Haibo Li

San Diego, CA | e: hal011@ucsd.edu | c: (858) 319-5227

EDUCATION

University of California, San Diego **06/2022**

- Bachelor of Science in Data Science **GPA: 3.80/4.00**
- Courses: *Principles of Data Science, Programming for Data Science, Theoretical Foundation of Data Science; Machine Learning Algorithms, Deep Learning, Data Structure and Algorithm, Database System*

SKILLS

- Programming: Python (Pandas, NumPy, PyTorch, Dask, PySpark, scikit-learn, Matplotlib); R; MySQL, SQLite3; Java
- Language: Chinese (native or bilingual), English (full-professional)

EXPERIENCE

Nokia Shanghai Bell Co., Ltd. Hangzhou Branch Co. **06/2019–08/2019**

Security System Intern

- Investigated data encryption algorithms that include DES, 3-DES, AES, AES-GCM, and HMAC; conducted literature reviews and presented workshops to senior engineers regarding the application and limitation of each algorithm.
- Implemented HMAC algorithm with Python in a data encryption protocol to support Nokia's engineering effort.

RESEARCH

Rose-Spatiotemporal Machine Learning Lab, UCSD **07/2021–10/2021**

Lab Researcher

- Led the research of a dynamic neural relational inference model that uses deep learning to map the trajectory of honeybees based on their interaction data; quantified information exchange mechanisms such as trophallaxis.
- Mitigated the noise and inconsistency within the raw data; constructed new features such as directional velocity.
- Benchmarked models that include Dynamic Multi-agent Relational Inference (DYARI), Neural Relational Inference (NRI), and Dynamic NRI models in PyTorch; tweaked models and parameters to determine optimal model fitting.
- Achieved 40.8% less mean square error in the improved DNRI than NRI, and 100 times less time than DYARI.

Rating Prediction for Restaurants on Yelp, UCSD **12/2020**

Course Project

- Implemented TF-IDF, similarity functions, and recommender systems to predict the rating of a restaurant on Yelp.
- Analyzed over 30K Yelp reviews to construct a regression model to recognize differences in content and reviewers.
- Used WordCloud to visualize the reviews of restaurants with different ratings; extract keywords to generalize the reviews into 1 to 5 star ratings; identified the 40 most important variables using Pearson correlation.
- Achieved an improvement 13% in prediction accuracy at 72.4%, compared with the baseline model.

Predict the Cost of Political Campaign Ads on Snapchat, UCSD **06/2020**

Course Project

- Designed a prediction model based on random forest regression to estimate the cost of political ads on Snapchat.
- Performed ETL tasks on 20,000 data points; introduced features to lower bias and boost model performance.
- Conducted PCA on over 20 features and tuned model parameters using GridSearchCV to improve model accuracy.
- Assessed missingness of data by identifying Non-missing at Random (NMAR), Missing at Random (MAR), or Missing Completely at Random (MCAR); tweaked the model to achieve over 70% in prediction accuracy.

PATENT

Method, Equipment and System Model to Locate Motion Trail of Cellphone Users (ID: 2016105550295) **09/2017**

HONORS

AP Scholar with Distinction Award, College Board **07/2017**

2017 Award for Excellence (only recipient), National High School Media and Communication Competition **06/2017**

Global 31, National 11, Division 1, American Regions Mathematics League **04/2017**