Steven Liu — Software Engineer / Data Scientist New Grad

San Francisco, CA | (415) 671-5701 | steventheliu@gmail.com | GitHub | Portfolio Website | LinkedIn

EDUCATION

University of California, San Diego (2018 - 2021) – *B.S. of Data Science – Provost Honors*

PROGRAMMING LANGUAGES *(Libraries)*

- **Python** *(Pandas, NumPy, Matplotlib, Seaborn, SciPy, Scikit-Learn, Spark, Dask, Flask, BeautifulSoup)*
- **JavaScript** *(D3js, Highcharts, ggplot, GoogleCharts)*
- **SQL** *(PostgreSQL, MySQL)*

PROJECTS

AirBnB Units Price Prediction — *Python*

- Developed a regression model to predict the price of AirBnB listings in New York City
- Data wrangling with lemmatization and ETL to clean and tokenize descriptions and data of AirBnB units
- Imputed missing square feet data with average by neighborhood
- Tuned hyper-parameters with GridSearch Cross-Validation for 8 different machine learning models
- Applied individual creativity and problem-solving skills to place top 25% of participants

Global Climate Analysis — *C, Python, SQL, JavaScript*

- Programmed ESP32 device to record stats such as temperature and humidity with C onto PostgreSQL
- Managed records from multiple ESP32 devices with a total of over 1 million records accessible from cloud
- Built an API with PyMySQL to write optimized SQL queries from database in Python
- Implemented API in JavaScript with XMLHttpRequest to query records from PostgreSQL database
- Data visualization with interactive global map of every ESP32 device as a marker with plots dynamically

NBA Team Rankings w/ Graph Neural Networks — *Python*

- Implemented Graph Convolutional Neural Networks w/ PyTorch for classification of NBA Team Rankings
- Web Scraped data with BeautifulSoup for individual player stats, team schedules and team rosters
- GraphSAGE achieved 88% accuracy w/ much faster runtime and scalable to larger networks and datasets
- For comparison, classic GCN (75%), GCN-LPA (82%), and the benchmark Logistic Regression (63%)
- Time management for a 6 month long project to consistently communicate and collaborate with partners
- Deployed with docker containers and images to maintain a constant environment
- Prepared a research paper and presentation: https://sdevinl.github.io/Project-NBA-Rankings-Prediction/

RELEVANT COURSEWORK *(Skills): Description*

- **Data Structures & Algorithms** *(Java, Python, C++)* – Recursion, Higher-Order, OOP, complexity, and data types
- **Practice & Application of Data Science** *(Python, Git, Github)* – Fundamental spanning algorithms, statistics, machine learning, visualizations, and data systems on real world data.
- **Database Management** *(SQL)* – Relational database, schema design, query language and optimization
- **Systems for Scalable Analytics** *(Spark, Dask, AWS-EC2, MapReduce/Hadoop)* – Big Data, Memory hierarchy, distributed systems, model selection, and deployment of large datasets at scale.
- **Data Visualization** *(Tableau, JavaScript, HTML, CSS)* – Computer graphics, human-computer interaction, cognitive psychology, statistical graphics to create effective visualizations.
- **Modeling & Machine Learning** *(Python, R)* – Models, Natural Language Processing, and Robotics.