

ANDREW WU

DATA SCIENTIST

✉ andywuswc@gmail.com
☎ (858) 757-3375
📍 San Francisco, CA
94016
in andrew-wu-939746171/
🔗 andywzz?tab=repositories

Skills

LANGUAGES

Python
SQL
Pandas
Scikit-Learn
NumPy
Matlab
Streamlit
Heroku

MACHINE LEARNING

Regression
Classification
Clustering
Natural Language
Processing (NLP)
CoREX
SpaCy
Neural Networks
Keras
Tensorflow
Transfer Learnng

DATA MANAGEMENT

Git
PostgreSQL
MongoDB
Bash
MySQL
Spark

Education

University of California San Diego

B.S Chemical Engineering 2020

Experience

Humanity Forward Arizona Data Team Lead

Phoenix, AZ
July 2020 to Current

- Work closely with local politicians and a team of 50 volunteers to optimize campaign outreach using data science and data analysis
- Creating machine learning models to do topic modeling and sentiment analysis on twitter data relevant to the campaign.
- Using several dimension-reduction and resampling techniques to deal with the large amount of data effectively.
- Visualizing campaign statistics and making presentations to convey findings in a concise and effective manner.

Metis Data Scientist

San Francisco, CA

Completed an immersive 12-week data science bootcamp with a strong emphasis on end-to-end project oriented skill-building in supervised/unsupervised machine learning, statistical modeling, data wrangling and data visualization through Python. See projects section.

Turning Point Therapeutics Chemistry Intern

La Jolla, San Diego
June 2018 to Sept. 2018

- Performed lab maintenance work(washing lab equipment, making solutions, consolidating waste, refilling solvents, etc.)
- Assisted the researchers on experiments(running NMRs offsite, organic synthesis, purification and characterization of compounds, performing data entry, query and cleaning).

Projects

Amazon Recommender System

- Built an item recommender system for Amazon electronics section using a hybrid method to avoid the “cold start” problem of traditional recommender systems
- Downloaded 2.6 million item metadata and user review data and processed it using Jupyter Notebook on Google Cloud Computing
- Using the Scikit-Learn package, 6 different supervised machine learning model(Naïve Bayes, XGBoost, Decision Tree, Random Forest, Logistic Regression, K-Nearest-Neighbors) were implemented
- Built an unsupervised machine learning model using Singular Value Decomposition.

Humanity Forward

- Scraped ~10000 tweets using the official twitter API related to Andrew Yang from Arizona and stored it into a MySQL database
- Using NLTK, a pipeline for pre-processing text data was built. This included a tokenizer, lemmatizer, and TF-IDF Vectorizer
- Performed time-based sentiment analysis using TextBlob
- Performed topic modeling using None-Negative Matrix Factorization(NMF).

Wildventures

- A photo-based travel recommender system using Convolutional Neural Networks and Transfer Learning
- Webscraped 87 countries, over 6000 photos + descriptions.
- Labeled the photos to 14 distinct landscape categories.
- Used Transfer Learning to train a pre-trained ResNet50 model on the photos and achieved a training accuracy of **87 percent**.
- Built a web app using **Streamlit + Heroku**.