

Project Report: House Pricing Prediction

Introduction: The aim of this project was to develop a predictive model for house pricing using two different regression methods: Linear Regression and Gradient Boosting Regression. The chosen project, available on GitHub, served as a foundation for exploring and implementing machine learning techniques.

Project Overview:

- **Project Title:** House Pricing Prediction
- **Source (Before modification):** GitHub Project
[GitHub - Shreyas3108/house-price-prediction](https://github.com/Shreyas3108/house-price-prediction)
- **Models Used:** Linear Regression, Gradient Boosting Regression

Features Used for Model Training:

1. **Bedrooms:** The number of bedrooms in the house.
2. **Bathrooms:** The number of bathrooms in the house.
3. **Sqft_living:** Total living area in square feet.
4. **Sqft_lot:** Total lot area in square feet.
5. **Floors:** The number of floors in the house.
6. **Waterfront:** Whether the property has a waterfront (1 if true, 0 if false).
7. **View:** An index representing the quality of the view from the property (0 to 4).
8. **Condition:** An index representing the condition of the house (1 to 5).
9. **Grade:** An index representing the overall grade given to the housing unit (1 to 13).
10. **Sqft_above:** Square footage of the house apart from the basement.
11. **Sqft_basement:** Square footage of the basement.
12. **Yr_built:** The year the house was built.

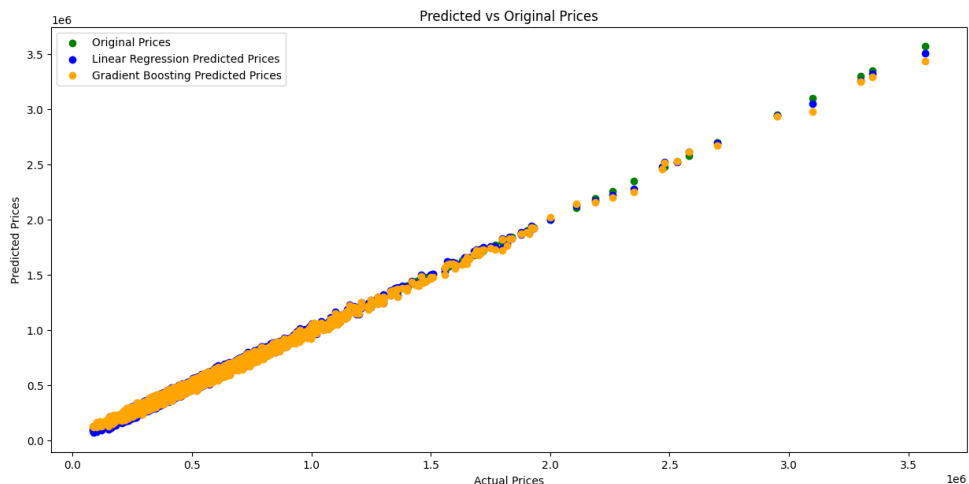
13. **Yr_renovated:** The year the house was last renovated.
14. **Zipcode:** The ZIP code of the property.
15. **Lat:** Latitude coordinate of the property.
16. **Long:** Longitude coordinate of the property.
17. **Sqft_living15:** Average living area of the 15 closest neighbors in square feet.
18. **Sqft_lot15:** Average lot area of the 15 closest neighbors in square feet.

Enhancements Made:

- Eliminated unnecessary code for clarity.
- Introduced two additional features to improve predictions:
 - **Safety_score:** A score representing the safety of the area around the property.
 - **Distance_to_public_transportation:** The distance to the nearest public transportation facility.

Model Evaluation:

- Utilized both Linear Regression and Gradient Boosting Regression models.
- Evaluated model performance using the score method.



Model Usage:

- Predicted prices for new, unseen data to assess model generalization.

Completed Project: The following link is the project with all modifications

[GitHub - hamreen1/House-Price-Prediction](#)

Conclusion: This project provided valuable hands-on experience in implementing regression models for house price prediction. The addition of safety_score and distance_to_public_transportation further improved the model's predictive capabilities. Feedback and comments on the project are eagerly awaited.