC
Data Scientist — *May 2020 - Present*
- **Data Analytics**: Designed the pipelines to analyze the annual census of student veterans and visualize the results in descriptive charts.
- **Webpages**: Implemented interactive webpages to present research results and data visualizations by Tableau and HTML/CSS.
- **AWS**: Designed the pipelines to transform data into analytics friendly schema and update data to AWS S3 and RDS.

Fresh Air DC Project - uRADMonitor, Inc. — WASHINGTON DC
Data Scientist Intern — *May 2019 - October 2019*
- **ETL**: Designed and built new ETL pipelines to transform data into analytics friendly schema by Python (Pandas and Numpy) resulting in a 90% reduction in ETL time.
- **Database**: Designed and managed the remote MySQL database (PyMySQL) to collect uRAD monitors' data updating every 30 seconds, storing at the George Washington University's cloud servers.
- **Comparative Analysis**: Conduct Time Series Analysis and 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 cartoon characters in pictures by using PyTorch.
- **Unity**: Developed a Unity program with the team to make characters have life-like breathing.

DATA ANALYTICS PROJECTS

Hybrid Content-based Movie Recommendition System — THE GEORGE WASHINGTON UNIVERSITY
Recommendation System — *October 2019 – December 2019*
- Designed distributed pipelines to automatically use YouTube API to download the movie trailers and use OpenCV and PIL (Python Image Library) to extract RGB information from 7,000 movie trailers.
- Built a hybrid content-based model that had significant improvement compared with the baseline model.

Blockchain Transaction Simulator — THE GEORGE WASHINGTON UNIVERSITY
Blockchain — *September 2019 – October 2019*
- Implemented Blockchain algorithms which create new blocks for transaction by python.

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 an SVM-based machine learning model to predict the market trend and stocks' price. Achieved average 78% accuracy.

Spotlight 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 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 twitter.