David Madriz

Data Scientist

(415) 724-9614 | dmadriz@berkeley.edu

EDUCATION

University of California, Berkeley | *B.A. in Data Science*

Honors: Regents and Chancellor's Scholarship, Students Rising Above Scholarship, and SEED Honors Program.
Saint Ignatius College Preparatory, San Francisco, CA | Class of 2020

SKILLS

Programming Languages: Python (Advanced), SQL (Advanced), Java(Intermediate), NoSQL (MongoDB, JSON)
Frameworks & Platforms: Scikit-learn, Pandas, Matplotlib, Numpy, SciPy, SQL Server, Power BI, Tableau.
Statistical Data Analysis: Statistical Inference, Bootstrapping, A/B Testing, Algorithms, Data Structures.
Machine Learning: Regression (OLS, LASSO, Ridge), Classification (Logistic Regression, SVM, GDA, Decision Trees, Random Forests), Unsupervised Learning (PCA, Clustering), Neural Networks (CNN, RNN, LSTMs), AI Topics (Q-Learning, Markov Decision Processes)

RELEVANT EXPERIENCE & PROJECTS

Data Analyst & Product Management Intern, Remedly **Summer 2024**

- Conducted market research and roadmap planning to support data-driven decisions for product development.
- Performed exploratory data analysis on users' engagement, identified the top 3 key customer issues, and proposed solutions, improving customer support response times from up to 30 days to less than 1 week with an automated bot system implementation.
- Built a real-time dashboard in Gainsight to monitor engagement metrics and flag churn risk, reducing potential churn from 70% to 35%.

Metropolitan and Non-metropolitan Mortality Predictor **Fall 2024**

- Developed and evaluated predictive models (Generalized Linear Models & Random Forests) to estimate premature mortality, achieving an R^2 of ~.80 with Random Forest.
- Implemented causal inference methods (Weighted Least Squares and Inverse Propensity Weighting) to assess the effect of metropolitan residency on life expectancy, controlling for socioeconomic and health confounders.
- Built a cleaned, merged dataset with over 3,000 counties by using County Health Rankings and USDA Rural-Urban Continuum Data, and normalizing metrics to adjust for population differences.
- Executed feature importance analysis using data visualizations and statistical summaries; identified poverty, smoking, and poor health prevalence as strong mortality indicators for both groups.
- Evaluated model robustness and prevented overfitting using cross-validation, AIC/BIC, and likelihood ratio tests.

Sensor Data Transformation **Fall 2024**

- Created SQL pipelines that detect outliers that are 3 Hampel X84 intervals away from the median values from the same sensor, and performed outlier winsorization.
- Conducted entity resolution to address duplicate data that contains inconsistencies (abbreviation/spelling variations).
- Interpolated missing information by identifying time-related gaps in each sensor, generating tuples to add missing sensor readings, and using linear interpolation to fill in values based on existing data points.

Customer Churn Prediction for Telco **Fall 2024**

- Identified key factors of customer churn (payment methods, internet service) by one-hot encoding variables to visualize feature importance, and understand correlations between features.
- Trained a Logistic regression model, evaluated its performance at 50% accuracy, and improved it by 20% to achieve 70% accuracy after addressing class imbalance.
- Trained a Neural Network model that achieves 75% accuracy as it captures nonlinear trends in the data.

2-D "MyWorld" Game Development **Fall 2024**

- Built a world exploration engine in Java with pseudo-random world generation, interactive objects, and a user avatar.
- Designed a Heads-Up Display for game management, including saving/loading worlds, and real-time state updates.

Spam Email Filtering System **Summer 2023**

- Built interpretable spam classifiers in Scikit-learn, achieving 91% accuracy; I used ROC curves and feature selection to prevent overfitting, and balanced performance with model simplicity for practical deployment and interpretation.