CRYSTAL HUANG

DATA SCIENTIST

✉ crystalhuang.ds@gmail.com
🌐 crystaldatasy.medium.com
📞 (917)698-5471
📍 Newport, RI 02840
in crystal-huang-ds
⌨ crystal-ctrl

SKILLS

PROGRAMMING LANGUAGES
Python
SQL
HTML
R

MACHINE LEARNING
Regression
Classification
Natural Language Processing
Cross Validation
Model Selection
Neural Network
Dimensionality Reduction
Computer Vision

VISUALIZATION
Tableau
Excel/PowerPoint
Matplotlib
Seaborn
Bokeh
Plotly
pyLDAvis

LIBRARIES
Scikit-Learn
NLTK
SpaCy
Keras/ TensorFlow
OpenCV
langdetect

DATABASES
MySQL
MongoDB
SQLite

DATA WRAGLING
Pandas
Numpy
PySpark

WEB SCRAPING
BeautifulSoup
Selenium

OTHER
Git
Streamlit
Flask
Dash
Google Colab
API

EXPERIENCE

Metis
Data Scientist
Remote
Mar. 2021 to July 2021

Completed an intensive 14 week accredited Data Science Bootcamp with a strong emphasis on Python programming, project design, problem-solving, data wrangling, statistical modeling, data visualization, analysis, machine learning, and communication of deliverables. Projects include:

Exploring the Vax Potential
Engineered an end-to-end data collection, storage, and processing pipeline that creates a dashboard to visualize Virginia COVID-19 case reports and vaccinations with real-time data. Leveraged API for data acquisition, SQLite for data storage, Plotly for visualization, and cron job for automation of daily data collection. Created an interactive dashboard web app with Dash for users to visualize real-time Virginia COVID-19 data.

If You Must Mask
Developed a face mask detection system using 12k face mask images dataset from Kaggle. Leveraged pre-trained Haar cascade classifier (OpenCV) to identify faces in the pictures and applied custom trained mask detector with convolutional neural network (Keras/TensorFlow) and cloud computing (Google Colab) to classify face images and achieved a model accuracy of 99%. Built an interactive web app (Streamlit) for users to upload pictures and use a webcam to identify mask usage.

Covipedia
Leveraged Natural Language Processing to analyze 260k COVID-19 research articles for topic modeling and building a recommendation system for scientists to navigate the current surge of COVID-19 research articles. Used langdetect, NLTK, spaCy, and gensim for text preprocessing, performed topic modeling (LDA) to identify the common traits among research articles, and visualize topics with pyLDAvis. Created an interactive web app (Streamlit) for users to explore research articles based on topics.

Success is No Accident
Created a supervised classification model (XGBoost) using Kickstarter campaign dataset featuring 870k campaigns to predict the success of a Kickstarter campaign. The model performed with an F1 score of 80% and was implemented in an interactive web app (Flask) for users to explore the model's prediction. Findings from the interpretation of the model provided recommendations for campaign creators to boost their chances for a successful campaign. Supervised algorithms included KNN, logistic regression, decision tree, random forest, Naive Bayes, and XGBoost.

Ratatouille Radar
Analyzed NYC restaurant inspection datasets and 311 call data of NYC rodent sighting reports to provided recommendations of restaurants to inspect for the NYC Health by prioritizing high-risk restaurants for inspection. Leveraged SQL for data storage of over 26m data points. Created an interactive dashboard (Tableau) for users to explore.

The Roger-Ebert Bot
Built reusable python script (BeautifulSoup, Selenium) for efficiently web-scrape data from the movie critic and IMDb website featuring 7k movies. Implemented linear regression, polynomial, Ridge, LASSO algorithms to predict renowned late movie critic's movie ratings.

MTA Traffic Analysis for Station Accessibility Upgrade
Analyzed the MTA dataset from 3/21/20 to 3/20/21 and built a reusable python script for efficiently gathering large amounts of MTA turnstile data and multiple visualizations to help gain insight into the subway traffic pattern. Provided recommendation of subway accessibility upgrade placement and work schedule for employees to minimize risk during the pandemic.

TridentCare
Lead Radiology Technologist/Sonographer
Virginia Beach, VA
Aug. 2020 to Jan. 2021

- Supervised operation to increase workflow efficiency and achieve customer satisfaction
- Designed feedback report that increased staff compliance with new COVID guidelines
- Developed and Organized spreadsheets to solve operational problems and improve communication among staff

Sonographer/Echocardiographer
Virginia Beach, VA
Aug. 2018 to Aug. 2020

- Mobilized to facilities and residences for performing various ultrasound testing procedures and echocardiograms
- Set up and adjusts all equipment accordingly in various settings and environments.
- Assisted in equipment and vehicle maintenance
- Assisted the patient during exams and positioned the patient
- Provided clinical instructions of other healthcare professionals

EDUCATION

Rutgers University
Sept. 2011 to May 2015
B.S. Exercise Physiology; Minor in Psychology 2011

Western Suffolk BOCES
July 2016 to June 2018
Certificate Diagnostic Medical Sonography 2018