Model Card—Housing Price Predictions

Model Details

- Developed by Allie Craddock (alliec45@vt.edu), Esther Kim (estherkdy@vt.edu), and Haley Fore (haleyrhiann4@vt.edu), 2023, v1.
- Linear regression model(s) that utilizes machine learning to predict housing prices. The pricing depends on five different variables: size of the house, number of bedrooms, number of bathrooms, location of the neighborhood the house is in, and the year it was built in.

Intended Use

- Primary intended purposes are to be used as an accessible pricing range for homes as well as for personal use by other data science enthusiasts.
- Not intended to be used as scientific evidence or verification and/or justification about housing prices.

Factors

- Includes environmental and geographical factors such as the model's performance in heavily populated areas.

Metrics

- This model uses accuracy linear regression and a decision threshold of 0.95.
- Evaluation metrics include Mean Squared Error and R-Squared values to measure variability and performance.
- Original data creator used an R-squared value, determining how significant the relationship is between two values that correlate with each other and discloses how much of the data is supporting the correlation of the house features versus the price.

Ethical Considerations

- Potential ethical considerations include data rooted in historical bias.
- Secondary research led to ethical concerns surrounding housing prediction model's potential contribution to gentrification. [1]
- Model does not take into consideration accessibility around the location of neighborhoods.
- With the dataset being synthetic, the model does not consider the social and political factors that determine the price of housing. [2]

Evaluation Data

- Original data *Housing Price Prediction Data* created by Muhammad Bin Imran. [3]
- Data was preprocessed by the original programmer and cleaned once again before evaluation.

Training Data

- Three datasets containing different combinations of house traits were used to create three housing price models: Model One based on square feet, Model 2 based on square feet and neighborhood, and Model 3 based on square feet, neighborhood, and number of bedrooms (see figures below).

Caveats and Recommendations

- Model may not include all changes in the housing market or unforeseen economic changes.
- Model does not take amenities into consideration.
- While the models can be informative and a good example of how to analyze the data, the dataset itself lacks the ability to make proper correlations between multiple variables.

Quantitative Analyses
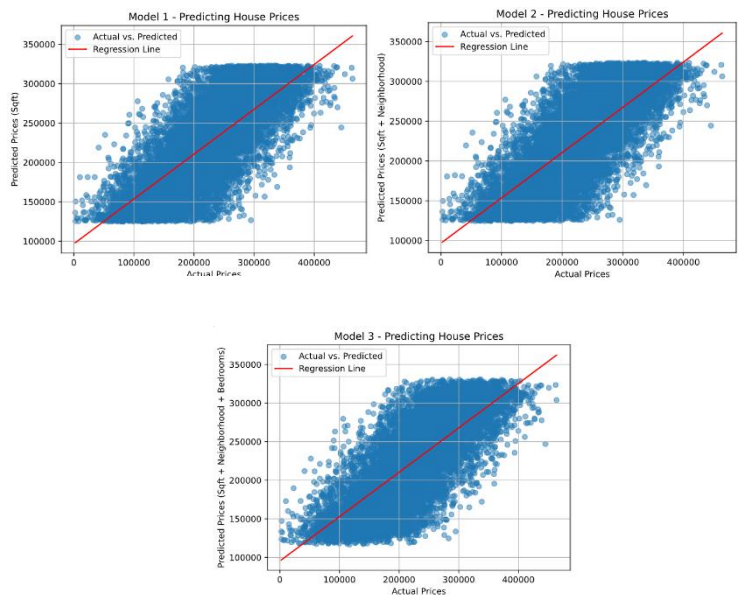
[1] Erb, Mike. (2020). *Machine Learning—An Ethics Review*. Towards Data Science. https://towardsdatascience.com/machine-learning-an-ethics-review-5f4c35112be5

[2] Jones, N. & Walton, R. (2019). *Technical Communication after the Social Justice Turn*. Routledge Publishing.

[3] Imran, M. *Housing Price Prediction Data*. (Version 1) [Data set]. https://www.kaggle.com/datasets/muhammadbinimran/housing-price-prediction-data/