House Price Prediction Using Decision Tree Regression

1. INTRODUCTION

House price prediction is an important problem in real estate analytics, as accurate estimation of property values helps buyers, sellers, investors, and policy makers make informed decisions. With the increasing availability of housing data, machine learning techniques have become a reliable approach for predicting house prices based on multiple influencing factors.

This project focuses on predicting house prices using a Decision Tree Regression model, which is a supervised machine learning algorithm capable of capturing non-linear relationships between features and the target variable.

2. OBJECTIVE

The primary objectives of this study are:

- To build a Decision Tree Regression model for house price prediction
- To visualize the structure of the trained decision tree
- To evaluate the model's performance using appropriate metrics
- To analyze how different features influence house prices

3. DATASET DESCRIPTION

The dataset used in this project consists of various housing-related attributes such as:

- Property size (e.g., area or square footage)
- Number of bedrooms and bathrooms
- Location-related features
- Other relevant structural or economic factors

The target variable is house price, which the model aims to predict based on the input features.

Before model training, the dataset was cleaned and preprocessed, ensuring that missing values and inconsistencies were handled appropriately.

4. METHODOLOGY

4.1 Data Preprocessing

- The dataset was divided into training and testing sets.
- Feature variables (X) and target variable (y) were separated.
- Only relevant numerical and encoded categorical features were used for modeling.

4.2 Model Selection

A Decision Tree Regressor was selected for this task due to:

- Its interpretability and transparency
- Ability to handle non-linear data
- Minimal need for feature scaling

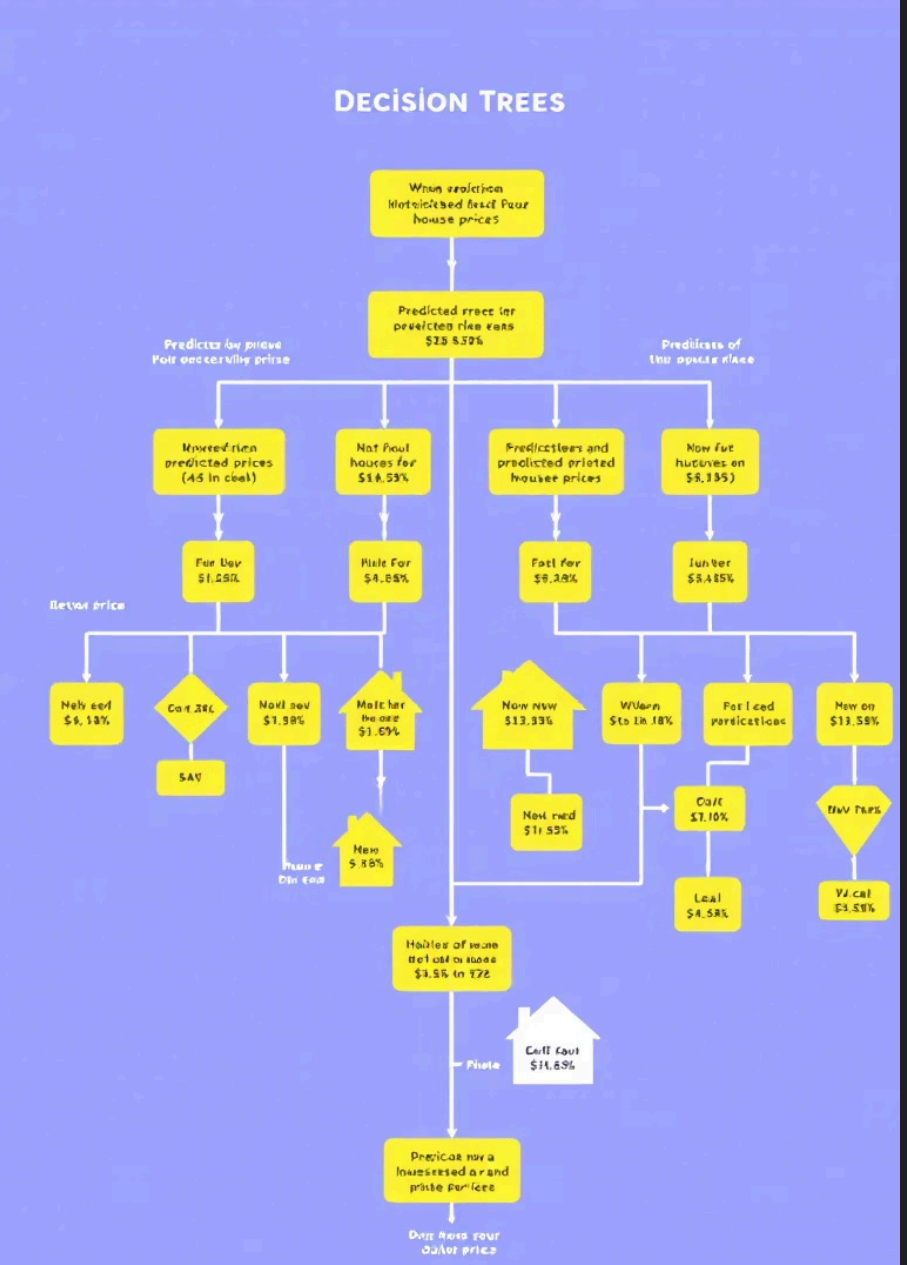
4.3 Model Training

The Decision Tree model was trained using the training dataset. During training, the model learned decision rules that split the data based on feature values to minimize prediction error.

4.4 Model Visualization

The trained Decision Tree was visualized using plot_tree from the sklearn.tree module. This visualization helps in understanding:

- Feature importance
- Decision rules
- Tree depth and complexity



5. DECISION TREE VISUALIZATION

The decision tree visualization clearly illustrates:

- Internal nodes representing decision conditions on features
- Branches showing the outcome of each decision
- Leaf nodes representing predicted house prices

Color coding was used to indicate the range of predicted values, making the structure easy to interpret.

6. MODEL EVALUATION

The performance of the Decision Tree Regression model was evaluated using standard regression metrics, such as:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- R-squared score (R^2)

These metrics indicate how closely the predicted house prices match the actual values. The results demonstrate that the Decision Tree model performs reasonably well on the dataset, capturing complex relationships between features and house prices.

7. RESULTS AND DISCUSSION

The Decision Tree model successfully identified key features influencing house prices. However, due to its tendency to overfit, the model's depth and complexity must be carefully controlled. Visualization of the tree revealed how specific features play a major role in determining price ranges.

While the model provides interpretable results, its performance can be further improved using ensemble methods such as Random Forest or Gradient Boosting.

8. CONCLUSION

This project demonstrates the effectiveness of Decision Tree Regression for house price prediction. The model offers strong interpretability and reasonable predictive performance. Visualization of the decision tree provides valuable insights into the decision-making process of the model.

Future work may include:

- Hyperparameter tuning
- Using ensemble learning techniques
- Comparing performance with other regression models