MARCUS BAMBERGER

DATA SCIENTIST

marcus@marcusbamberger.com 206-495-5053 marcus-bamberger-78a609170/ hedgehog612

SUMMARY

Aspiring data scientist with a background in environmental modeling and large dataset management. Experience working with machine learning methods and clear, direct data visualization in Python. Looking for further opportunities to contribute to data projects while continuing to grow my skills and experience.

SKILLS

LANGUAGES: Python, R, SQL
PLATFORMS: Jupyter, Anaconda, WebStorm, GitHub
MACHINE LEARNING: Random Forests, Naive Bayes
MODULES: Matplotlib, Scikitlearn, Rayshader, WRF-Chem, ISORROPIA II
FUNCTIONS: Search, Pattern-finding, Data cleaning, Data refinement
DATA: Modeling, Visualization, Tracking, Outlier handling
LARGE DATASETS: Interpretation, Management, Curating, Visualization
INSTRUCTION: Building procedures, Preparing documentation, System presentation, Visualization
OPERATING SYSTEMS: Mac, Windows, Ubuntu (Linux)
PACKAGES: Pandas, Matplotlib, Seaborn, Excel, Word

EDUCATION

Springboard Data Science Immersive · Jan. 2020 to July 2020
6 month intensive course in data science, machine learning, Python, SQL. Capstone Projects: https://github.com/Hedgehog612

Reed College · Sept. 2015 to May 2019
Bachelor of Arts 2019
Coursework: Calculus, Analytical Chemistry, Computational Biology, Economics, Statistical Thermodynamics, Organic Chemistry, Inorganic Chemistry, Environmental Chemistry

PROJECTS

Capstone Project 2: NYC property valuation
Mar. 2020 to Current
- Capstone Project 2: Predicted New York City property valuations from NYC OpenData dataset
- Curated 5.7 million data points, 110 columns of data
- Build Random Forest, SGD regressor models to predict final valuations
- Modeling with Matplotlib, Folium to map NYC data
- Classification model: .83 score on test data with minimal (<2%) overfit

Capstone Project 1: Poll Prediction
Jan. 2020 to Mar. 2020
- Capstone Project 1: Predicted election outcomes from initial polling using FiveThirtyEight data
- Examined over 10,000 polls and results from 20 years of polling
- Built Random Forest, Naïve Bayes machine learning models to handle large datasets
- Wrote, ran python programs to rapidly and independently select key variables from dataset
- Used classification, regression modeling to predict election outcomes from polling data
- OLS model: 0.862 adjusted R-squared score

Senior Thesis
Sept. 2018 to May 2019
- Used ISORROPIA II software to model the Columbia River Gorge atmosphere
- Reviewed previous papers to ground and direct my inquiries
- Identified effect of CO2, NO2, SO4 emissions from the Portland General Electric Boardman power plant
- Managed large datasets acquired from IMPROVE monitoring station tri-daily, 1992-2017
- Compiled and ran WRF-Chem modules and trained new students in its operation and use
- Future thesis work ongoing with prepared WRF-Chem model, instruction

EXPERIENCE

PlanEat
Python Back-End data scientist · Aug. 2020, Aug. 2020 to Current · Seattle, WA
- Built Python engine to calculate greenhouse gas cost, nutritional value of user's mealplan
- Integrated engine into back-end API for PlanEat webpage

Reed College
Research Assistant · Jan. 2018 to May 2018 · Portland, OR
- Researched Heck reaction with palladium catalysts and interaction with urban road dust
- Visualized, presented final findings for clarity to inform future work
- Quantified road dust inhibition of Heck reactions
- Identified an upper and lower bound on the inhibiting capacity of road dust

Voxa
Intern · June 2013 to May 2015 · Seattle, WA
- Helped develop and construct the Mochii, a portable, low-cost electron microscope
- Trained new hires in lab practices and procedures
- Established, refined documentation for future hires
- Managed inventory, updated company database with new component specifications
- Tested Mochii's imaging system on algae samples