

Steven Liu

Data Scientist

(415) 671-5701

San Francisco, CA

stevenheliu@gmail.com

<https://sdevinl.github.io>

PROJECTS

AirBnB Kaggle Challenge — *Python*

- Developed a regression model to predict the price of AirBnB listings in New York City
- Cleaned data by converting strings/arrays to floats
- Implemented lemmatization to clean and tokenize descriptions of AirBnB units
- Imputed missing square feet with average by neighborhood
- Tuned hyper-parameters with GridSearch Cross-Validation
- Developed and tuned 8 different models
- Placed top 25% of participants in test set

Climate Time Correlation — *C, Python, SQL, JavaScript*

- Programmed ESP32 device to record stats such as temperature and humidity with C onto PostgreSQL database
- Managed database records from multiple ESP32 devices with a total of over 1 million records accessible from cloud
- Created API with PyMySQL to write SQL queries in Python
- Implemented API in JavaScript with XMLHttpRequest to acquire records from PostgreSQL database
- Developed a map that contains the location of every ESP32 device in the database as a marker around the world
- Programmed the page to update records dynamically based on on-click and on-hover mouse behaviour from the viewer
- Visualized the map and data w/ Google Charts

NBA Rankings w/ GNNs — *Python*

- A Multiclass classification of NBA Team Rankings w/ Graph Neural Networks
- Used BeautifulSoup to webscrape player stats, schedules and team rosters data from www.basketball-reference.com/
- Implemented 3 Graph Convolutional Networks w/ PyTorch
- The classic GCN produced 75% accuracy w/ slow runtime
- GCN-LPA and GraphSAGE are variants to the classic GCN
- GCN-LPA 82% accuracy with w/ slow runtime
- GraphSAGE w/ 88% accuracy w/ much faster runtime and will scale to larger networks
- As a benchmark, Logistic Regression achieves 63% accuracy

EDUCATION

University of California - San Diego

Bachelor's of Data Science

2018-2021

Provost Honors 2020

PROGRAMMING LANGUAGES

Python ★★★★★

JavaScript ★★★★★

SQL ★★★★★

Java ★★★★★

R ★★★★★

TECHNICAL SKILLS

❖ Data Wrangling

- Pandas
- NumPy

❖ Data Visualization

- Matplotlib
- Seaborn
- D3js
- Highcharts
- ggplot

❖ Machine Learning

- SciPy
- Scikit-Learn

❖ Neural Networks

- PyTorch
- Keras/TensorFlow

❖ Database Management

- SQL

❖ Web Scraping

- BeautifulSoup

❖ Big Data

- Spark
- Dask
- AWS