

Technical Report: New York Housing Price Prediction

Data source (Kaggle): <https://www.kaggle.com/datasets/nelgiriwithana/new-york-housing-market?resource=download>

Business Problem:

Investing in a house is a big decision for every person's life especially in an expensive city like New York. So, having the knowledge of the right price for a house is crucial for a buyer, seller or an investor. In this project, data has been compiled to build a model that gives an estimation to the price of a home that an individual is interested, based on its location and feature which helps a person to have a better understanding of the market and avoid overpaying and underselling based on who is buying or selling.

Data Source and Preprocessing:

To build the model, a public dataset from Kaggle named "*New York Housing Market*" was used to predict the housing prices. The dataset included 4,801 property listings approximately which featured many details like location, property size (in sq. ft.) and number of bedrooms and bathrooms. This dataset had no missing values, hence it was easier to process the data. The analysis required scaling the numbers to a common range, and details for the location were converted into a format that could be understood by the model. The data for this model was split into 80/20 as per the requirement.

Model Selection and their Performance:

The models that were tested are:

1. Multiple Linear Regression
2. Random Forest
3. XG Boost
4. Polynomial Regression

Given the data, Random Forest model had the best accuracy with an R^2 of 0.64 which means that it explained about 64% of the variation in prices. XG Boost model performed in between but Linear Regression model was the least performing model with an R^2 of 0.34. Polynomial Regression had no improved results.

Key Insights and Recommendations:

Based on the model's predictions, all the features from location, size of the property and number of bedrooms and bathrooms commonly affected the price of a property. For example: the model predicted the price of a property sized 800 sq. ft. in Manhattan would roughly cost about \$2 million. Hence, these are rough estimates from the model which is useful and should not be relied to for exact values. Also, consulting with an expert is important and considering the market trends.

Limitations and Future Improvements:

The model is limited to the basic features and doesn't consider extensive features like condition of the house or the views from the property and other extra qualities that may or may not affect the price directly. Also, properties with a significant value are not considered by the model leading the prediction to be less accurate. To improve the model, other features can be added like neighborhood factors and house conditions.