

MARCOS DOMINGUEZ

DATA SCIENTIST

✉ md.ghsd@gmail.com
☎ (925) 448-7965
📍 Pleasanton, CA 94566
in marcosdominguez2018/
🌐 mdominguez2010

Skills

SUPERVISED MACHINE LEARNING

Linear/Logistic Regression
K nearest neighbor
Ensemble modeling

UNSUPERVISED MACHINE LEARNING

Topic Modeling (NMF/LDA)
Latent Semantic Analysis
TF-IDF
Stemming & Lemmatization

DATA TOOLS

Pandas
Numpy
Scikit-learn
Matplotlib
Seaborn

DATABASES

SQL
PostgreSQL

OTHER

Tableau
Beautiful Soup
Selenium
HTML
Git

Education

Brandman University Sept. 2016
M.B.A. Accounting

University of California, Davis Sept. 2010
B.S. Economics

Experience

Metis

Data Scientist

San Francisco, CA
Sept. 2020 to Dec. 2020

- Metis is an ACCET accredited 12week immersive data science bootcamp focused on project oriented learning
- The core curriculum is centered around Python, statistics, supervised and unsupervised machine learning, exploratory data analysis, databases, and visualization techniques
- Completed five self-designed data science projects from conception to presentation; including data collection, data management, exploratory data analysis, modeling, and visualizations
- Project highlights in projects section

BAC Community Bank

Senior Credit Analyst

Stockton, CA
Sept. 2016 to Sept. 2020

- Lead credit analyst team in reducing errors and improving turnaround time from 5 days to 2 per request
- Automated a data entry process using Excel sumif formulas and pivot tables, saving hours of manual labor
- Underwrote credit worthiness of borrowers for commercial loans
- Prepared loan presentations for committee approval
- Evaluated loan collateral

Financial Analyst

Stockton, CA
Sept. 2013 to Sept. 2016

- Prepared statements of condition such as income statement, balance sheet, and statement of cash flows with minimal error
- Prepared analytical reports leveraging Excel pivot tables and pulling data from a Cognos database
- Monitored and analyzed financial, statistical and operational trends
- Developed performance reports to track significant variances and trends

Projects

Loan Loss Risk

Dec. 2020

- Predicted a portfolio's loan loss with using both classification and regression supervised machine learning techniques
- Performed data exploration with matplotlib, seaborn, pandas, and numpy
- Executed a variety of models with scikit-learn's algorithms such as knn, logistic regression, XGBoost, Lasso/Ridge Regression, and Elastic Net
- Achieved optimal recall of 0.94 and R-squared of 0.73 with feature engineering

Trump Speech NLP

Nov. 2020

- Developed a recommendations from Donald Trump speech transcripts
- Web-scraped 8 speeches using BeautifulSoup
- Preprocessed text using spaCy and NLTK for lemmatization, removing stop words/punctuation, lemmatization
- Vectorized corpus using scikit-learn's CountVectorizer function
- Extracted insights using topic modeling methods such as NMF, LDA, and SVA

Classification of Credit Card Clients

Nov. 2020

- Predicted whether or not a client would default by utilizing random forest classification algorithm and optimizing recall score
- Implemented hyperparameter tuning with scikit-learn's RandomizedSearchCV methods
- Increased recall score by 15% by fixing class imbalance with imblearn and decreasing decision threshold

Predicting Sneaker Prices

Oct. 2020

- Built linear regression model to predict sneaker prices with RMSE of \$120
- Tested feature engineering's effects on accuracy, with no change observed
- Web-scraped 1,000 rows of data using BeautifulSoup and Selenium