

LUISELY DOZA, Shepherd University, Shepherdstown, WV, 25443, and JASON MILLER, Department of Computer Science, Mathematics, and Engineering, Shepherd University, Shepherdstown, WV 25443. Forecasting California Housing Prices Using a Linear Regression Model

With the prevalence of big data, we are able to implement simple Machine Learning techniques against datasets to solve many present issues. Utilizing the *California Housing Prices* dataset, we can apply linear regression models to forecast districts' future median housing prices. As a result, homebuyers could predict if they are getting a good deal on their home, investors could predict their potential return on investments and capitalize on undervalued properties, and sellers could gauge what they could potentially sell for on the open market. To find a starting base, our research began as an independent study guided by a textbook that provides hands-on machine learning training, including one that uses the *California Housing Prices* dataset. From here, we are able to choose the Least Square Regression Line as our performance measure in the model—which will result in a linear equation that looks for minimizing the variance between the training set's predictions and the actual points produced on the regression line. Once we are able to train our model to make accurate predictions from the testing set, we can expand upon this research by moving beyond linear regression and investigating different Machine Learning techniques further within the textbook to model the data in more sophisticated ways.