Satellite Imagery-Based Property Valuation

1. Overview

The objective of this project is to predict residential house prices by integrating **structured tabular data** with **satellite imagery**. Traditional house price models rely heavily on tabular attributes such as property size, location, and amenities. However, such data often fails to capture **neighborhood-level characteristics**—including road connectivity, housing density, and surrounding land use—which can significantly influence property value.

To address this limitation, this project adopts a **multimodal learning approach**, combining:

- A **Convolutional Neural Network (CNN)** trained on satellite images to extract spatial and visual features.
- A **tree-based regression model (XGBoost)** trained on tabular data and CNN-derived embeddings.

The CNN is used **not as a standalone predictor**, but as a **feature extractor**, enabling the fusion of visual context with structured data. The final prediction is produced by XGBoost, which learns optimal interactions between tabular and visual features.

2. Exploratory Data Analysis (EDA)

2.1 Price Distribution

The house price distribution is right-skewed, with a majority of properties clustered in the lower-to-mid price range and a small number of high-value outliers. This behavior is typical in real estate datasets and motivates the use of robust regression models such as XGBoost.

Visualizations used:

- Histogram of house prices
- Boxplot to identify outliers

These plots confirm the presence of extreme values and justify careful evaluation using RMSE and R^2 rather than relying solely on mean-based metrics.

2.2 Satellite Image Inspection

Each property is associated with a fixed-size satellite image tile centered at its geographic coordinates. Visual inspection of random samples shows that the images consistently capture:

- Road networks
- Housing density and layout

- Vegetation and open spaces
- Urban vs. suburban structure

The images are not focused on individual houses but rather on **neighborhood context**, which aligns with the intended modeling objective.

3. Financial and Visual Insights

3.1 Quantitative Insights (SHAP Analysis)

To quantify the contribution of satellite imagery, SHAP (SHapley Additive exPlanations) was applied to the final XGBoost model.

The contribution breakdown was:

- **Tabular features:** ~79%
- **CNN-derived visual features:** ~21%

This demonstrates that satellite imagery provides **meaningful complementary information**, even though tabular features remain the dominant predictors of price.

3.2 Qualitative Insights (Grad-CAM Analysis)

Grad-CAM was used to interpret the CNN's learned visual representations.

The activation maps indicate that the CNN focuses on:

- Road intersections and connectivity
- Housing density patterns
- Neighborhood layout regularity
- Boundaries between built-up areas and open spaces

Notably, the model does **not** focus on isolated rooftops or arbitrary pixels. Instead, it captures **spatial patterns associated with neighborhood desirability**, such as infrastructure accessibility and organized urban planning.

This confirms that the CNN is learning **economically meaningful visual cues**, rather than noise.

4. Architecture Diagram

Model Pipeline :

The pipeline and the flow of the project is depicted in the image below:

```
Satellite Image
      |
      ▼
Convolutional Neural Network (CNN)
      |
      ▼
Image Embeddings (cnn_0 ... cnn_127)
      |
      ├───────────────┐
      |               |
Tabular Features      |
(size, location,      |
amenities, etc.)      |
      |               |
      ▼               ▼
     XGBoost Regressor
             |
             ▼
       Predicted House Price
```

Design Rationale

- The CNN learns spatial representations from images.
- Image embeddings are treated as additional numerical features.
- XGBoost performs final regression, allowing flexible nonlinear interactions between tabular and visual features.

This modular design improves interpretability, robustness, and scalability.

5. Results

5.1 Model Performance Comparison

Model Type	R^2 Score	RMSE (Approx.)
Tabular Data Only (XGBoost)	~0.89	~118K
Tabular + Satellite Images	~0.888	~118K

5.2 Interpretation

- Tabular data alone provides strong predictive power.
- Adding satellite imagery does **not dramatically increase R^2**, which is expected for real-estate pricing problems where structural attributes dominate.
- SHAP analysis confirms that visual features still contribute significantly (~21%) to the prediction process.

The results indicate that satellite imagery adds **contextual refinement rather than dominant predictive power**, aligning with real-world housing economics.

6. Conclusion

This project demonstrates a **successful multimodal learning pipeline** for house price prediction. By integrating satellite imagery with tabular data, the model captures both **property-level attributes** and **neighborhood-level context**.

Key takeaways:

- CNNs are effective **feature extractors** for spatial data.
- Tree-based models naturally balance tabular and visual information.
- SHAP and Grad-CAM together provide comprehensive interpretability.
- The approach is realistic, scalable, and aligned with industry practices.