YUTAO ZHOU

https://yutao-zhou.github.io/CV/ https://www.linkedin.com/in/yutao-zhou/ yutaozhoucolumbia@gmail.com (805) 637-1617

EDUCATION

Columbia University New York City, NY

M.S. in Electrical Engineering GPA: 3.835/4.0 Expected Dec 2023

University of California - Santa Barbara: B.S. in Physics GPA: 3.67/4.00 **Dean's** Honors List in 2021 Winter. Sep 2018 - Dec 2021

SKILLS

Python, Java, HTML, CSS, JavaScript, React.JS, Django, Flask, FastAPI, MySQL, Neo4j, MongoDB, Solidity, Full-Stack Development, Socket Programming, Spark/Hadoop, Tensorflow, Git, GCP, AWS, Airflow, D3.js, NoSQL, REST APIs, TypeScript, MATLAB

WORK EXPERIENCE

Amazon(AWS): SDE Intern **Cupertino, CA** **May 2022 - Aug 2022**

- Working in a team that builds software to prevent and mitigate DDoS attacks on AWS infrastructure and customers.

Deepchem Co., Ltd: Python Intern **Beijing, China** **Feb 2022 - May 2022**

- Designed and built calculation task distribution systems. Distributing calculation jobs from the distribution server to different calculation servers (Group project, 4 people in total (including one manager)).
- **Communicated** and **collaborated** with front-end, and other co-walkers to create a web-based platform. **Represented team** to communicate with manager Finished building in 1 month.
- Checked job status on the platform and handled manual stop from user with **GET**. Handling exceptional cases e.g. distribution server offline. Stress tested on all 4 calculation servers.
- Check front-end job status and **submit log** content from calculation to the distribution server in **real-time** with **GET** and **POST**. Zip needed calculation results and uploaded files to the distribution server with POST (used for more than 10 jobs in business).
- **Increased** overall calculation **efficiency** by **50% - 200%** (By keeping calculation servers busy during nonbusiness hours).
- Created **algorithms** to **find missing tuples** in the database from id queries CSV. **Data filtering** and **aligning**. Extract 3D Cartesian coordinates and get SMILEs with **Pybel**(OpenBabel) python package.
- **Constructed** and **maintained SQL database**. Extract data from XYZ file, CSV file, and convert SMILE and insert it into SQL database(including **checking redundant** data in database).
- Developed an **algorithm** to automatically audit two-way connections between PC and lab equipment (**Heartbeat**) with **SOCKET** (Individual project finished in 1 day).

PROJECTS

Independent Project: Used Car Data Visualization WebApplication **Jun 2022 - Jul 2022**

- Built with **streamlit**, dealing **large data sets**(365K data points) with **Desk**, **Pandas**, and **NumPy** for data filtering and cache data.
- **Visualize data** with scatter plot on the heat map (with more than 100 selectable base maps), pie chart, scatter plot with the trend line, with packages e.g. **plotly**, **leafmap**, **pydeck**.
- Added VIN lookup function with Get from NHTSA (National Highway Traffic Safety Administration)'s **API**.
- Designed AI key phrase extraction from listing description with **spacy**, and visualization with **wordcloud** with VIN query results (VIN query, key phrase generation, and word cloud should take less than 5 seconds, usually 2 seconds).
- Implemented **geocoding** and filtering data with user input distance from user query location with **geoencoder** in **GeoPy** (Entire query should take 3 seconds depending on the setting, usually less than 0.5 seconds).
- Added Login page with the **cookie**. **Hosting** web applications on a personal server with **domain redirection**.

Full stack Course Project: NYC Subway Traffic Analysis **Oct 2022 - Dec 2022**

- Full stack **RESTful** web application that displays the entry and exit of each subway station on an interactive map.
- Write **Frontend JavaScript** that would let the user choose a different time with a slider. Then, corresponding data will be visualized on the map. A ranking of stations with the top 10 throughputs at the selected time will be displayed on the side.
- All data presented in the interactive map are fetched in real-time from the **backend REST API** written with **Python Flask**.
- The data was batched data from the MTA website and pre-processed with **Spark**.

Full stack Course Project: MBTI Personality Analysis and Prediction **Oct 2022 - Dec 2022**

- We used **Flask** as the backend and **HTML**, and **CSS** as the **frontend**. When a user enters their username we will fetch the user's Tweets using **Twitter API(tweepy)**. Then we would process fetched data and use the **pre-trained model** to make predictions. Then we would present corresponding results to the user.

Course Project: My Own Internet **Nov 2022 - Dec 2022**

- Configure **OSPF** and **iBGP** to connect 8 routers and 6 hosts in my **Autonomous System**.
- Configure **eBGP** to perform different routing policies for **inter-AS** connection with my **provider**, **customer**, and **peers**. e.g. Achieved **no valley routing**. Achieved **preferred customer routing**(preference in this order: customer, peer, provider). **Inbound traffic engineering**: prefers traffic coming from one link of a provider(that has multiple links). Guide traffic to prefer coming from one provider over others. Successfully fetched data across our internet(with working policy) formed with my classmates.