David J. Thorne / thorneinsight@gmail.com / 860-485-5370 / Plymouth, CT
My GitHub https://github.com/djthorne333 / My LinkedIn https://www.linkedin.com/in/david-thorne-7b2216228/

Summary

I am a recently graduated physics Master's student with full stack analytics and ML experience.

Education

University of California Santa Cruz, Santa Cruz, CA 95064

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

Central Connecticut State University, New Britain, CT 06053

- B.S in Physics, B.A. in Mathematics, Graduated Jan 2019; GPA: 3.69/4.0

Skills

- Python (NumPy, Pandas, MatplotLib, Seaborn, Scikit-learn, Pytorch, Statsmodels, Mlxtend, Flask, PyMysql), R (Ggplot, Dplyr, Tidyr), SQL, Docker Containerization, html/CSS, Git, Power BI, Excel.
- AWS Cloud Computing (EC2, RDS, S3), Boto3, Linux CLI.
- Web-Scraping, Databases, Exploratory Analysis, Dashboards, Classification, NLP, Time Series, Deployment.
- Technical writing, oral presentations of research, team-oriented research, former university TA lab instructor.

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

Full Stack AWS Self Updating Time Series Weather Forecast Dashboard http://3.236.77.132:8080/

- A full stack SARIMA/SVAR weather forecasting application done entirely within AWS cloud computing architecture (EC2, RDS, S3) using Boto3, PyMySQL, Docker and Flask. I compare SARIMA and Seasonal VAR time series models as weather forecasters. It is a completely autonomous pipeline (via crontabs) that extracts/transforms data (API calls), updates/pulls from the RDS database, trains weekly models, saves forecast visuals to an S3 bucket, and updates the weekly dashboard via Flask/gunicorn.
    - *(ETL, Frontend/Backend, AWS, EC2, RDS, S3, Boto3, PyMySQL, Docker, SARIMA, VAR, html/CSS).*

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

- Exploratory data analysis 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, EDA, feature engineering, classification, regression, clustering, knn, random forest).*

NLP CNN Subreddit Sorter Heroku App https://datascience-reddit-post-sorter.herokuapp.com

- End-to-end development of an app using a neural network that suggests to users/moderators which subreddit a reddit post belongs to according to its title. The subreddits chosen for training are all technical with similar content, and the app could benefit reddit 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, Flask/Heroku, html/CSS).*

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)

Check out my Physics CV and Research: https://github.com/djthorne333/Physics-publications-posters-cv