Ethan Zhou
Data Science @ UCSD

San Diego, CA | xethanhzhou@gmail.com | Github: @ethannhzhouu | LinkedIn: @ethannhzhouu

EDUCATION

University of California, San Diego   September 2019 - June 2023
B.S. Data Science; Minor: Cognitive Science
Provost Honors; Cumulative GPA: **3.5**

Udemy   June 2021 - August 2021
Data Science and Machine Learning Bootcamp

COURSEWORK

- SQL/NoSQL Databases
- Natural Language Processing
- Data Visualization
- Systems for Scalable Analytics

- Advanced Data Structures and Algorithms
- Object Oriented Algorithms and Design
- Machine Learning Algorithms
- Data Mining & Web Scraping

PROJECT EXPERIENCE

Movie2Watch   November 2021 **-** December 2021
Python, HTML, CSS, Flask
- Implemented a collaborative-filtering movie recommendation system based on **100,000** movie user ratings
- Utilized Cosine similarity as the metric to measure the similarities between user-inputted movie and recommended movies

Amazon Reviews Rating Prediction   November 2021 - December 2021
Python, Machine Learning
- Performed feature engineering using **bag-of-words** model on review text and one-hot encoding on review date
- Developed a **XGBoost** classification model to predict user ratings, improving F1 score by **4%** from baseline model

NYPD Allegations Dataset Demographic Prediction   April 2021 - May 2021
Python, Machine Learning
- Feature engineered (one-hot encoding, data imputation) to transform data and improve predictive model accuracy
- Implemented decision-tree classification model with GridSearchCV in **SKLearn** to predict complainant gender
- Improved model's accuracy score by **5%** by utilizing random forest classifier and hyper-parameter tuning

Movie Popularity and Revenue Analysis   January 2021 - March 2021
Python
- Analyzed the positive correlation between Google search movie popularity and movie revenue in **Jupyter Notebook**
- Scraped dataset of **750+** movies from BoxOfficeMojo and Google with **BeautifulSoup** and **Pytrends**
- Applied the **Matplotlib**, **Numpy**, and **Seaborn** libraries to create **9** data visualizations
- Developed linear regression model and conducted variable relationship analysis with **Patsy**

SKILLS

Programming Languages: Python, SQL, R, Java, MATLAB
Frameworks & Technologies: SKLearn, Pandas, Tensorflow, Dask, Spark, PostgreSQL, AWS
Industry: Git, Unit Testing, Tableau, Excel

WORK EXPERIENCE

Burlingame Country Club   May 2019 - September 2019
Server
- Collaborated with chefs and managers through peak business summer season to maintain customer satisfaction
- Conducted efficient ordering system to deliver **2000+** orders and serve **100+** members

ORGANIZATIONS

UCSD eSports   October 2019 - March 2021
Collegiate Esports Athlete
- Joined UCSD's League of Legends Varsity Collegiate team as a top **1000** player in North America
- Placed **4th** in the Riot Games 2020 Western Conference tournament