

Kean Gabriel E. Salvahan

My First End-to-End Machine Learning Project

My first full machine learning project was about predicting housing prices in California using the California Housing Dataset. It was exciting and helped me understand how all parts of a machine learning project fit together, from loading data to training the model and checking how well it works. Before this, I only knew how to do small parts like training a model or checking accuracy. But this project showed me how each step affects the next one and how small mistakes early on can change the final results.

One important thing I learned was how crucial data cleaning and exploration are. At first, I thought the dataset would already be perfect since it's a popular one. But when I looked closely, I found that some features were unevenly distributed and had extreme values that could confuse simple models like Linear Regression. This made me realize that even clean datasets need checking and preparation before use.

The hardest part was figuring out why my Linear Regression model didn't perform as well as the Decision Tree model. I thought I made a mistake in the code, but after checking everything, I found that the problem was with the model type. Linear Regression assumes straight-line relationships, but housing prices depend on more complex, non-linear patterns. To understand this better, I plotted graphs comparing the actual and predicted prices, which helped me see where Linear Regression went wrong. This taught me that visualizing results is just as important as looking at accuracy scores.

This kind of workflow can be used in real life. For example, a real estate company could use the same steps to build a tool that estimates house prices. They could gather data like house size, location, number of rooms, school ratings, and more, then clean the data, train models, compare their performance, and finally deploy the best one.

In short, this project helped me see machine learning as more than just running algorithms. It's about understanding data, testing ideas, and improving step by step. I now know that being good at ML means being curious, checking your assumptions, and always refining your work. This experience gave me a strong foundation to handle bigger and more challenging ML projects in the future.