House Price Prediction Using Satellite Imagery

By - Prachi Bhagat 23115106

Abstract

In this project I tried the applicability of Convolutional Neural Networks (CNNs) for predicting the cost of houses with an inclusion of visual and non-visual elements. Traditional end-to-end patterns of supervised machine learning usually incorporate a finite set of numerical or categorical features, ignoring spatial and aesthetic information. CNNs, which can parse complex patterns in unstructured data, represent a way to improve predictive accuracy at the price of including property images with numerical and geographical coordinates. Analysis of the outcomes obtained in experiments showed that CNN-based models performed better than the benchmark models, such as regression, gradient boosting, and random forest regression, in identifying the subtle interconnection between attractiveness and property values. The results further support the ability of CNNs to use multimodal information to solve diverse prediction problems within real estate.

Introduction

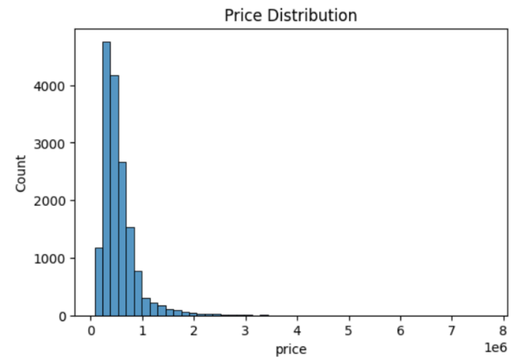
The goal of this project is to predict residential property prices by integrating **tabular data** with **visual embeddings** derived from satellite imagery. Traditional real estate valuation models rely heavily on tabular features such as size, construction quality, and location. This project investigates whether satellite imagery can provide information about neighborhood characteristics that influence property values.

The modeling strategy follows a **multimodal learning paradigm**, consisting of:

1. Strong tabular baselines using gradient-boosted decision trees.
2. Visual feature extraction using pretrained convolutional neural networks.
3. Late fusion of tabular and visual representations.
4. Quantitative evaluation and qualitative explainability using Grad-CAM.

EDA

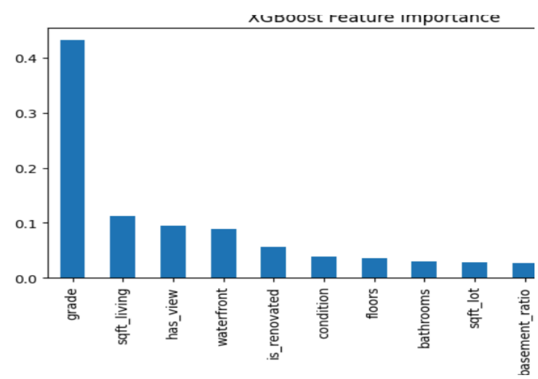
1. Price distribution : The distribution is right skewed , which is usually expected from the price data , since there are a small number of high price houses. Since models work better when the distribution is nearly normal , log transformation is required which results in more symmetrical distribution.



2. Tabular feature correlation : Correlation with price and different features was calculated and after training verified by various models (Random Forest and XGboost). Prices show high correlation with grade , sqft_living and view.

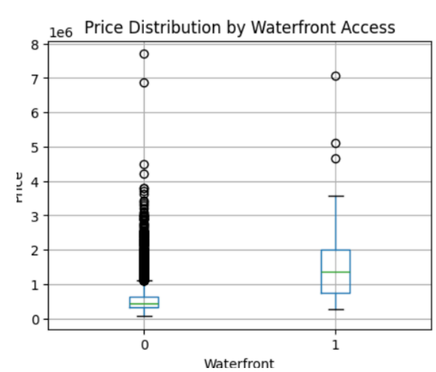
```
corr = df.corr(numeric_only=True)["price"].sort_values(ascending=False)
corr

price          1.000000
sqft_living    0.700933
grade          0.664266
sqft_above     0.602648
sqft_living15  0.581781
bathrooms      0.525487
view           0.390534
sqft_basement  0.320301
lat            0.310008
bedrooms       0.304454
floors         0.251428
waterfront     0.245221
```

```
importance = pd.Series(
    rf.feature_importances_,
    index=X.columns
).sort_values(ascending=False)

importance

grade                 0.418380
sqft_living           0.241921
sqft_lot              0.085468
living_vs_neighbors   0.068612
lot_vs_neighbors      0.047451
basement_ratio        0.033199
condition             0.023213
```



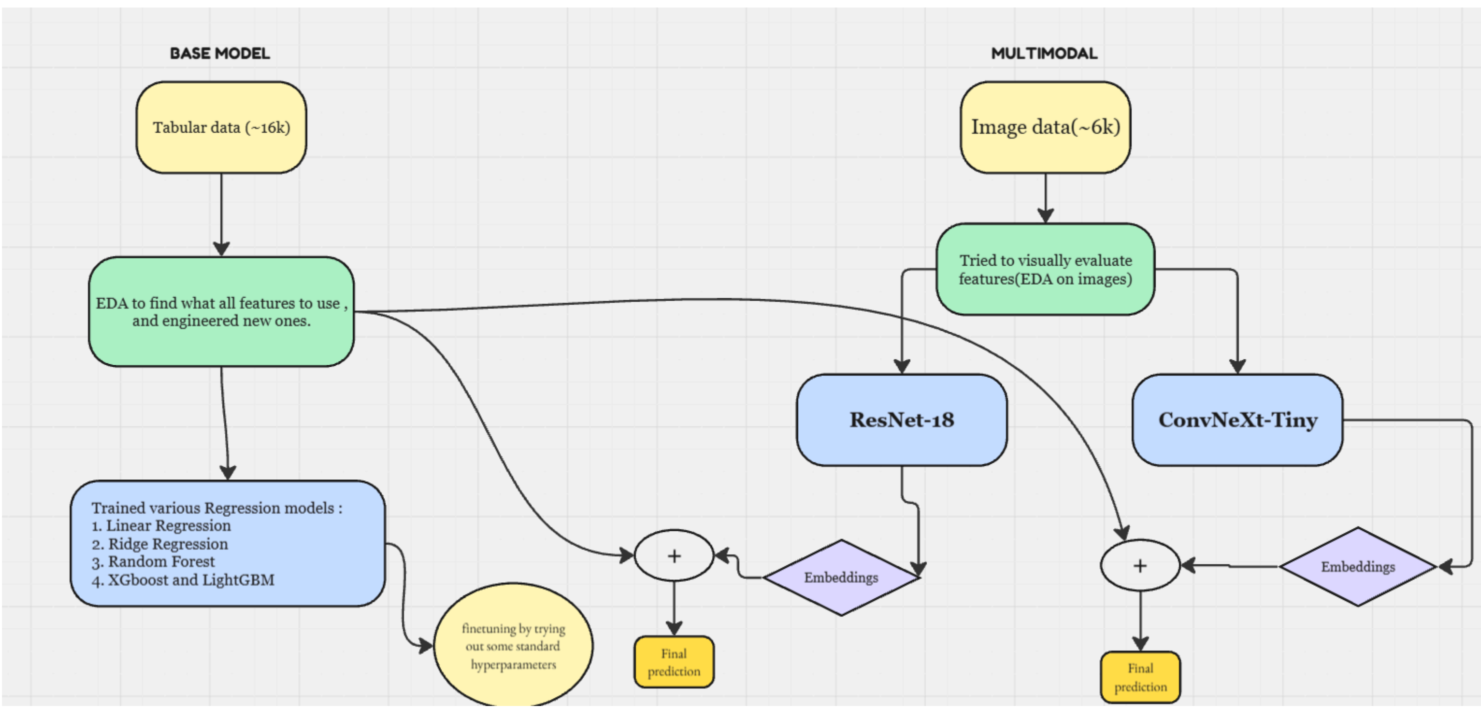
3. Neighborhood context and relative size : Engineered features comparing a property to its neighbors show meaningful associations with price. Houses that are larger than their surrounding neighborhood average tend to be priced higher. This confirms that relative positioning within a neighborhood matters, not just absolute size.

4. Waterfront and view : Waterfront properties show a clear price premium compared to non-waterfront houses. Presence of view further increases the expected price.



5. Visual exploration of satellite imagery : Satellite images show substantial variation in green cover and open land , road compactness , housing density , proximity to water bodies . Simple handcrafted visual proxies (like green color more than blue and red) show weak correlation therefore promoting use of CNN.
6. Grad-CAM : Grad-CAM confirmed that the CNN focuses on semantically meaningful regions such as waterfront proximity and green space, aligning with known real-estate valuation drivers and improving trust in the multimodal pricing model.

Architecture Diagram



Baseline Tabular Models

As a strong baseline, several tabular-only (~16k data) regression models were trained using engineered housing features:

- Linear and ridge regression models to establish a benchmark.

- Gradient boosting models (XGBoost and LightGBM) to capture non-linear interactions.
- Tree-based models achieved substantially higher performance than linear models, confirming the presence of complex relationships between housing attributes and price.

Visual Feature Extraction Using CNNs

Rather than training convolutional networks from scratch, pretrained convolutional neural networks were used as feature extractors. Specifically:

- **ResNet-18** and **ConvNeXt-Tiny**, pretrained on ImageNet, were employed to extract high-dimensional visual embeddings from satellite images.
- The classification heads were removed, and the output of the final convolutional layers was used as a compact representation of the image.

To mitigate overfitting PCA was applied to reduce embedding dimensionality prior to modeling.

Multimodal Fusion Strategy

A late fusion approach was adopted to combine visual and tabular information (~6k). In this framework:

- Tabular features and CNN-derived image embeddings were processed independently
- The two feature sets were concatenated at the feature level and lightGBM was used as a model to predict final prices.

Due to time constraints ~6000 images were used which surpassed the results of 16K tabular data (RMSE and R2 score) , and thereby it is concluded that taking more image data will further improve the prediction score .

Results

Tabular Data only (16K samples) :

RMSE (log prices) = ~0.29

R2 = ~0.68

Tabular + Satellite data(6k samples) :

RMSE (log price) = ~0.26

R2 = ~0.73

Satellite imagery provides some environmental context that refines prediction . Here , small gains are also meaningful since tabular data was already strong. Importantly multimodal achieves higher accuracy even with less data which is true success for the project.

Financial and visual insights :

Analysis of tabular data confirms that house price formation is primarily driven by property size, construction quality, and neighborhood context, which align with established real estate valuation principles. Gradient-boosted models trained solely on these attributes achieve high predictive performance, indicating that most price variation is already captured by structured data.(already proved in images above)

Satellite imagery contributes complementary information related to environmental quality and neighborhood layout rather than property interiors. Through Grad-CAM it was observed that the presence of open green spaces , low housing density , proximity to water bodies and more organized road structure contributed to high property value.