Shahid Beheshti University
Data Science (2023-2024)

House Price Estimation from Visual and Textual Features

Mehrab Kalantari

Department of Computer Science

Shahid Beheshti University

mehrab.kalantari.22@gmail.com

Abstract

Most existing automatic house price estimation systems rely only on textual data, such as the area and the number of rooms. The final price is often estimated by a human agent who visits the house and assesses it visually. In this project, visual features from house photographs are extracted and combined with the house's textual information. These combined features are used in machine learning and deep learning algorithms to estimate the house price as a single output. To train and evaluate the models, a dataset that includes both images and textual attributes is used, consisting of 535 sample houses from California, USA.

1 Introduction

The housing market plays a significant role in shaping the economy. Housing renovation and construction boost the economy by increasing house sales, employment, and expenditures. It also impacts demand in related industries, such as construction supplies and household durables. The traditional price prediction process relies on sales price comparison and cost, which is often unreliable and lacks a standard or certification process. Therefore, precise automatic price prediction is needed to help policymakers design effective policies and control inflation, as well as to assist individuals in making wise investment plans. This process requires combining both visual and textual attributes for accurate price estimation.

Project Contributions:

- a dataset provided by Eman Ahmed that combines both visual and textual attributes for price estimation is used.
- Both machine learning and deep learning approaches are proposed to estimate prices and report the results of these models.

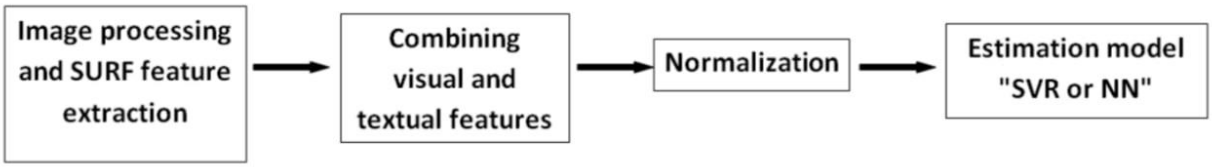
The remaining of this project is organized as follows:

I begin with a review of related work, then provide a detailed overview of the dataset by Eman Ahmed. Next, proposed methods and baselines for solving this problem are presented, followed by the results for each baseline.

The complete code for this project can be found on my GitHub repository: [GitHub Link]

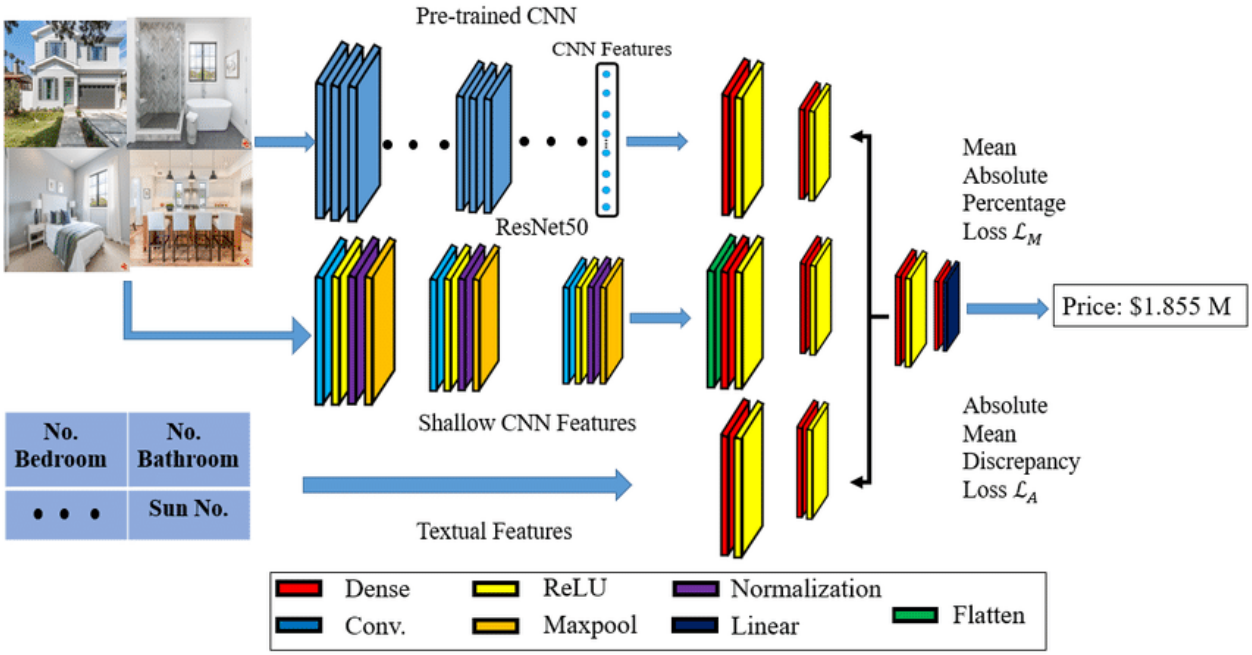
2 Related Works

For the first time, estimating house prices using both visual and textual data was proposed by Eman Ahmed and Mohamed Moustafa in their 2016 paper, House Price Estimation from Visual and Textual Features. Their work focused primarily on feature engineering and preprocessing rather than on training complex models. The proposed pipeline includes the following steps:



The SURF method was used to extract features from each image, with an example shown on the right. Min-Max normalization was then applied to the combined data. Finally, a multilayer perceptron and a support vector regressor were used to estimate the price.

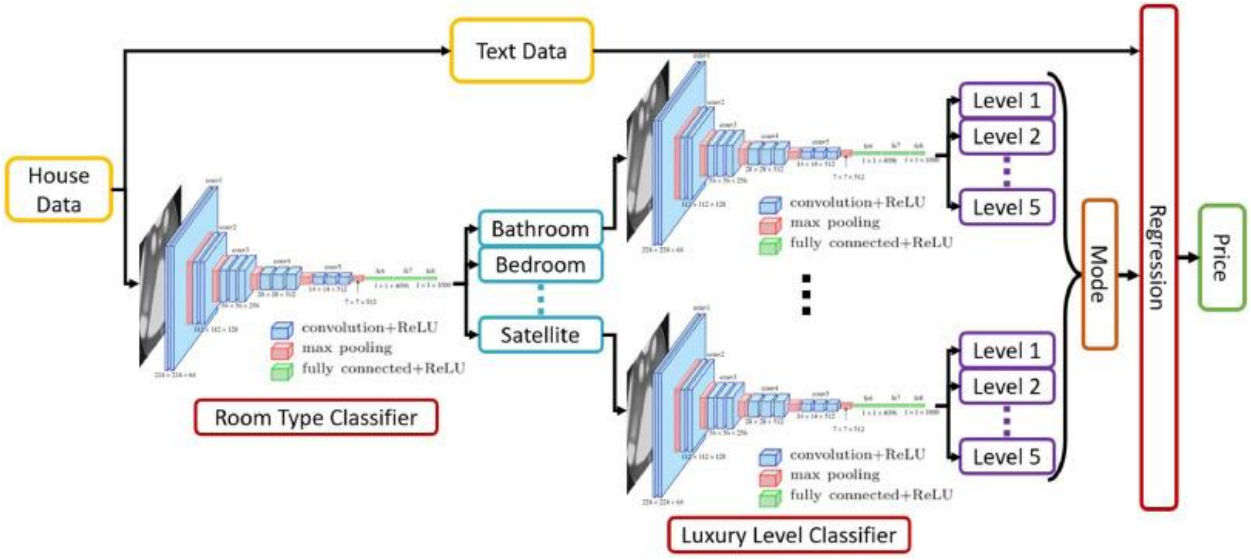
In 2020, Youshan Zhang published a new paper proposing a deep learning architecture for house price estimation using a three-channel convolutional neural network.



The model consists of three modules: a deep feature module, a shallow CNN feature module, and a textual feature module. In the deep feature module, features are extracted using a pre-trained ResNet50 network. The shallow CNN feature module obtains features directly from raw images through three repeated blocks, while the textual feature module primarily includes two layers.

The model uses two loss functions: mean absolute percentage error (MAPE) loss and absolute mean discrepancy loss. MAPE measures the average percentage difference between the predicted and actual prices, while absolute mean discrepancy loss ensures that the mean predicted price closely approximates the mean actual price.

Ultimately, in 2022, a paper titled Vision-Based Housing Price Estimation Using Interior, Exterior, and Satellite Images was published to tackle a more advanced price estimation problem. This new deep learning architecture can be adapted to address our specific issue:



The model includes 6 stages:

1. Classify images to room categories (bathroom, bedroom, kitchen, living room, dining room, front, satellite images).
2. Classify each category of images by luxury levels.
3. Take the mode of the luxury levels across all room types as a level feature.
4. Combine this feature with text data.
5. Input the combined data into a regression model.
6. Compare the performance with existing methods in the literature.

3 Dataset Description

3.1 Understanding Dataset

The collected dataset consists of 535 sample houses from California, United States. Each house is represented by both visual and textual data. The visual data includes four images: a frontal view of the house, the bedroom, the kitchen, and the bathroom, as shown in the figure below. The textual data represent physical attributes of the house, such as the number of bedrooms, number of bathrooms, area, and zip code of the location. This dataset was manually collected and annotated from publicly available information on real estate websites. There is no duplicate or missing data entries.

Details about the dataset and its features are provided in the following tables. The dataset includes five numerical features and four images for each house.

Bedrooms	Bathrooms	Area	Zipcode	Price
4	4.0	4053	85255	869500
4	3.0	3343	36372	865200
3	4.0	3923	85266	889000
5	5.0	4022	85262	910000
3	4.0	4116	85266	971226

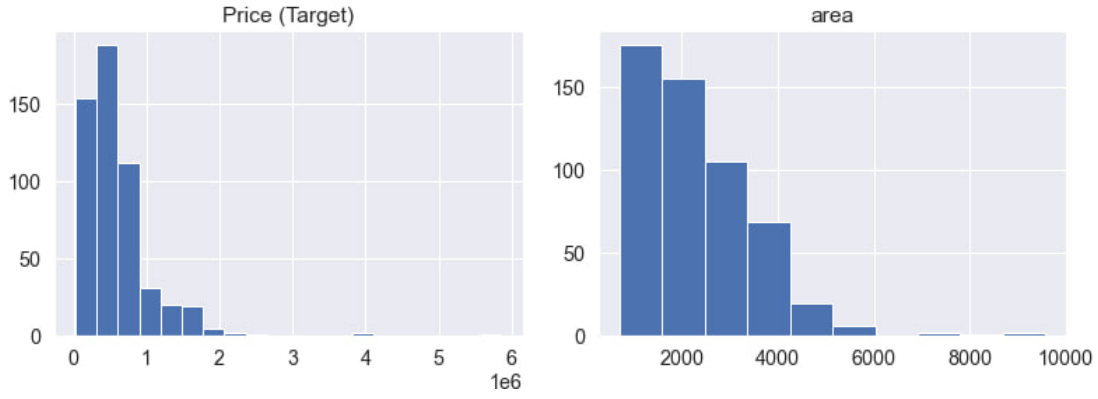
Feature	Dtype	Unique Numbers	Description
bedrooms	int64	9	Number of bedrooms
bathrooms	float64	14	Number of bathrooms
area	int64	435	Area of the house
zipcode	int64	49	Zipcode
price	int64	369	House price

3.2 Statistical Details

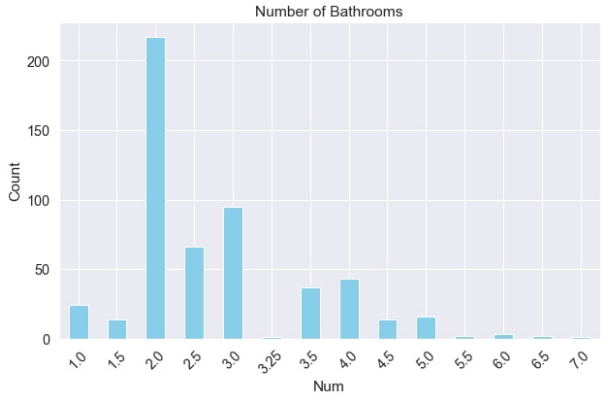
The chart below provides statistics related to the dataset. The house prices range from $22,000 to $5,858,000. The minimum image resolution is 250x187 pixels, which is important and will be discussed later. Additionally, the number of bathrooms ranges from 1 to 7, while the number of bedrooms ranges from 1 to 10.

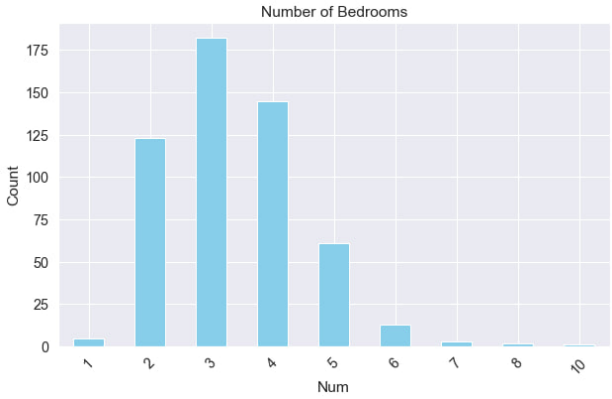
Detail	Average	Minimum	Maximum
House price (USD)	589,360	22,000	5,858,000
House area (sq. ft.)	2364.9	701	9583
Number of bedrooms	3.38	1	10
Number of bathrooms	2.67	1	7
Images resolution	801x560	250x187	1484x1484

The provided histograms illustrate the distributions of the price and area features, which exhibit normal distributions with skewness. The range and average for each feature can be compared in the table. Additionally, there are some high values that may be related to each other.



The bar charts provide information about the distribution of the number of bedrooms and bathrooms. It is evident that most houses have 2 bathrooms and 3 bedrooms. Notably, there is only one house with 10 bedrooms and 7 bathrooms.



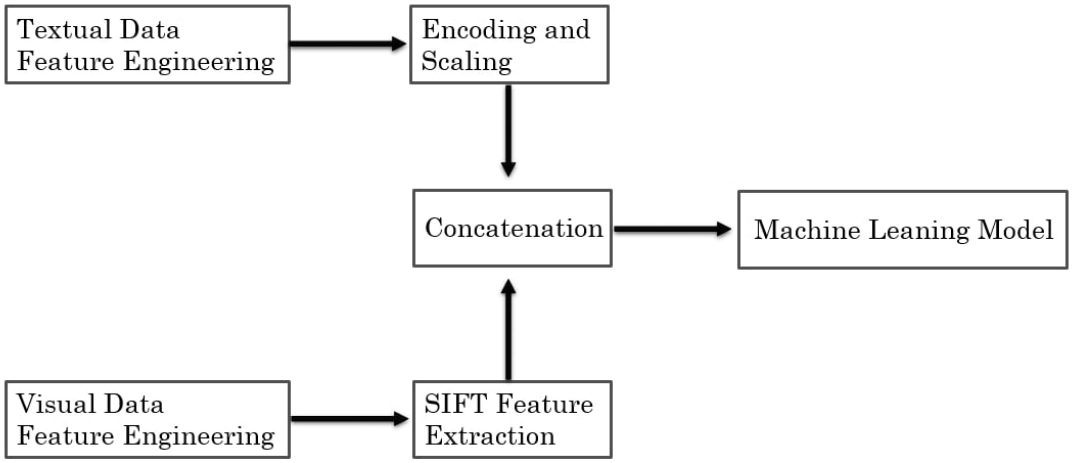


4 Proposed Methods

In this section, methods used to estimate house prices are discussed. Generally, proposed approaches can be categorized into machine learning-based and deep learning-based methods.

4.1 Machine Learning Approach

The overall pipeline used in this section is illustrated in the following flowchart.



Two feature engineering phases for both textual and visual features were employed. Then, I applied a feature extraction algorithm to the visual data and a preprocessing module to the textual data. Finally, the results were concatenated and input into machine learning algorithms.

Textual data feature engineering includes three stages:

Stage 1: Creating and extracting new features from the dataset

- Total rooms: The sum of the number of bedrooms and bathrooms, plus 1 for the kitchen.
- Average room area: The ratio of the total area to the total number of rooms.
- Area size: A categorical classification of the area, categorized as very small, small, medium, large, or very large.

Stage 2: Ordinal encoding for categorical columns
This stage focuses on area size, as it is a feature that is inherently comparable.

Stage 3: Log transformation for skewed features
Log transformation was applied to skewed features, including area, average room area, and price. This choice was based on tests showing that log transformation yielded better results compared to normalization and standardization.

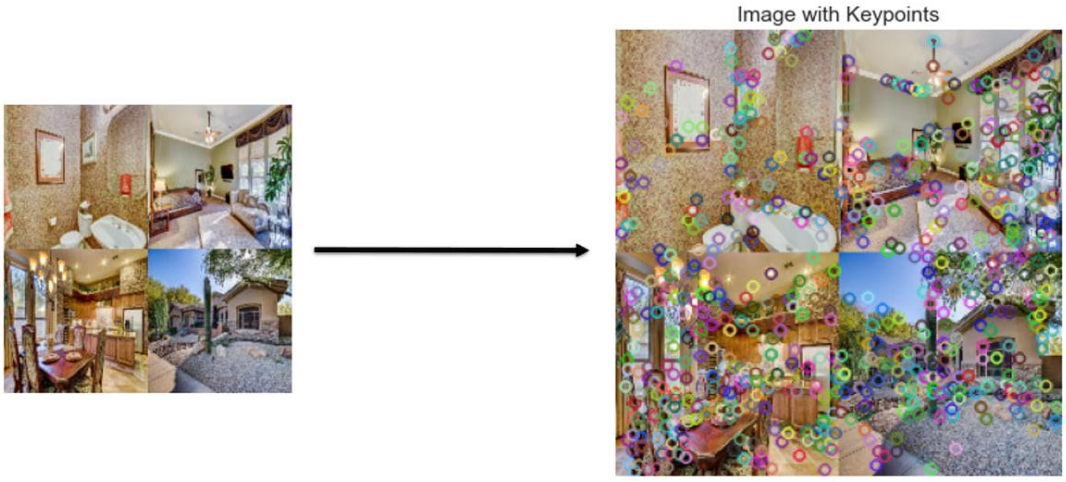
Visual data feature engineering includes two stages:

Stage 1: Representation of the images
In this stage, a matrix was created to represent all four images of a house, allowing us to refer to each house by a single n x n matrix. Selecting the size of the matrix was challenging; several tests were conducted to determine the optimal dimensions while ensuring the minimum image resolution of 250 x 187 was not exceeded. Initially, a 256 x 256 matrix was chosen, with each image sized at 128 x 128, which was appropriate. After further testing, this initial choice yielded the best results.

At the end of this stage, there were 535 matrices, each sized 256 x 256 x 3, containing four images with RGB channels.

Stage 2: Feature Extraction Using SIFT
In this stage, the SIFT method was applied to extract features from each group of images. This step was crucial for reducing dimensionality, allowing us to utilize machine learning models and achieve better performance. I opted for SIFT primarily because previous works used SURF, and I wanted to explore new methods that might outperform existing algorithms. By the end of this stage, there were 535 images, each with a dimensionality of 128, representing the extracted features from the houses.

An example of this method is as follows:



4.1.1 Classic Machine Learning Models

In this section, several classic machine learning models and their training processes are discussed.

1. **Linear Regression**
   A simple built-in linear regression model was trained on the dataset, but it achieved a poor score.

2. **Polynomial Regression**
   For this model, polynomial regression with degrees 2 and 3 was used. Unfortunately, the results were very disappointing, ranking as the worst among all methods.

3. **Ridge Regression**
   This model, with a regularization parameter set to 1.0, performed better than the previous two methods, but it still did not achieve the best score among the classic methods.

4. **Decision Tree Regressor**
   The decision tree model, with a maximum depth of 3, yielded the best results among the classic methods. Its performance was significantly better than that of the other models.

4.1.2 Advanced Machine Learning Models

This section focuses on more advanced machine learning models that utilize bagging and boosting methods. Extensive tests were conducted to find the best hyper-parameters for these models, and the highest results achieved were recorded.

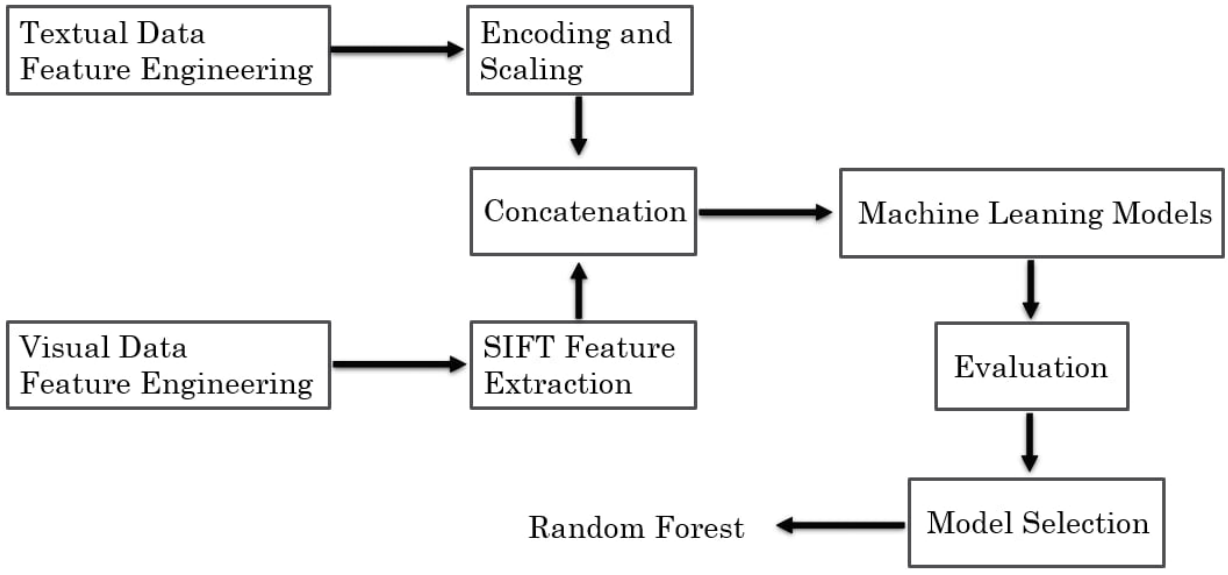
1. **Random Forest Regressor**
   This method achieved the best result among all the models tested for this task, using only 50 estimators.

2. **Support Vector Regressor**
   The SVR achieved a relatively high score compared to the classic methods, but it did not perform as well as the advanced methods.

3. **CatBoost Regressor**
   After thorough testing and evaluation to fine-tune this algorithm, it performed well, but the results were not better than the Random Forest.

4. **XGBoost Regressor**
   Similar to CatBoost, after extensive fine-tuning, the XGBoost regressor also achieved a high score. However, like CatBoost, it did not surpass the performance of the Random Forest Regressor.

At this point, the proposed diagram can be completed to illustrate the overall workflow of the house price estimation process using machine learning models.



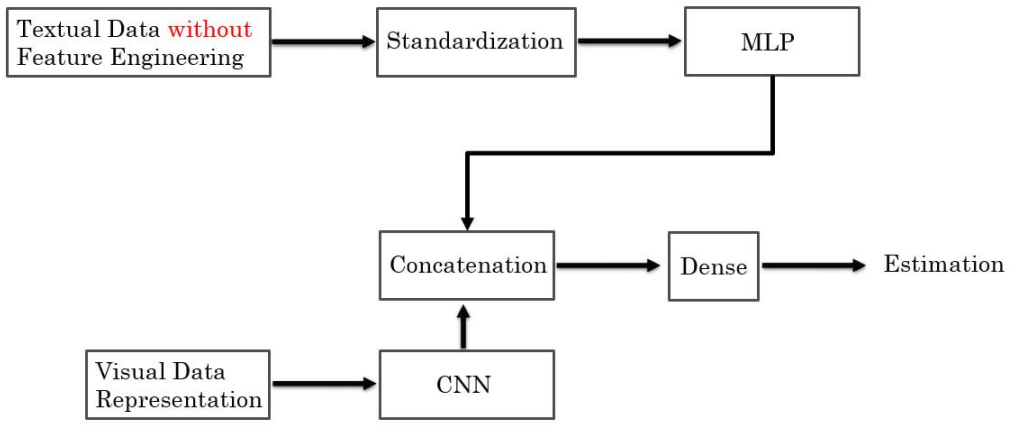
For implementation details, please refer to the code available on my GitHub repository: [GitHub Link]

4.2 Deep Learning Approach

In this section, neural networks with different architectures are utilized to estimate house prices. Two types of neural networks are employed: Multilayer Perceptron (MLP) and Convolutional Neural Networks (CNNs). The rationale behind choosing MLP for textual data and CNNs for visual data is to create a multi-channel neural network that can process both types of information effectively.

For this approach, the same preprocessing and feature engineering techniques as before are not applied. Instead, I allow the network to extract features directly from the raw data and attempt to solve the problem independently.

The overall workflow for this approach is illustrated in the following diagram:



For the neural network approach, standard scaling was applied to textual data to ensure optimal performance. Various preprocessing methods were explored, including those used in the machine learning phase. Ultimately, standardization yielded the best scores for the network.

Another challenge was determining how to represent the visual data. I chose to use matrix representation from the previous section but needed to identify the most effective size. After conducting tests, it was evident that a matrix size of 64 x 64, with each image being resized to 32 x 32, produced the best results.

Following these preprocessing steps, textual data is processed through the MLP channel, while visual data is handled by the CNN channel. The outputs from both channels are concatenated and passed through a hidden dense layer, with the final output layer estimating the house price.

A significant consideration in this section was the architecture of the network. I aimed to avoid creating a deep and complex network due to the limited amount of data. Therefore, I opted for a shallower architecture consisting of only three convolutional layers.

4.3 Performance Evaluation

4.3.1 Mean Absolute Error

The Mean Absolute Error (MAE) measures the average absolute difference between predicted values and actual values, serving as a key metric for assessing the effectiveness of a regression model.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} \left| y_i - \hat{y}_i \right|$$

4.3.2 R2 Score

The R^2 Score, or coefficient of determination, quantifies the closeness of the predicted model to the actual model. It is calculated using the following formulas, which involve various error metrics:

- SSE (Sum of Squares of Errors): This measures the total deviation of the predicted values from the actual values.
- SST (Sum of Squares Total): This captures the total deviation of the actual values from their mean.

$$\text{SSE} = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$$

$$\text{SST} = \sum_{i=1}^{n} (\bar{y} - y_i)^2$$

The value of R^2 ranges from 0 to 1, with higher values indicating a more accurate estimation model. The R^2 score is calculated as follows:

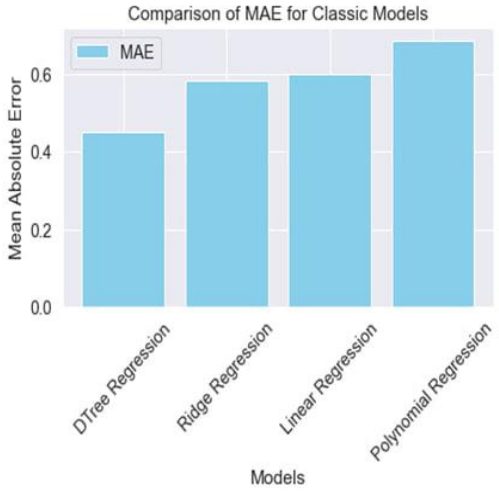
$$R^2 = 1 - \frac{SSE}{SST}$$

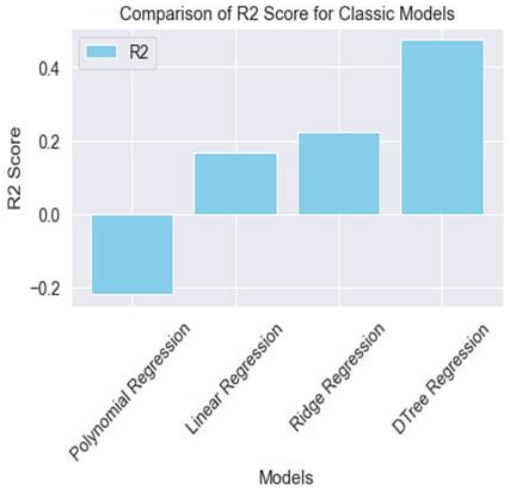
5 Results

In this section, the results achieved by the employed methods are presented and discussed.

5.1 Classic Methods Results

The results for classic machine learning algorithms are as follows:

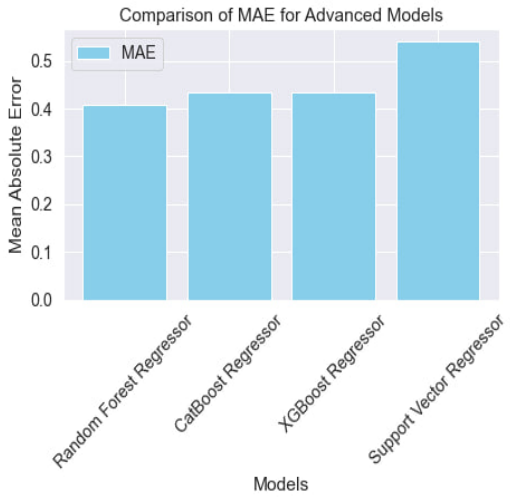


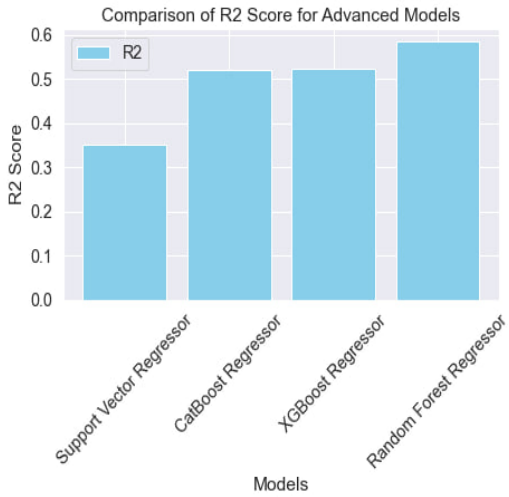


It is evident that the best score was achieved by the decision tree regressor, while polynomial regression recorded the poorest performance among all the classic methods.

5.2 Advanced Methods Results

The results for advanced machine learning algorithms are as follows:



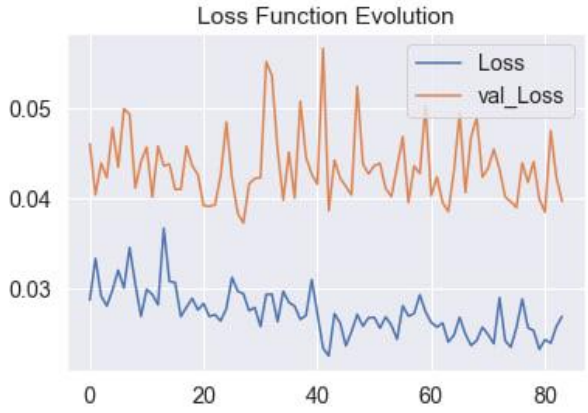


We can see that the best performance overall was achieved by the Random Forest Regressor. Additionally, the scores for CatBoost and XGBoost methods are nearly equal and outperform the Decision Tree Regressor. On the other hand, the Support Vector Regressor (SVR) performed weaker than the Decision Tree.

5.3 Neural Network Results

For the neural network, the best score achieved after 100 epochs was 40%, which is not as competitive as the advanced models. This performance can be attributed primarily to the limited amount of data and the simple desinged architecture. Additionally, unlike the machine learning models, which utilized feature engineering and preprocessing, the neural network operated directly on the raw data.

The following chart illustrates the training loss and test loss for the network:

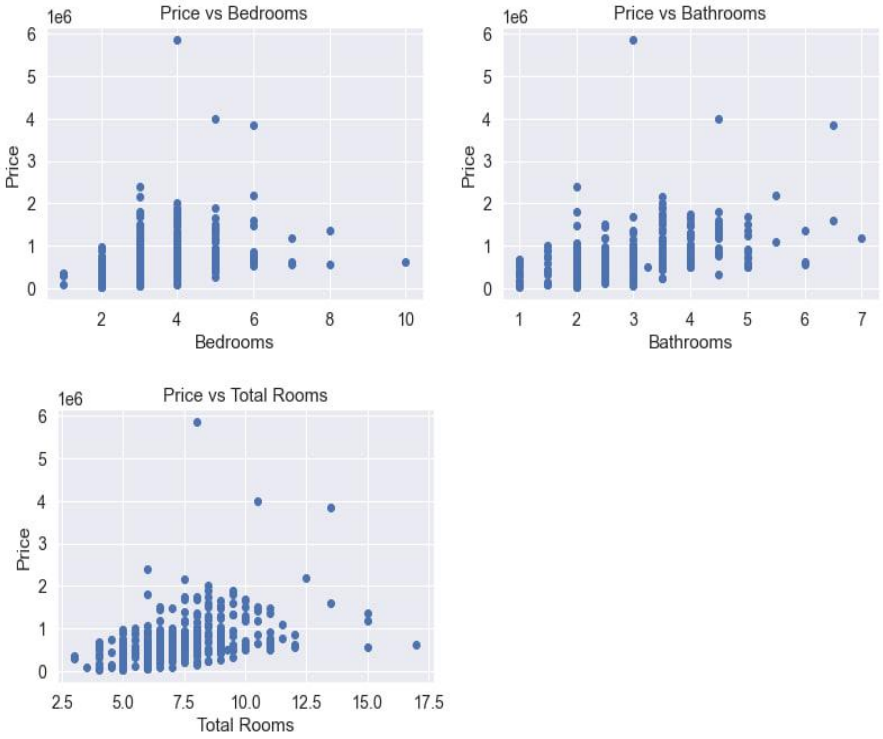


6 Further Analysis

6.1 Feature Importance

In this section, feature importance and the impact of each feature on the target variable, house price, are addressed. Understanding which features significantly influence predictions is crucial for refining trained models.

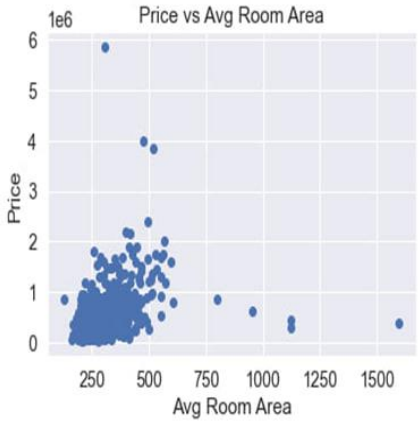
The following charts illustrate the relationship between the target (house price) and the number of rooms, specifically focusing on bedrooms, bathrooms, and total rooms:



It is evident that each feature exhibits a strong correlation with the target variable, particularly total rooms, the derived feature from the sum of bedrooms and bathrooms. This correlation suggests that larger homes, indicated by a higher total room count, tend to command higher prices.

I will delve into a statistical analysis of these correlations in the following sections, providing further insights into their significance and how they can be leveraged to improve model performance.

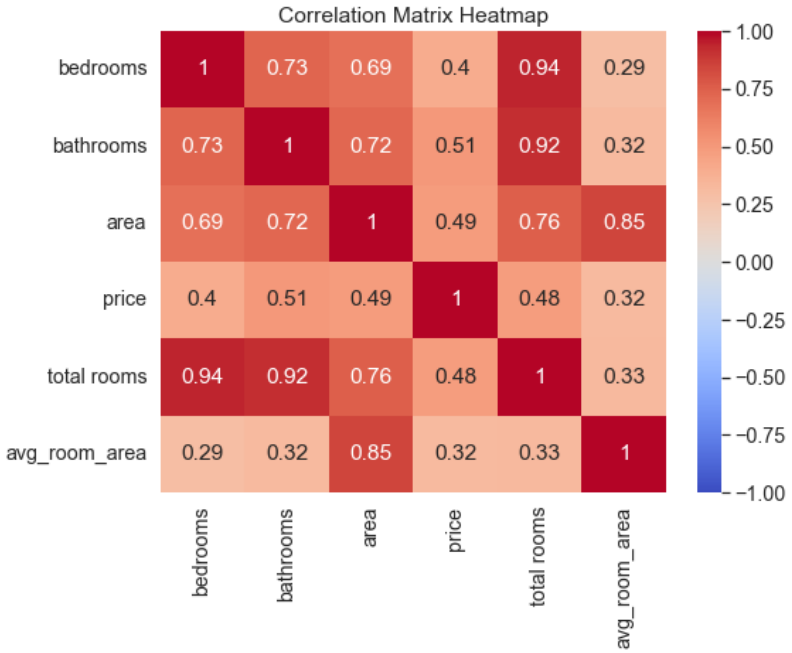
The two charts provided below illustrate the relationship between total area and average room area with the target variable, house price.



From the charts, it is evident that there is a relatively strong correlation between total area and average room area with the target.

6.2 Correlation Analysis

In this section, a statistical validation of the relationships observed earlier is provided. The chart below illustrates the correlation among all features in the dataset.



Overall, this correlation analysis provides a solid foundation for trained model, confirming that key features not only influence the target variable but also relate to each other in meaningful ways.

7 Summary

Through experiments, it was apparent that the results for classic machine learning algorithms did not meet the initial expectations. Despite their established effectiveness in various applications, the classic methods struggled to deliver competitive performance in this particular housing price estimation task.

In contrast, advanced machine learning methods significantly outperformed the classic approaches, achieving the highest scores among all models tested. Notably, algorithms like Random Forest and CatBoost exhibited robust predictive capabilities, validating their suitability for complex regression tasks in this context.

Even though the neural network aimed to leverage deep learning's potential, its performance lagged behind the advanced methods. This outcome highlights the challenges of working with limited data and the simplicity of the network architecture.

Overall, the results suggest that while classic algorithms have their place, advanced machine learning techniques are better suited for tackling intricate prediction problems like housing price estimation.

References

https://www.sciencedirect.com/science/article/pii/S2667305322000217

https://www.researchgate.net/figure/The-architecture-of-our-proposed-mixing-deep-visual-and-textual-features-for-house-price_fig3_343856615

https://arxiv.org/abs/1609.08399

https://github.com/emanhamed/Houses-dataset

GitHub Repository: Multi-Modal House Price Estimation