SATELLITE IMAGERY BASED PROPERTY VALUATION

1. Introduction

Accurate prediction of house prices is an important problem in real estate analytics, financial planning, and urban development. Traditional prediction models primarily rely on structured tabular data such as house size, number of rooms, and construction year. While these features effectively describe intrinsic property characteristics, they often fail to capture neighborhood-level and environmental factors that significantly influence property value.

This project proposes a multimodal machine learning approach that combines tabular housing data with satellite imagery. Satellite images provide contextual information about the surrounding environment, such as greenery, road connectivity, building density, and neighborhood layout. By integrating these two data modalities, the model aims to improve prediction accuracy and robustness compared to tabular-only approaches.

2. Dataset Description

The dataset consists of residential housing records containing structured numerical attributes along with geographic coordinates. The training dataset includes house prices as the target variable, while the test dataset contains only input features.

Each record includes attributes such as living area, number of bedrooms, number of bathrooms, number of floors, construction year, renovation year, and location-based information. Latitude and longitude values are used to retrieve satellite images corresponding to each house location.

3. Satellite Image Collection

Satellite images are collected using the Mapbox Static Satellite Images API. For each property, a satellite image centered on its geographic coordinates is downloaded at a fixed resolution of 224 × 224 pixels.

The notebook **DATA_FETCHER.ipynb** handles this process. Images are saved locally and indexed consistently to ensure accurate mapping between tabular records and image files. Image downloading is executed only once, and stored images are reused in later stages to avoid repeated API calls and unnecessary computational overhead.

4. Data Preprocessing

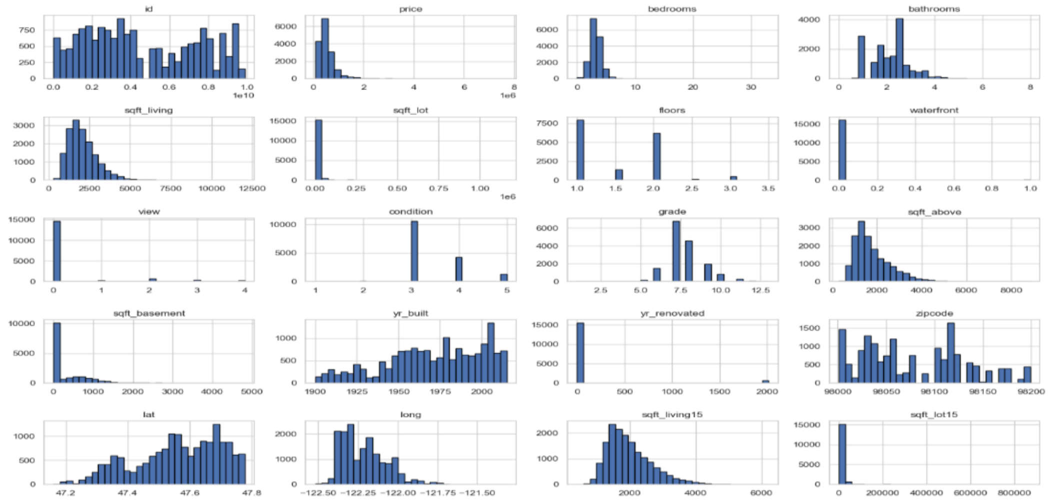
Tabular data preprocessing is performed in **01_preprocessing (2) (1).ipynb**. This step includes handling missing values, removing irrelevant or non-numeric columns, and validating geographic coordinates. Features such as date fields that cannot be directly used by regression models are removed.

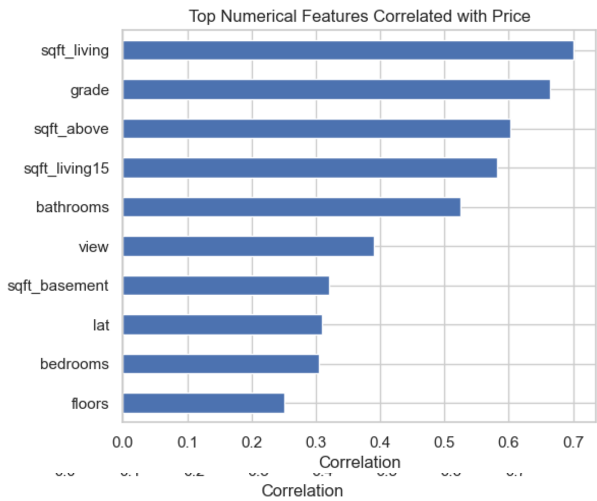
Numerical features are standardized to ensure uniform scaling, which improves model convergence and performance. After preprocessing, the dataset is clean, consistent, and suitable for machine learning model training.

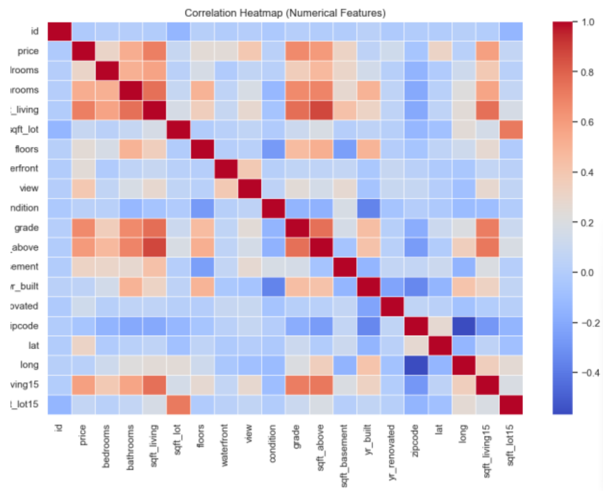
5. Exploratory Data Analysis (EDA)

Exploratory Data Analysis is conducted to understand data distributions and feature relationships. The distribution of house prices is observed to be right-skewed, indicating the presence of high-value properties that exceed the median price.

Numerical feature distribution







Tabular Data–Based House Price Prediction

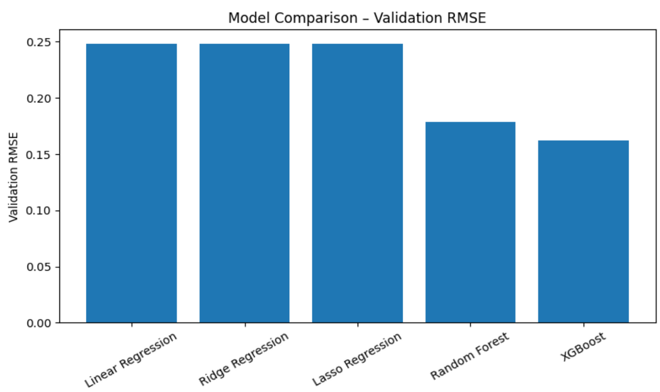
In the initial phase of the study, house price prediction is performed using only structured tabular data. The objective of this stage is to evaluate how well numerical housing attributes alone can explain variations in property prices. The tabular features include house size, number of bedrooms and bathrooms, construction year, renovation year, and other property-specific characteristics.

Exploratory Data Analysis of the tabular dataset reveals that house prices are right-skewed, with a small number of high-value properties significantly exceeding the median price. Scatter plots between price and key attributes such as living area and house grade show strong positive correlations, indicating that larger and higher-quality houses generally command higher prices.

However, considerable variation is observed among houses with similar sizes, suggesting that numerical features alone do not fully capture all price-determining factors.

Several regression models are trained using the tabular dataset, including Linear Regression, Random Forest, and XGBoost. Linear Regression demonstrates limited performance, as it assumes linear relationships between features and price. Tree-based ensemble models significantly improve prediction accuracy by capturing nonlinear interactions among features. Among all tabular-only models, XGBoost achieves the best performance, with a validation RMSE of approximately 0.162 and an R^2 score of approximately 0.905. These results establish XGBoost as the strongest baseline model for tabular data.

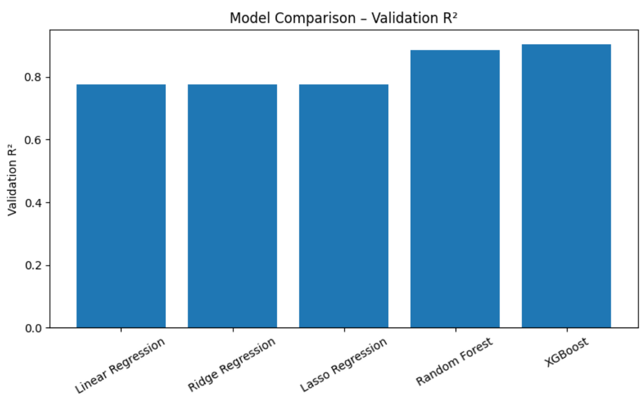
	Model	Train_RMSE	Val_RMSE	Train_R2	Val_R2
4	XGBoost	0.116484	0.162093	0.950488	0.904788
3	Random Forest	0.066149	0.178443	0.984033	0.884612
0	Linear Regression	0.252557	0.248095	0.767246	0.776951
1	Ridge Regression	0.252555	0.248146	0.767250	0.776859
2	Lasso Regression	0.252613	0.248354	0.767143	0.776486



Shows which model predicts prices most accurately

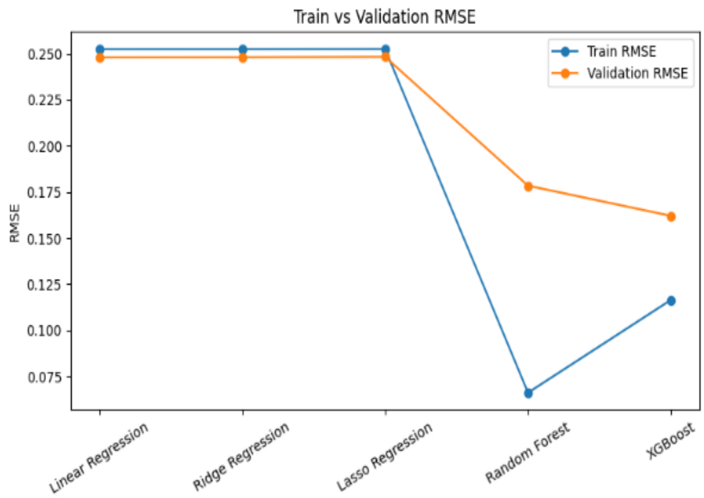
The shortest bar is your best model

Usually → XGBoost



hows explained variance

Higher R² → model captures price behavior better



Multimodal House Price Prediction Using Tabular Data and Satellite Imagery

In the second phase of the study, satellite imagery is incorporated to enhance house price prediction by providing neighborhood-level contextual information. While tabular features describe intrinsic property characteristics, satellite images capture external factors such as surrounding infrastructure, greenery, road connectivity, and spatial layout, which are not explicitly represented in numerical data.

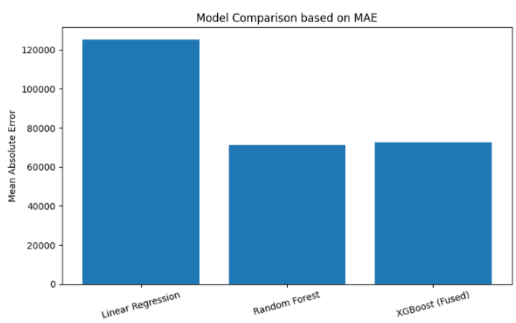
Exploratory analysis of satellite images reveals visible differences between high-priced and low-priced properties. High-value houses are often located in areas with more open spaces, better road networks, and higher levels of greenery, whereas lower-priced houses tend to be situated in denser or less-developed regions. These observations motivate the inclusion of visual features in the prediction model.

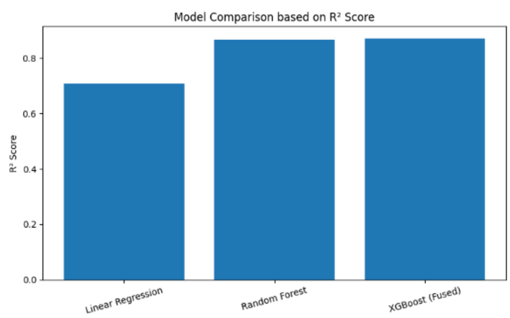
Satellite images are processed using a pretrained convolutional neural network to extract high-level visual features. These image features are then fused with tabular features to form a multimodal representation. An XGBoost regression model is trained on this fused dataset to learn relationships between numerical attributes and visual neighborhood context.

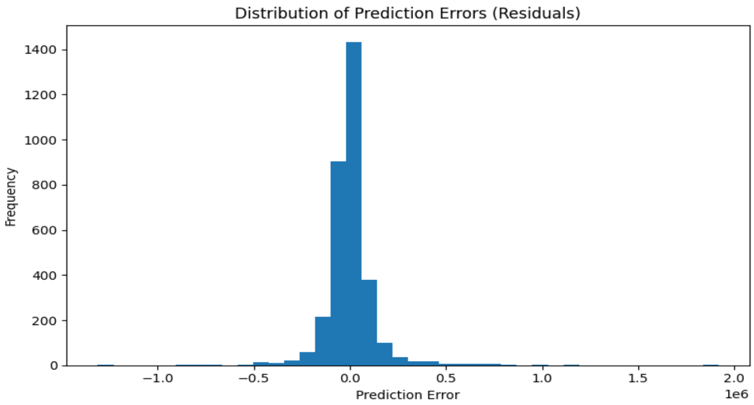
The multimodal model achieves a Mean Absolute Error of approximately 72,751 and an R^2 score of approximately 0.87. Although the tabular-only XGBoost model slightly outperforms the multimodal model in terms of R^2, the inclusion of satellite imagery provides complementary information that enhances interpretability and robustness. The multimodal approach demonstrates that neighborhood context plays a meaningful role in house price prediction, particularly for properties with similar structural attributes.

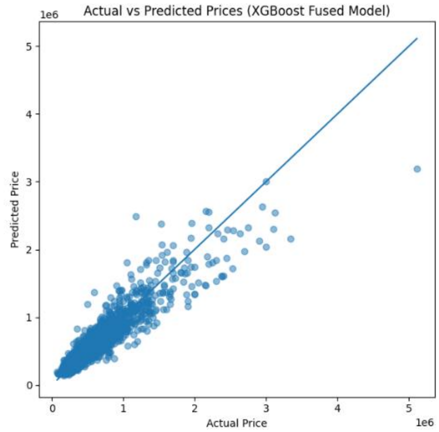
[31]:

	Model	MAE	R2 Score
0	Linear Regression	125224.702297	0.707661
1	Random Forest	71436.940998	0.866234
2	XGBoost (Fused)	72751.359602	0.870395









6. Baseline Models Using Tabular Data

Baseline regression models are trained using only tabular features in **02_Model_Training_Baseline (2).ipynb**. The following models are evaluated:

- Linear Regression
- Ridge Regression
- Lasso Regression
- Random Forest
- XGBoost

Model performance is evaluated using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 score.

Among linear models, Linear Regression, Ridge, and Lasso show similar performance, with limited ability to capture nonlinear relationships. Random Forest significantly improves prediction accuracy but shows mild overfitting.

XGBoost achieves the best tabular-only performance, with a validation RMSE of approximately 0.162 and a validation R^2 score of approximately 0.905. Based on these results, XGBoost is selected as the baseline model for further multimodal integration.

7. Image Feature Extraction

To extract meaningful visual information from satellite images, a pretrained ResNet50 convolutional neural network is used in **04_Image_Feature_Extraction.ipynb**. The network is used as a fixed feature extractor by removing the final classification layer.

Each satellite image is converted into a 2048-dimensional feature vector that captures high-level spatial and visual patterns such as neighborhood layout and surrounding land use. This step is computationally expensive and requires several hours to complete.

To improve efficiency and reproducibility, extracted image features are saved to disk and loaded by default in subsequent runs, preventing unnecessary recomputation.

8. Multimodal Feature Fusion

After image feature extraction, tabular features and image features are concatenated to form a fused multimodal feature set. This representation allows the model to learn relationships between numerical property characteristics and visual neighborhood context.

Feature fusion is performed after scaling tabular features to ensure compatibility with the high-dimensional image feature vectors.

9. Multimodal Model Training and Evaluation

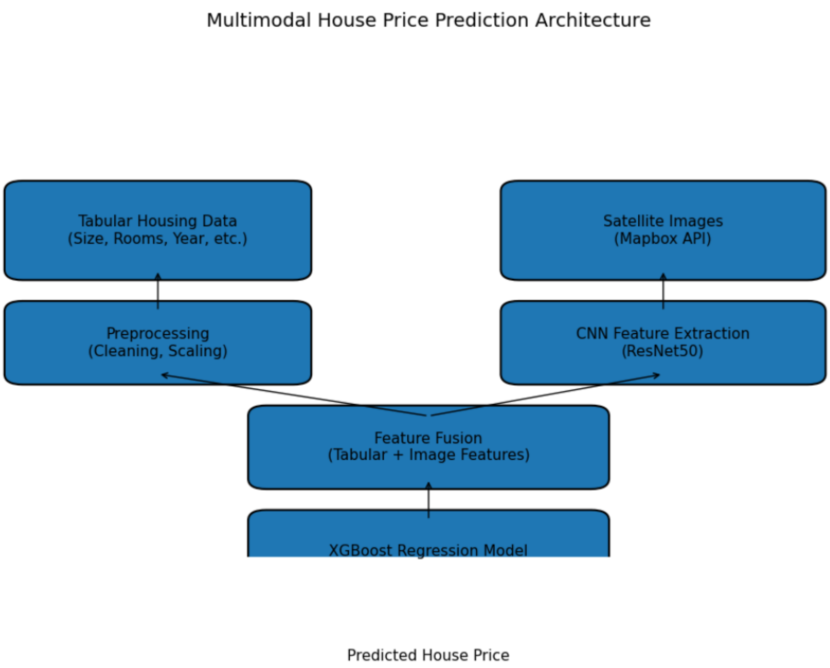
An XGBoost regression model is trained on the fused multimodal dataset. Model performance is evaluated using Mean Absolute Error (MAE) and R^2 score.

The multimodal XGBoost model achieves the following performance:

- Mean Absolute Error (MAE): approximately **72,751**
- R^2 Score: approximately **0.87**

These results indicate that the multimodal model effectively incorporates visual information and maintains strong predictive performance. While tabular data remains a dominant contributor, satellite imagery provides complementary contextual information that enhances model robustness.

Architecture diagram -



10. Final Prediction and Submission

Final predictions on the test dataset are generated in **Prediction file download.ipynb**. The trained multimodal model is applied to the test data, and predicted house prices are exported as a CSV file.

The final submission file **23113075.csv** contains house identifiers and corresponding predicted prices in the required format. The file is generated programmatically to ensure correctness and compatibility with evaluation systems.

11. Conclusion

This project demonstrates the effectiveness of a multimodal machine learning framework that integrates tabular housing data with satellite imagery for house price prediction. The approach leverages the strengths of structured numerical data and unstructured visual data to capture both intrinsic and contextual factors influencing property value.

The results confirm that satellite imagery provides valuable neighborhood-level information that complements traditional features. The proposed multimodal framework can be extended to other real-world regression problems involving geographic and environmental context.

Thanking you

NAME- JAMPA SRI PRAMODHITHA

ENROLL NO. – 23113075

BTECH-3Y, Civil Department