Zhaoyi Hou (Joey)

E-mail z9hou@ucsd.edu **Phone** 858-729-8929

Education

09/2017 - present

University of California, San Diego

- Major: Data Science (B.S.)
- Minor: Communication
- Selected Courses: Web Mining and Recommender Systems, Systems for Scalable Analytics, Deep Learning for Natural Language Understanding, Advanced Data Structure, and Interaction with Technology

Experience

06/2020 - present

Research Fellow - *Data Mining Lab@UCSD*

Text Mining for Patent Approval Prediction

- Implemented a customized **Bert** model and an **LSTM** model to predict whether a patent application will be approved;
- Improved accuracy **from 60% to 85%** at the abstract level.

07/2019 - present

Data Scientist Intern - *Salk Institute for Biological Studies*

Data Analysis for a Circadian Rhythm Studies

- Formatted data (up to **10,000 records**), conducted **hypothesis tests**, and **visualization** for two circadian rhythm projects;
- Built an **analysis pipeline** in Python for food logging data from a mobile app (up to **500,000** records);

02/2019 - 11/2020

Research Fellow - *UC San Diego Shiley Eye Institute*

Computer Vision for Medical Purposes

- Built an eyelid measurement pipeline for patient facial images with **OpenCV**;
- Achieved a success rate of **80%** for measurement result within a **0.5 mm** margin.

03/2018 - 08/2019

Data Structure Tutor - *Halıcıoğlu Data Science Institute, UCSD*

Intro to Data Structure & Principle of Data Science

- Led individual tutor hours to help students on homework and write tests for homework.

Projects

07/2020 - present

Food Parser

An Open Source Text Processing Project

- Automatically Correct **typo** and **extract food & beverage phrases** from user's input text with **85% parsing success**.
- **Open-source** project available for nutritional analysis studies.

11/2019 - present

Machine Learning for Ophthalmological Diagnosis

Awarded UCSD Halıcıoğlu Data Science Undergraduate Scholarship

- Built an **image classification pipeline** for common eye diseases pre-diagnosis with **CNN in PyTorch**;
- Currently achieved **75%** accuracy for distinguishing involutional ptosis, thyroid eye disease, and normal eyes.

04/2019 - 06/2019

Beijing Housing Price

A Data Analysis & Machine Learning Project

- Analyzed more than **300,000 pieces** of house data with hand-crafted features;
- Built an **XGBoost** model for **housing price evaluation**.

Skills

Machine Learning

Decision Trees, Perceptron, Boosting, PyTorch, K-Mean etc.

Natural Language Processing

BERT, LSTM, Transformer, etc.

Data Analysis

Hypothesis Tests, Bootstrap, Panda, SciPy etc.

Data Structures

Binary Tree, KD Tree, Heap, Hash Map, Graphs, Priority Queues, etc.

Data Visualization

Seaborn, D3, JavaScript, CSS, HTML etc.

Leadership

President - UCSD Chinese Computer Community (Triple C)

- Project-based student community with more than **100 members**;
- Designed the onboarding **technical training** for data science members;
- Implemented the **pipeline** from idea brainstorming and user research to developer recruitment, development, and product marketing.