Available and Hope to Start Immediately

David J. Thorne / Thorneinsight@gmail.com / 860-485-5370 / https://github.com/djthorne333

Objective

I am a recently graduated physics Master's/former PhD student making the exciting leap into data science.

Education

University of California Santa Cruz, 1156 High St, Santa Cruz, CA 95064

- Graduated with Master's from physics PhD program (Spring 2021).

Central Connecticut State University, New Britain, CT 06053

- Graduated winter 2019; GPA: 3.69/4.0
- Bachelor of Science in Physics, Bachelor of Arts in Mathematics, Numerous Mechanical Engineering Credits.

Skills

- Python (NumPy, Pandas, MatplotLib, Seaborn, Scikit-learn, Pytorch, Mlxtend, Flask), R (Ggplot, Dplyr, Tidyr), SQL.
- Exploratory analysis, Frequent patterns, NLP, Classification (Regression, Clustering, Knn, Neural networks, Random Forest, Parameter optimization, Text), Web-Scraping (Selenium), deployment (Flask/Heroku).
- Proficiency in all Microsoft Office applications, Entry level Matlab.
- Technical writing, oral presentations of research as poster or PowerPoint, team-oriented research, former university physics lab instructor and teacher's assistant.

Experience (GitHub Link: https://github.com/djthorne333)

EDA of Professional (CS:GO) Gamer's Gear and Settings, and Modeling Player Accuracy Performance

- Exploratory data analysis via visualizations and statistics to describe trends in what gear/settings pro players prefer, and modeling whether a player has above average aim based only on their gear/settings.
    - Managed an increase in model classification accuracy of 17% (to 67%, without overfitting) over raw web-scraped data through feature engineering (binning, frequent patterns, clustering, separation by player role), filter methods (chi2 tests with class target), wrapper methods (sequential feature selection) and parameter optimization (Gridsearch). *(Python, NumPy, Pandas, MatplotLib, Seaborn, Scikit-learn, mlxtend, EDA, classification, regression, clustering, knn, feature engineering).*

NLP-CNN-Subreddit-Sorter-Application

- End-to-end development of an application that suggests to users/moderators which subreddit (of similar topics) a reddit post belongs to according to its title. The subreddits chosen for training are all technical with shared content, and the app could benefit users/moderators interested in data science and related fields.
    - Managed an increase in model classification accuracy of 10% over glove embeddings through use of custom word2vec embeddings, and quickly found optimal number of filters to use for CNN through novel method. *(Web-scraping, exploratory analysis, feature engineering, custom word2vec embeddings, convolutional neural network, text classification, Pytorch, deployment:Flask/Heroku).*

Exploratory Data Analysis of Stroke Dataset in R

- Exploratory data analysis of a Kaggle stroke dataset using Rstudio *(R, Ggplot, Dplyr).*

Relevant Coursework

- CSE 242: Data Mining / CSE 243: Machine Learning (Theory, Stats, EDA, Classification, Deep Learning)
- Math 218: Discrete Mathematics / Math 377: Intro to Real Analysis / Math 366: Intro to Abstract Algebra
- ENGR 240: Computational Methods for Engineers (Matlab and advanced functions of Excel)