

## **Summary**

Physics Researcher with Data Engineering, Analytics, and Machine Learning 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, Jupyter, Docker Containerization, html/CSS, Git, Power BI, Excel, Statistics.
- AWS Cloud Computing (EC2, RDS, S3), Boto3, Linux CLI, Flask, Heroku, frontend, backend, ETL.
- Web-Scraping, Exploratory Analysis, Visualizations, Classification, NLP, Time Series, Deployment.
- KNN, Clustering, Random Forest, Regression, Regularization, Lasso, CNN, NLP, SARIMA, VAR, Apriori.
- Technical writing, oral presentations of research, team-oriented research, former university TA lab instructor.

## **Work Experience** (Physics CV: <https://github.com/djthorne333/Physics-publications-posters-cv> )

### **Teacher's Assistant (2019 – 2021), University of California Santa Cruz**

- As a teacher's assistant I taught physics labs and graded exams.

### **Student Independent Research (2017- 2018), Central Connecticut State University**

- Synthesized nanoparticle supercapacitor electrode materials and performed characterizations.

### **REU Intern (June 2016-August 2016) SRI International Laboratories, Menlo Park, CA**

- Operated vacuum systems and performed analyses on spectrometer data for heat shield simulations.

### **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.

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

- Exploratory data analysis with visualizations 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 text 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).

### **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.

### **Exploratory Data Analysis of Stroke Dataset in R**

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