Dylan Kasanders

5/14/24

MATH 342 Project Proposal

1. Dylan Kasanders, I plan on doing this project alone.
2. I am using the California Housing dataset. This dataset is using in ‘Hands-On Machine learning with SciKit-Learn and TensorFlow’ Each individual in the dataset represents a housing block in California.
3. We aim to predict median home value for households within a block.
4. Longitude, Latitude, housing, median housing age representing the average age of a house within the block, total rooms representing the total number of rooms within the block, total bedrooms representing the total number of bedrooms within a block, population representing the total number of people residing within a block, households representing the total number of houses within a block, median income representing the median income for households within a block.
5. The data was featured in the paper “Sparse spatial autoregressions.” By Pace, R. Kelly and Ronald Barry. I retrieved the dataset through Kaggle.
6. There are 20641 datapoints.
7. I believe that through the features mentioned and feature extraction we can model a relationship between the aforementioned features and the median home value within the block. Feature extraction can be done specifically on the latitude and longitude values to measure distances the blocks have between certain key areas in California. I would argue that houses closer to “hotspot” areas such as San Deigo, Los Angeles, and other locations with high cost of living would have higher median home values.
8. I want to find the relationship between housing information within a block and the median housing price within the block, and what variables are most attributing to housing prices within California.