

YOUNG SUH

Experience

Metis Data Scientist

Completed an immersive 12-week data science bootcamp with a strong emphasis on project-oriented skill-building in problem solving, data wrangling, statistical modeling, machine learning, and communication of deliverables. **Project highlights include:**

Remote / CA
Jan. 2021 to Mar. 2021

New York Turnstile Data Analysis

- Analyzed the time series turnstile usage data of New York subway stations, for the client to maximize the email collection for the annual gala.
- Used Matplotlib and Seaborn to visualize the turnstile usage.
- Utilized Pandas module to determine the time and location of station with highest foot traffic.
- Implemented Autoregressive Integrated Moving Average (ARIMA) model to predict turnstiles usage in between provided time frames.

Wine Quality Prediction

- Scraped data from wineenthusiast.com using BeautifulSoup and Scrapy to predict the quality of wine using features provided.
- Utilized Scikit-learn module to perform feature engineering and feature extraction.
- Applied Lasso and vanilla linear regression models to predict the wine rating.

Spotify Music Genre Classification

- Predict the genre of music between 15 different classes using characteristic of features provided by Spotify API.
- Used PostgreSQL to store and query data needed.
- Classified genres of music using XGBoost Classifier with 0.67 F1 score.

Subreddit Post Analysis

- Analyzed subreddit group /r/AmTheAsshole in order to classify between the asshole group and non-asshole group.
- Performed natural language processing using SpaCy, NLTK, and Scikit-learn.
- Performed topic modeling using Tf-Idf and was able to get 7 topics that were well defined.

Reci-py Recommender: Minimizing Food Waste

- Recommend recipe based on presented list of ingredients and user preference, in order to minimize food wastes and to have a decent meal.
- Implemented hybrid of content-based and collaborative-based filtering system to recommend recipes.
- Utilized Surprise recommender to evaluate and fine tune the best collaborative-filtering system.

Clinomics USA Inc Head Computer Engineer

La Jolla / CA
July 2019 to Aug. 2020

- Performed data analysis on DNA dataset mainly using Python to identify the correlation between genotypes and sequence mutations.
- Developed web scraping tools to automate data gathering extraction for animal DNA mutation using Python.
- Analyzed Covid-19 sequences to design primers and probes to develop Covid-19 testing kit.
- Collaborated with branches in Korea and London to restore data and maintain servers.

UCSD Computational & Cognitive Neuroscience Lab Student Research Lab Assistant

La Jolla / CA

- Preprocessed facial image data before loading on computation models using Python.
- Devised effective normalization techniques on facial image data to create AAM that reconstructs 3D facial structure from 2D image data.
- Assisted setting up and designing web based social experiments using AWS, cloud computing EC2, to collect facial feature data.

Undergraduate Project

EEG Transfer Learning · <https://bit.ly/3gk7J0L>

- Evaluated transfer learning effectiveness between 3 different deep learning models and different numbers of training subjects, in order to apply BCI over shorter period of time to end users.
- Identified signals that can be used in models with less noise using EEG data from Kaggle.
- Utilized TensorFlow and Python to train and evaluate training and testing accuracy.
- Used matplotlib to visualize EEG data, and the train and test accuracies of models.

Contact

✉ yahoyoung@gmail.com

☎ 909.646.2007

📍 Rancho Cucamonga, CA

in www.linkedin.com/in/young-suh

🔗 [yahoyoung](https://github.com/yahoyoung)

Education

University of California, San Diego

B.S. Cognitive Science, Specialization in

Machine Learning and Neural

Computation 2019

Minor Mathematics 2019

Skills

PROGRAMMING LANGUAGES & LIBRARIES

Python

Java

Matlab

SQL

Pandas

Numpy

Seaborn

BeautifulSoup

Selenium

SpaCy

NLTK

Gensim

Surprise Recommender

PySpark

OpenCV

MACHINE LEARNING / DEEP LEARNING

Classification

Natural Language Processing

Neural Networks

Regression

Clustering

Recommendation System