Last Updated on 29th April 2022

Margaret Sabelhaus

Data Scientist | Denver, CO

margaretsabelhaus.com msabelhaus msabelhaus margaretsabelhaus@gmail.com 3015252903

Experience

HITACHI ENERGY | Data Acquisition Development Intern

May 2021 – Dec 2021 | Broomfield, CO

- Integrated multi-threading code to production data scrape acquiring up to 12,000 rows of data daily for each of over 16 reports.
- Saved more than 5 person-hours/week by using Python's BeautifulSoup module to scrape California ISO data based on HTML structure.
- Optimized in-house energy demand model by converting from R to Python, reducing daily runtime by 11 minutes.

BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM | Research Assistant

June 2018 – June 2020 | Washington, DC

- Routinely used R, SQL, and command-line interface to manipulate large financial datasets (over 1,000,000 observations).
- Designed and implemented a model of daily reserve volume in R and maintained year-out forecast data in PostgreSQL.
- Oversaw a team of four to restructure the Money Market briefing book and distributed it to the head of the Monetary Affairs division before FOMC meetings.

Selected Projects

ENERGY DEMAND FORECAST FOR CONNECTICUT ISO ZONE | Python

2021

- Combined and cleaned 200,000+ rows of data on hourly average temperature and energy usage gathered from Oracle database.
- Developed a Long Short-Term Memory neural network in Tensorflow to predict next hour's energy demand for Connecticut Independent Systems Operator region, outperforming the mean absolute error of baseline model by 72%.

PREDICTING SUCCESS OF PLANNED GENERATING UNITS | Python

2021

- Aggregated and prepped SQL data for over 20,000 generating units.
- Employed grid search algorithm to optimize a random forest model in Python in pursuit of classifying construction success of planned generating units.
- Confirmed business suspicions and communicated findings to Hitachi Energy stakeholders.

CLASSIFYING OPTIMAL LOSSY COMPRESSION OF CLIMATE DATA | R, Python

2021

- Feature engineered 20+ variables to explain variability in optimal compression level of NCAR climate data.
- Constructed a multi-stage classification model using a boosted tree and support vector machine in R, producing the highest overall accuracy across all considered models.
- Collaborated with peers to report and present findings.

Skills

SOFTWARE

Python • R • SQL • Git/Github • Agile • • LaTeX• Command-Line • SAS • MATLAB

AREAS OF EXPERTISE

Machine Learning • Probability and Statistics • Data Analytics • Data Visualization • Production-Ready Code Development • Web scraping • Data Mining

Education

COLORADO SCHOOL OF MINES

Master's in Data Science

Expected May 2022 | Golden, CO

THE PENNSYLVANIA STATE UNIVERSITY

Bachelor's in Statistics and Economics

Aug 2014 - May 2018 | State College, PA

Volunteer Work

HOWARD UNIVERSITY

Project Coordinator/Teaching Assistant

Jan 2019 - May 2020 | Washington, DC

Coursework

Machine Learning
Statistical Methods
Linear Algebra
Time Series Analysis
Stochastic Modeling
Probability Theory
Survey Sampling