ELINOR L. VELASQUEZ, Data Scientist (Entry Level)
elinor.velasquez@Berkeley.edu, 925-532-7670
www.linkedin.com/in/elinor-l-velasquez
https://github.com/elinor2020

SUMMARY

My GitHub repositories feature projects written in Python, using SciKit-Learn, SQL, Bayesian optimization, CatBoost, and other code, involving data wrangling, exploratory data analysis, preprocessing, training, modeling, slides for presentations, and analytical reports. I have taught and tutored in mathematics, statistics (R), and data science (MatLab).

EDUCATION

Certificate in Data Science, Springboard (Expected June, 2021)
M.S., Bioinformatics, University of California, Santa Cruz (2010)
M.S., Cell and Molecular Biology, San Francisco State (2008)
Ph.D., Mathematics (1991)

EMPLOYMENT

Instructor, University of California, Berkeley Extension (3/2021 - present). Teach biostatistics (Fall semester): Lecture, assign and grade homework, midterm, R-based projects, and final exam.

Instructor, University of California, Santa Cruz Extension (4/2013 - 1/2015). Taught statistics for the biosciences: Lecture, assign and grade homework, midterm, R-based projects, and final exam.

Principal Investigator, University of California, Santa Cruz (6/2013 - 8/2013; 6/2014 - 8/2014). Supervised high school students to conduct cancer research, using support vector machines and R (2014), and whole-cell modeling research using MatLab (2013).

Institute Scholar, University of California, Santa Cruz, Center for Biomolecular Science and Engineering (9/2008 - 6/2010). Conducted research in bioinformatics for cancer diagnosis, using R. Created a new algorithm for feature selection for Parkinson's disease, using R. Studied RNASeq using the Bowtie, TopHat, Cufflinks pipeline and other bioinformatics algorithms.

Scholar, National Institutes of Health Research in Science and Engineering, San Francisco State University (7/2004 - 8/2008). Conducted bioinformatics research on Pre-eclampsia using Weka software (Bayesian networks). Conducted computer science research on drug discovery using Java.

RECENT PROJECTS

1. Big Mountain Guided Capstone Ski Ticket Price Optimization (12/2020 - 1/2021). Ski ticket price modeled using Random Forest Regressor. Computed metrics for optimized price. Missing values imputed, data normalized, cross validation employed to estimate best number of features for final optimization.

2. Capstone Two: Rental Bike Share Optimization (3/2021 - 4/2021). The goal was to predict rental bike count per day for Seoul, South Korean company. Data came from the University of California, Irvine Data Repository. The time series data time feature was converted relative to Epoch Time (1/1/1970) to train the model using Epoch Time as a feature. Hyperparameter tuning was employed to optimize the model. The best model (SciKit-Learn metrics) was CatBoost Regressor. Future work will be to model the data using advanced time series algorithms.

3. SQL Mini-Project (12/2020 - 1/2021). SQL commands were used to study a country club’s database. Three tables of data, Bookings, Facilities, Members were investigated using SQL.