Winston DeGraw I Data Scientist

Seattle, WA · (818) 635-0013 · winston.degraw@gmail.com · Portfolio · GitHub · LinkedIn

SUMMARY

I am a data scientist with a background in physics who is eager/determined to combine my academic experience in working with data as a research assistant with my data skills in machine learning and data pipelines to help companies in climate/green energy or other scientific application industries achieve their goals. Additionally, my experience as a sailor has taught me how to work effectively under pressure and collaboratively across dynamic teams.

SKILLS

Skills: Python, SQL, Jupyter, Git, Scikit-Learn, Numpy, Pandas, Matplotlib, Seaborn, AWS, Machine Learning, Tensorflow, Keras, PyTorch, Computer Vision, Natural Language Processing, Regression models (Linear, Ridge, Lasso), Classification models (Logistic, K-Nearest Neighbors, K-Means, Decision Trees, Random Forests), Docker

EXPERIENCE

Data Science Fellow I General Assembly I Remote 11/2022 - 02/2023

Successfully completed 500+ hours of expert led instruction in machine learning, data visualization, and data engineering. Demonstrated ability to quickly learn new tools and use them effectively in classwork. Communicated findings for both technical and non-technical audiences. Developed projects, including:

- **Event Reconstruction in He6-CRES with Faster R-CNN** | GitHub Repo
    - Aimed to test the feasibility of using R-CNN deep learning models for reconstructing events in the He6-CRES experiment at the University of Washington.
    - Adapted an existing simulation and wrote a PyTorch implementation of a Faster R-CNN model to train (on an AWS GPU instance) and lay the groundwork for future systematic studies into such a model's performance.
    - Achieved a mean Intersection over Union (IoU) of 0.8, demonstrating that these types of models are applicable to the data seen in the experiment and worth further examination.
- **Classifying Excess Low Birth Weight Incidence with Air Quality Data** | GitHub Repo
    - Group project that examined the effects of air quality on excess rates of low birth weight (LBW) in infants.
    - Scraped 15 years of EPA data from the Air Quality Service API, and 15 years of data from the CDC's WONDER system to build a classification model that determined on a county level whether or not an excess LBW rate would be seen based on pollutants measured in the prior year.
    - Neural Net classifier produced results with accuracy of 75%, showing there is a connection in the measured pollutants and LBW, though no exact insights as to the most important pollutants could be extracted with this model type.
- **Subreddit Post Classifier with NLP** | GitHub Repo
    - Used PushShift API to scrape ~15k posts from r/NoStupidQuestions and r/AskReddit to build an NLP classification pipeline to distinguish between the two subreddits.
    - Compared several types of models including Random Forest, Naïve Bayes, and Logistic Regression, grid searching over hyperparameters to find the best performer.
    - Random Forest performed best with an accuracy of 77% - more than 25% improvement over baseline, and determined most common words/types of questions to allow for distinguishing between the two subreddits

Research Assistant I University of Washington (He6-CRES) I Seattle, WA 05/2021 - 11/2022

- Developed and implemented a radiation detector broadcasting real-time measurements via ethernet to be used in analysis of data in experiment.
- Assisted in adapting C++ analysis framework for study of spectrogram images, alongside DBSCAN clustering of track segments inside of the images.
- Implemented analysis pipeline on computing cluster with Docker container for much faster/automated analysis.
- Operated and maintained a unique experiment involving high vacuum systems, superconducting magnets, and radioactive gasses in a dynamic team environment.

Research Assistant I Lawrence Berkeley National Lab I Berkeley, CA 02/2018 - 08/2020

- Used supercomputing cluster to perform Monte Carlo simulations of detector response to guide design of tracking detectors in a future Electron Ion Collider.
- Part of a team that built 60 detector units that form the middle two layers of the Inner Tracking System at CERN's ALICE experiment.

EDUCATION

Certificate, Data Science **Remote**
General Assembly *11/2022 - 02/2023*

Master of Science, Physics **Seattle, WA**
University of Washington *09/2020 - 11/2022*
GPA: 3.89

Bachelor of Arts, Physics **Berkeley, CA**
University of California, Berkeley *08/2015 - 12/2017*
GPA: 3.84

Associate of Arts, Math **Woodland Hills, CA**
Los Angeles Pierce College *08/2011 - 06/2015*
GPA: 4.00