Jiahe Feng

j1feng@ucsd.edu | (484)-782-4770 | San Diego, CA | LinkedIn: jiahe-feng | j1feng.github.io

EDUCATION

Carnegie Mellon University 08/2022 - 05/2022

- Incoming Master of Computational Data Science at School of Computer Science.

University of California San Diego 08/2018 - 03/2022

- B.S. Data Science; B.S. Cognitive Science Specialized in Machine Learning. Minor in Japanese Studies.
- GPA: **3.93/4.00**; Provost Honor for 10 consecutive quarters; Warren College Honor Society.

SKILLS

Coursework: Data Structures, Algorithms, Discrete Math, Statistics, Data Management, NoSQL Database, Machine Learning, Recommender Systems, Deep Learning, Text Analytics, Distributed/Cloud Computing, Robotics, Signal Processing.

Programming: (Proficient) Python, Java, SQL, R, (familiar) HTML, CSS, JavaScript, C, C++, Matlab.

Framework/Tool: NumPy, Pandas, Scikit-learn, Matplotlib, PyTorch, Dask, Apache Spark, Selenium, React.js, Highcharts (js), Tableau, AWS, IBM Cloud, ArcGIS, PostgreSQL, MongoDB, Neo4j, TigerGraph.

Certifications: Cloud Application Development (IBM); Big Data, Predictive Analytics (UCSD); C/C++ Coding (UCSC).

WORK EXPERIENCE

UCSD School of Global Policy & Strategy | Research Assistant | La Jolla, CA 09/2020 - Present

- Building open-source **Python** package **Geodata** for earth system modeling with various geospatial datasets.
- Managed TB-scale datasets on **HPC** with **Dask** data parallelism, created modules that build off **Xarray**, Rasterio, and Shapely.
- Examined the PM2.5 change in China with **Matplotlib** animations and generated land capacity factor for solar/wind energy.
- Working as 2nd author for the 38th USAEE Conference presentation/paper on designing renewable energy plans in India.

The Scripps Research Institute | Data Science & Bioinformatics Intern | San Diego, CA 01/2021 - 03/2022

- Conducting genome-wide association studies using Bedtools, Regenie to better understand and prevent alcoholic diseases.
- Processed large data (500k+ participants, and 10k+ phenotypes and biosamples features) through Scripps **HPC** Jupyterhub.
- Performed **EDA** and statistical analysis in **Python** and **R** to explore correlations and extracted covariates for GWAS.
- Conducted survivorship analysis, use **SHAP** to calculate feature importance, and built **ML** classifiers for risk/behavior prediction.

HOVE Social Good Intelligence | Data Science Intern | San Diego, CA 06/2021 - 09/2021

- Researched on social impacts of corporations and evaluated ones that may meet the social good requirements.
- Performed data **ETL** by capturing public data with **Selenium** and **bs4** and creating **PostgreSQL** databases on **DataGrip**.
- Conducted sentiment analysis on social signals with **Azure** cloud service and **NLTK** in Python for customer feedback.

UC San Diego | Teaching Assistant | La Jolla, CA 09/2019 - 03/2022

- Taught 6 Quarters (~**225 students** per quarter) Created/graded student homework/exams, led discussions, and held office hours.
- Topics included OOP, unit testing, data structures, algorithms, complexity, A/B and statistical testing, text analysis, and ML.

PROJECTS/OTHER EXPERIENCES

San Diego Supercomputer Center | HPC Trainee, Research Volunteer

- High Performance Computing & Cyberinfrastructure Training (01/22 - 05/22): Hands-on training on Expanse supercomputer. Topics include parallel computing, cloud computing, containers, CPU/GPU computing and profiling, CUDA, deep learning.
- GeoAct Project (09/21 - 10/21): Supervised by Dr. Ilya Zaslavsky, helped build an **ArcGIS** dashboard and automate the update of cases and vaccination information in its **Python** API to provide reopening policy evaluation for schools in San Diego.

Actionable Recourse | Data Science Capstone Project

- Supervised by Dr. Berk Ustun, help users who were rejected from loans take actions to flip the ML prediction result.
- Examining different **Python** packages (recourse, SHAP, LIME) for loan default classifiers and judging the actionability.
- Developed **front-end** for an **interactive** web app that generates personalized action sets using **react.js**.

Amazon Product Review Analysis | Cloud Computing Project

- Stored 25GB of Amazon review data on **AWS EC2**, used **Dask** to perform feature exploration and data consistency check.
- With optimized runtime, used **Spark**'s **SQL** and **MLlib** libraries to create **Word2Vec** feature matrix and predict user rating.

Steam Game Play Recommendation | Recommender System Kaggle Competition

- Implemented NLP/similarity algorithms with review text to predict if users would play a game given 170k user-game pairs with few features in **JSON**. Accuracy ranked **top 13%** in the class competition (~400 undergrads & 250 graduate students).