House Price Prediction: A Machine Learning Model

Kritika Bansal[1]
[1]*Vellore Institute of Technology, 632014, Vellore*
kritika.bansal2022@vitstudent.ac.in,

***Abstract*—This study presents the development, implementation, and comparison of three machine learning models—Random Forest, Decision Tree, and Linear Regression—for predicting house prices using the Boston Housing Dataset. The objective is to determine which model best captures the underlying patterns in the data while minimizing prediction errors. The dataset was pre-processed and analysed, including feature selection, correlation assessment, and overfitting detection. The models were evaluated using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) to assess their predictive performance. Results indicate that Random Forest outperformed the other models due to its ability to handle non-linearity, complex interactions, and reduce the influence of outliers. While Decision Tree also performed well, it exhibited a tendency toward overfitting, potentially affecting generalization. Linear Regression, constrained by its assumption of linearity and independent variables, showed the weakest performance. The findings highlight the importance of ensemble methods like Random Forest in real-world price prediction tasks, where housing prices depend on multiple interacting factors.**

Index Terms— Decision tree, linear regression, machine learning, mean absolute error, random forest, root mean squared error

I.INTRODUCTION

The housing market is a key indicator of a nation's economic health, reflecting trends in urbanization, population growth, and infrastructure development. As more people migrate from rural to urban areas, the demand for housing rises, leading to increased property prices. Additionally, factors such as improved infrastructure—better roads, public transport, and access to utilities—can significantly impact house prices in specific regions. Addressing issues like poor road conditions and inadequate power supply often results in sudden price hikes, making housing affordability a crucial concern. House prices are influenced by multiple factors, including physical characteristics, location, and market trends. The physical condition of a house—such as the number of rooms, size, age, and design—affects its value. Location is another significant determinant, as properties in well-connected areas with access to educational institutions, markets, and healthcare facilities tend to have higher prices. Moreover, real estate developers use strategic marketing tactics to increase demand, such as emphasizing proximity to major roads, shopping centers, and business hubs.

To better understand the trends and influences on housing prices, machine learning (ML) models are increasingly used for predictive analysis. ML enables data-driven decision-making, assisting homeowners, buyers, policymakers, and real estate analysts in evaluating market trends and making informed property-related decisions. This study leverages ML techniques to develop a house price prediction model using the Boston Housing Dataset, a well-known dataset containing 13 influential features such as crime rate (CRIM), average number of rooms per dwelling (RM), and property tax rates (TAX).

The primary objective of this research is to compare the performance of three machine learning models—Random Forest (RF), Decision Tree (DT), and Linear Regression (LR)—in predicting house prices. The models are evaluated using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to determine the most effective approach. Through this comparative analysis, the study aims to identify the best-performing model that can capture complex relationships within housing data and provide accurate price predictions.

II.LITERATURE REVIEW

Housing price prediction plays a significant role in the real estate sector, aiding buyers, sellers, and policymakers in making informed decisions. With the advancements in machine learning (ML), researchers have explored various models to improve predictive accuracy. ML techniques allow for the discovery of complex relationships between housing features and prices, making them highly effective for price estimation [1-2].

Several studies have investigated different ML models for housing price prediction. One such study examined regression-based models, including Linear Regression (LR), Polynomial Regression (PR), Lasso Regression (LaR), and Ridge Regression (RR), applied to the Boston Housing Dataset [4]. The results indicated that Lasso Regression performed better than other regression models, as it could handle complex data and reduce the influence of irrelevant features. This highlights the importance of feature selection in improving predictive performance.

Beyond traditional regression models, other research has explored the effectiveness of decision trees (DT) and ensemble techniques like Random Forest (RF) [5-6]. Studies have found that RF generally outperforms individual decision trees, as it reduces overfitting and captures non-linearity in data [7-8]. Another comparative study on housing price prediction using ML examined multiple algorithms, including LR, DT, and RF, and concluded that RF delivered better accuracy on both training and test datasets [9].

Furthermore, alternative ML approaches such as k-Nearest Neighbors (k-NN) have been investigated [10]. While k-NN is useful for identifying patterns in small datasets, its efficiency can decline when dealing with high-dimensional data. On the other hand, RF has shown greater robustness in handling diverse housing attributes, making it a strong candidate for price prediction tasks [11].

Studies have also emphasized the role of external factors such as location, accessibility to infrastructure, and economic conditions in determining house prices [12]. One investigation into housing price trends suggested that incorporating location-based attributes significantly improves model performance [13]. Additionally, a broader analysis of real estate ML models highlighted that ensemble methods, particularly RF and gradient boosting techniques, tend to yield the most reliable predictions [14].

While previous research has provided valuable insights into the effectiveness of different ML models, many studies focus on only one or two approaches. This study aims to conduct a comparative evaluation of LR, DT, RF, and k-NN models on the Boston Housing Dataset to identify the most effective model for predicting house prices. By analysing feature importance, model accuracy, and generalization ability, this study seeks to contribute to the ongoing development of data-driven solutions for real estate price prediction.

III. ARCHITECTURE

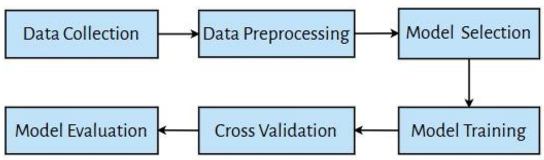

Fig. 1: System Architecture

IV. METHODOLOGY

A. Data Collection

The Boston Housing Dataset is widely used for predictive modeling in real estate, containing key attributes that influence housing prices. The dataset includes various socio-economic, geographical, and structural features of houses, such as crime rate, number of rooms, accessibility to highways, and property tax rates. Understanding these features is crucial, as each contributes differently to the final house price prediction. Below is a detailed explanation of the dataset's variables.

TABLE I. DATASET FEATURES AND DESCRIPTION

S. no.	Feature	Description
1	CRIM	per capita crime rate by town
2	ZN	proportion of residential land zoned for lots over 25,000 sq.ft.
3	INDUS	proportion of non-retail business acres per town
4	CHAS	Charles River dummy variable (1 if tract bounds river; 0 otherwise)
5	NOX	nitric oxides concentration (parts per 10 million)
6	RM	average number of rooms per dwelling
7	AGE	proportion of owner-occupied units built prior to 1940
8	DIS	weighted distances to five Boston employment centers
9	RAD	index of accessibility to radial highways
10	TAX	full-value property-tax rate per $10,000
11	PTRATIO	pupil-teacher ratio by town
12	B	1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town
13	LSTAT	% lower status of the population
14	MEDV	Median value of owner-occupied homes in $1000's

After loading the dataset into Python, the initial step was to inspect the data by displaying a sample view of the first 5 rows. This provided an immediate overview of the dataset's structure, allowing for a quick assessment of the data's format, column names, and the values within each column.

TABLE II. DATASET STRUCTURE

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0	0.469	NaN	61.1	4.9671	2	242	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222	18.7	396.90	5.33	36.2

A detailed exploratory analysis on the dataset derived a useful summary statistic and provided a comprehensive overview of the data's central tendencies and dispersion. Various measures such as mean, median, mode, standard deviation, and range were calculated, shedding light on the dataset's distribution and variability (see Table III).

TABLE III. DATASET'S DISTRIBUTION AND VARIABILITY

	count	mean	std	min	25%	50%	75%	max
CRIM	506.0	3.613524	8.601545	0.00632	0.082045	0.25651	3.677083	88.9762
ZN	506.0	11.363636	23.322453	0.00000	0.000000	0.00000	12.500000	100.0000
INDUS	506.0	11.136779	6.860353	0.46000	5.190000	9.69000	18.100000	27.7400
CHAS	506.0	0.069170	0.253994	0.00000	0.000000	0.00000	0.000000	1.0000
NOX	506.0	0.554695	0.115878	0.38500	0.449000	0.53800	0.624000	0.8710
RM	506.0	6.284634	0.702617	3.56100	5.885500	6.20850	6.623500	8.7800
AGE	506.0	68.574901	28.148861	2.90000	45.025000	77.50000	94.075000	100.0000
DIS	506.0	3.795043	2.105710	1.12960	2.100175	3.20745	5.188425	12.1265
RAD	506.0	9.549407	8.707259	1.00000	4.000000	5.00000	24.000000	24.0000
TAX	506.0	408.237154	168.537116	187.00000	279.000000	330.00000	666.000000	711.0000
PTRATIO	506.0	18.455534	2.164946	12.60000	17.400000	19.05000	20.200000	22.0000
B	506.0	356.674032	91.294864	0.32000	375.377500	391.44000	396.225000	396.9000
LSTAT	506.0	12.653063	7.141062	1.73000	6.950000	11.36000	16.955000	37.9700
MEDV	506.0	22.532806	9.197104	5.00000	17.025000	21.20000	25.000000	50.0000

Next, the histograms of all the independent features were visualized. Histograms provide a visual representation of the frequency distribution of numerical features in a dataset. By analyzing histograms, we can identify whether a feature follows a normal distribution, is skewed, or has multiple peaks. This insight is crucial in determining the nature of the data and selecting appropriate preprocessing techniques.

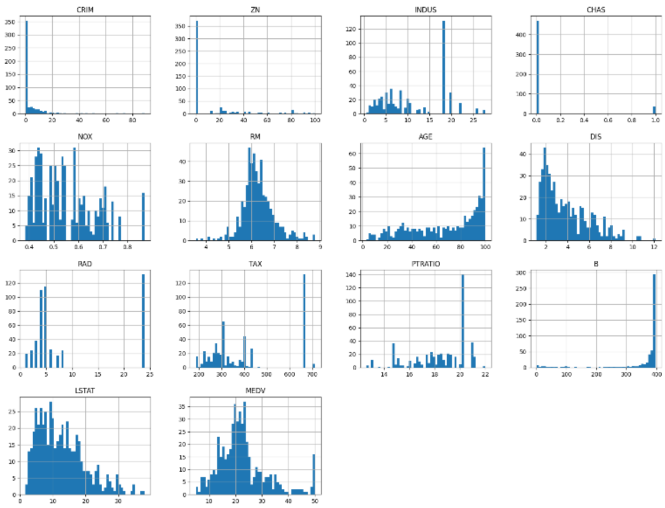

Fig. 2: Histograms of all the independent features

B. Data Preprocessing

To ensure the dataset is clean and optimized for machine learning models, several preprocessing techniques were applied. These steps help improve the accuracy and efficiency of the model.

Handling Missing Values
Missing values in the dataset can negatively impact model performance. A thorough check was conducted to identify missing values, and appropriate imputation technique i.e. median was applied to replace missing values.

Feature Scaling
Since the dataset contains numerical features with different ranges, scaling was performed to standardize the values. Standardization (Z-score Normalization) was applied to ensure that the features have comparable distributions, improving model performance.

Correlation Analysis
Correlation analysis was conducted to understand the relationship between the target variable (*MEDV*, the median house price) and other independent features in the dataset. A positive correlation indicated that as the independent variable increased, *MEDV* also tended to increase, whereas a negative correlation suggested that an increase in the independent variable led to a decrease in *MEDV*. Below is a table depicting the correlation values between *MEDV* and various features in the dataset.

TABLE IV. CORRELATION VALUES

MEDV	1.000000
RM	0.677546
B	0.361761
ZN	0.339741
DIS	0.240451
CHAS	0.205066
AGE	-0.364596
RAD	-0.374693
CRIM	-0.393715
NOX	-0.422873
TAX	-0.456657
INDUS	-0.473516
PTRATIO	-0.493534
LSTAT	-0.740494

Scatter plots were used to visually explore these relationships. Features with strong correlations, such as RM (positive) and LSTAT (negative), exhibited clear linear trends with *MEDV*. Meanwhile, features with weaker correlations, such as CHAS and DIS, displayed more scattered distributions. This analysis helped in selecting the most relevant features for the predictive model, ensuring better accuracy.

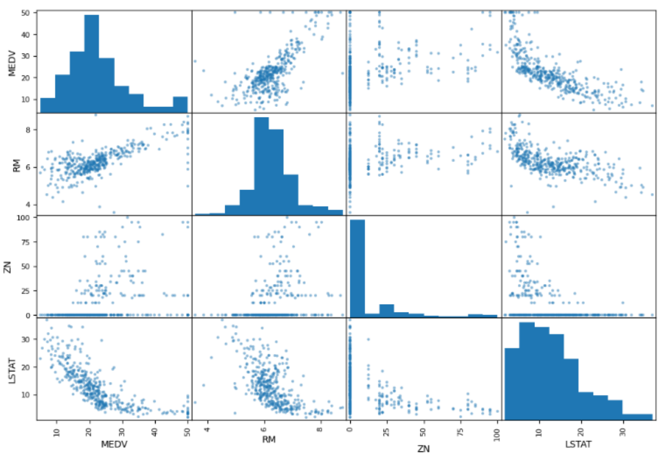

Fig. 3: Scatter plots of all the independent features

To gain further insights into the relationships among the variables, a correlation heatmap was generated. This heatmap visually represented the strength and direction of linear relationships between different features using color intensity.

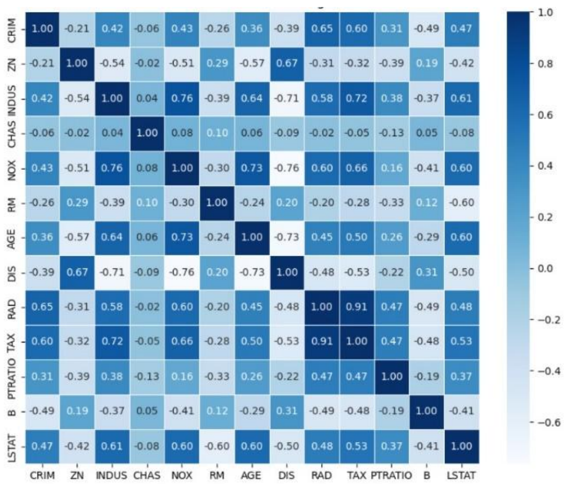

Fig. 4: Correlation Matrix of housing data

C. Model Selection and Training

The importance of model selection lies in the fact that different models have different strengths and weaknesses. Some models may perform well on certain types of datasets while underperforming on others. Furthermore, different models have different hyperparameters that can be tuned to improve their performance on a given dataset. Choosing the wrong model or failing to optimize the hyperparameters can lead to poor predictions, decreased model performance, and adversely affect the overall results prediction accuracy. Therefore, when adopting ML model for predicting the housing prices, selecting an appropriate model is crucial to achieve accurate predictions and maximize the overall performance of the model. These models were trained on the processed training dataset to learn the relationship between the independent features and the target variable, MEDV.

(a) Linear Regression

Linear Regression is a simple yet widely used machine learning algorithm for predicting a continuous target variable based on one or more independent variables. In the context of Boston housing price prediction, LR is used to model the relationship between features such as crime rate, number of rooms, and distance to employment centres, and the target variable, i.e., the median value of owner-occupied homes (MEDV).

The relationship between the target variable Y and independent features $X_1, X_2, \ldots, X_n$ can be expressed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_n X_n + \varepsilon$$

Where β_0 is the intercept, $\beta 1, \beta 2, \ldots, \beta n$ are the coefficients that represent the weight of each feature and ε is the error term capturing noise or unexplained variance.

(b) Decision Tree

Decision Tree regression is a non-linear model that splits the dataset into smaller subsets based on feature values to predict the target variable. In this method, the data is divided at each node using the feature and threshold that result in the greatest reduction of prediction error. For the Boston Housing Dataset, DTs are capable of learning non-linear relationships between variables like number of rooms, crime rate, and accessibility to highways, and the target house price. The DT algorithm recursively splits the dataset until it reaches a stopping criterion (e.g., maximum depth). The prediction for a new data point is computed as the average value of the target variable in the corresponding leaf node:

$$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Where y_i are the values of training samples in the same leaf and n is the number of those samples. Although DTs are interpretable and easy to implement, they are prone to overfitting if not pruned or regularized.

(c) Random Forest

Random Forest is an ensemble learning method that builds multiple decision trees during training and outputs the average prediction of the individual trees. By introducing randomness in feature selection, Random Forest reduces the variance observed in single decision trees, thereby improving accuracy and generalization. It is highly effective in handling non-linear relationships, managing missing values, and identifying feature importance.

The general form of the Random Forest regression model can be written as:

$$y = f(X) + \varepsilon$$

Where f(X) is the function learned by the forest and ε is the random error term.

To compare the performance of different models, the following evaluation metrics were used:

(a) Root Mean Square Error (RMSE)

RMSE measures the standard deviation of the prediction errors and is useful for understanding how much the predicted prices deviate from the actual prices. It is calculated using the formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i + \hat{y}_i)^2}$$

where y_i is the actual house price, $\hat{y}_i$ is the predicted house price and n is the total number of predictions.

RMSE is particularly useful because it penalizes larger errors more than smaller ones. This is due to the squaring of differences before averaging, which ensures that predictions with large deviations from actual values contribute more to the final error. A lower RMSE indicates that the model's predictions are closer to actual house prices, making it a crucial metric for evaluating model accuracy. However, RMSE can be sensitive to outliers, meaning that a few extreme mispredictions can significantly increase the error value.

(b) Mean Absolute Error (MAE)

MAE measures the average magnitude of errors in the predictions without considering their direction (overestimation or underestimation). The formula for MAE is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i + \hat{y}_i|$$

MAE does not square the differences, meaning that all errors contribute equally to the final value. This makes MAE less sensitive to outliers, providing a more realistic measure of how far, on average, the predicted house prices deviate from actual values. A lower MAE means that the model is making accurate predictions with smaller deviations from actual house prices.

V. ABLATION STUDY

An ablation study was conducted to compare the performance of three machine learning models—Linear Regression, Decision Tree, and Random Forest—on the Boston Housing Dataset. The purpose of this study was to assess the contribution of each algorithm in terms of prediction accuracy and error, and to identify the most suitable model for house price prediction.

A. Model Performance Evaluation

To evaluate the predictive accuracy of our models, we calculated two key performance metrics — RMSE and MAE:

TABLE V. MODELS AND THEIR RMSE AND MAE VALUES

Model	RMSE	MAE
Linear Regression	4.84131	3.41550
Decision Tree	0.0	0.0
Random Forest	1.15061	0.81565

From the table, we observe that the Random Forest model outperforms both Linear Regression and Decision Tree models by achieving the lowest RMSE and MAE score. This suggests that the Random Forest model provides the most accurate predictions for house prices.

The Linear Regression model performed the worst because it assumes a linear relationship between features and the target variable, which may not always hold true in real-world housing data. The Decision Tree model improved performance by capturing non-linear relationships but was still prone to overfitting. The Random Forest model, being an ensemble technique, reduced overfitting and provided the best overall performance.

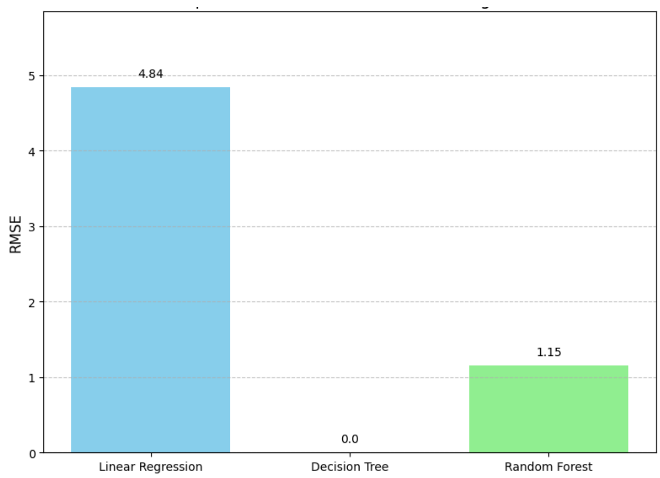

Fig. 5: Comparison of Model Performance using RMSE

The bar chart presents the Root Mean Square Error (RMSE) values for each model. Random Forest achieves a significantly lower RMSE compared to Linear Regression, indicating better prediction accuracy. The Decision Tree appears to have an RMSE of zero, which suggests overfitting—likely memorizing the training data without generalizing well. Overall, Random Forest demonstrates strong performance with minimal error.

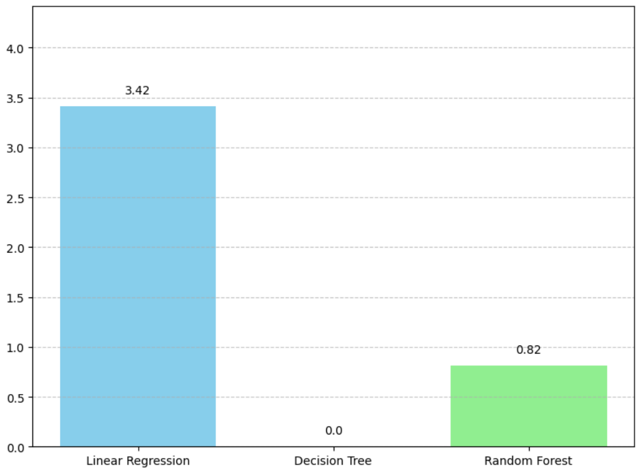

Fig. 6: Comparison of Model Performance using MAE

This bar chart compares the Mean Absolute Error (MAE) of the three models. Random Forest again shows the lowest MAE, confirming its accuracy and reliability in predicting house prices. Linear Regression has a higher MAE, indicating larger average prediction errors, while Decision Tree shows an unusually low MAE, which may reflect overfitting rather than true performance.

B. Cross-Validation Results

To ensure that our models generalized well across different data subsets, we applied k-fold cross-validation and computed both the Mean RMSE and Mean MAE, along with their respective standard deviations. This approach helps assess how consistently each model performs across different data splits. The results are summarized in the table below:

TABLE VI. CROSS-VALIDATION RESULTS TABLE

Model	Mean RMSE	SD (RMSE)	Mean MAE	SD (MAE)
Linear Regression	5.03773	1.05626	3.58292	0.52991
Decision Tree	4.50372	1.08047	3.17326	0.70151
Random Forest	3.35434	0.67942	2.35300	0.34140

Random Forest achieved the lowest Mean RMSE and Mean MAE, confirming its superior predictive accuracy. The standard deviation of Random Forest is also the lowest, indicating that its performance remains consistent across different data folds. Decision Tree showed moderate performance, but its higher standard deviation suggests some degree of overfitting in certain folds. Linear Regression had the highest error values, reinforcing the fact that it struggles with capturing non-linear relationships in the data.

VI. RESULTS AND DISCUSSIONS

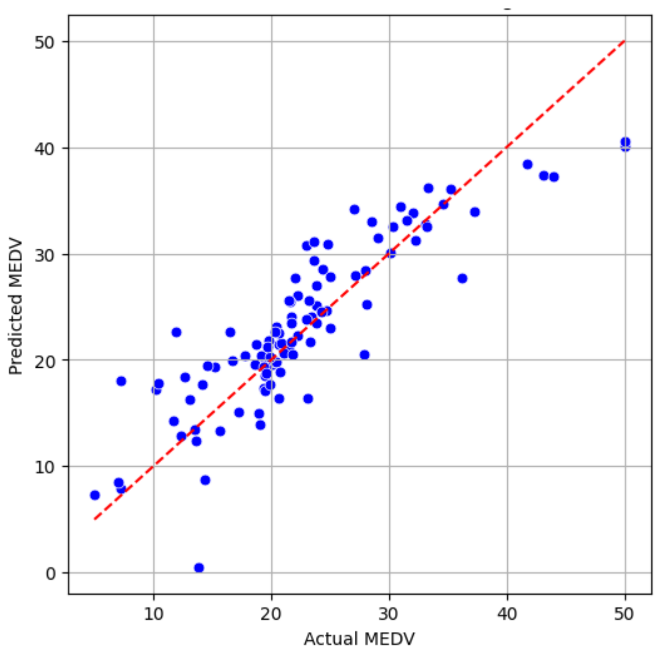

Fig. 7: Actual vs Predicted Prices (Linear Regression)

In the actual vs predicted plot for Linear Regression, the data points are scattered somewhat broadly around the diagonal reference line. This suggests that while the model is capturing a general trend between the input features and the target variable (MEDV), its predictions are not consistently accurate for all ranges. Some points deviate significantly from the diagonal, showing that the linear model tends to either overestimate or underestimate values at the extremes.

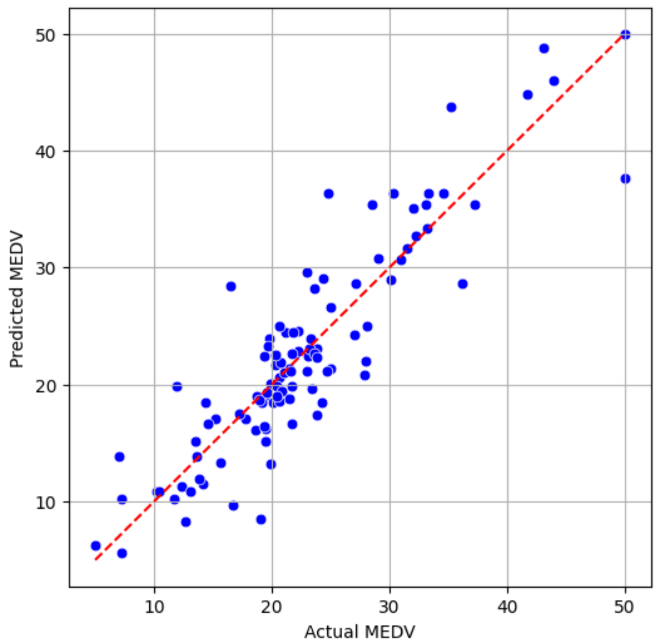

Fig. 8: Actual vs Predicted Prices (Decision Tree)

The plot for the Decision Tree model shows a different pattern. The predictions appear as horizontal clusters, where many actual values are mapped to a limited set of predicted values. This blocky behavior is a result of how decision trees split data into regions and assign a constant value within each. While some predictions are close to the diagonal, the model exhibits higher variance and signs of overfitting, especially for values on the lower and higher ends of the target range. The clustering also suggests less smooth prediction transitions, especially compared to ensemble models.

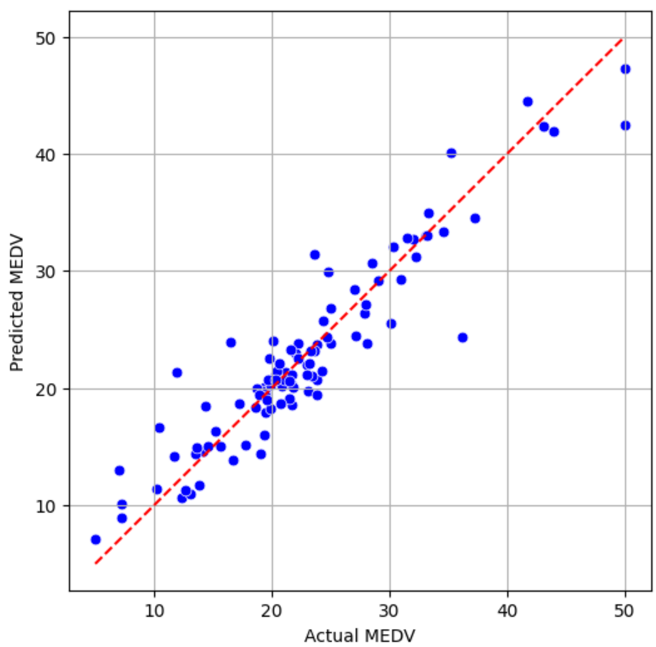

Fig. 9: Actual vs Predicted Prices (Random Forest)

The Random Forest plot shows a tight cluster of data points around the diagonal line, indicating strong agreement between the predicted and actual values. The scatter is denser and more linear compared to the other models, with fewer extreme deviations. This reflects Random Forest's ability to generalize better by combining predictions from multiple decision trees, reducing overfitting and increasing accuracy. The closeness of most points to the diagonal demonstrates that this model has captured both linear and non-linear patterns in the dataset more effectively.

Among the three models, Random Forest shows the most accurate and consistent predictions, with points closely aligned along the diagonal in the actual vs predicted plot. Linear Regression captures general trends but struggles with complex patterns, while Decision Tree displays step-wise predictions and higher variability due to overfitting. Overall, Random Forest handles non-linear relationships better and generalizes well, making it the most reliable model for predicting housing prices in this study.

VII.CONCLUSION AND FUTURE SCOPE

This project successfully implemented and compared three machine learning models—Linear Regression, Decision Tree, and Random Forest—to predict housing prices using the Boston Housing Dataset. After performing thorough data preprocessing, correlation analysis, and feature scaling, each model was trained and evaluated using both cross-validation and performance metrics such as RMSE, MAE, and R^2 Score. Visual tools like scatter plots and boxplots were used to assess prediction accuracy and consistency.

The results clearly indicate that Random Forest outperformed the other models in terms of both accuracy and stability. Its ensemble learning approach allowed it to effectively capture complex, non-linear relationships while reducing overfitting. While Linear Regression provided a strong baseline and Decision Tree captured some non-linear behaviour, both showed limitations in generalization and consistency.

Overall, the study highlights the effectiveness of ensemble methods, particularly Random Forest, in real estate price prediction tasks. These findings can be valuable for building reliable, data-driven decision support systems in the housing and real estate sectors.

While this study demonstrated the effectiveness of machine learning models—particularly Random Forest—in predicting housing prices using the Boston Housing Dataset, there are several opportunities for further enhancement. Future work could involve experimenting with larger and more recent real-world datasets that include additional features such as neighborhood amenities, crime statistics, economic indicators, and infrastructure quality. Incorporating geospatial data through GIS integration and interactive map visualizations could also provide deeper location-based insights.

Additionally, exploring advanced ensemble models like Gradient Boosting Machines (e.g., XGBoost or LightGBM) may further improve prediction accuracy. The use of hyperparameter optimization techniques such as Grid Search or Bayesian Optimization can also help fine-tune model performance. Finally, deploying the trained model into a web-based or mobile application would allow real estate professionals and potential buyers to interactively estimate property values, making the solution practical and user-friendly.

VIII.REFERENCES

[1] Glaeser, E.L., Gyourko, J., Morales, E. and Nathanson, C.G., 2014. Housing dynamics: An urban approach. Journal of Urban Economics, 81, pp.45-56

[2] UN-Habitat (United Nations Human Settlements Programme). (2008). "The State of the World's Cities 2008/2009: Harmonious Cities.

[3] Kreuzberger, D., Kühl, N. and Hirschl, S., 2023. Machine learning operations (mlops): Overview, definition, and architecture. IEEE Access.

[4] Masini, R.P., Medeiros, M.C. and Mendes, E.F., 2023. Machine learning advances for time series forecasting. Journal of economic surveys, 37(1), pp.76-111.

[5] Pham, T.Q.D., Le-Hong, T. and Tran, X.V., 2023. Efficient estimation and optimization of building costs using machine learning. International Journal of Construction Management, 23(5), pp.909- 921.

[6] Møller, S.V., Pedersen, T., Montes Schütte, E.C. and Timmermann, A., 2023. Search and predictability of prices in the housing market. Management Science.

[7] M. Ahtesham, N. Z. Bawany and K. Fatima, "House Price Prediction using Machine Learning Algorithm - The Case of Karachi City, Pakistan," 2020 21st International Arab Conference on Information Technology (ACIT), Giza, Egypt, 2020, pp. 1-5

[8] A. P. Singh, K. Rastogi and S. Rajpoot, "House Price Prediction Using Machine Learning," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2021, pp. 203-206

[9] S. Sharma, D. Arora, G. Shankar, P. Sharma and V. Motwani, "House Price Prediction using Machine Learning Algorithm," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 982-986

[10] N. T. M. Sagala and L. H. Cendriawan, "House Price Prediction Using Linier Regression," 2022 IEEE 8th International Conference on Computing, Engineering and Design (ICCED), Sukabumi, Indonesia, 2022, pp. 1-5

[11] C. Chee Kin, Z. Arabee Bin Abdul Salam and K. Batcha Nowshath, "Machine Learning based House Price Prediction Model," 2022 International Conference on Edge Computing and Applications (ICECAA), Tamilnadu, India, 2022, pp. 1423-1426

[12] A. Gupta, S. K. Dargar and A. Dargar, "House Prices Prediction Using Machine Learning Regression Models," 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, Karnataka, India, 2022, pp. 1-5

[13] S. Bai, "Boston house price prediction: machine learning," 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2022, pp. 1678-1684

[14] S. Sanyal, S. Kumar Biswas, D. Das, M. Chakraborty and B. Purkayastha, "Boston House Price Prediction Using Regression Models," 2022 2nd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2022, pp. 1-6

[15] Adetunji, A.B., Akande, O.N., Ajala, F.A., Oyewo, O., Akande, Y.F. and Oluwadara, G., 2022. House price prediction using random forest machine learning technique. Procedia Computer Science, 199, pp.806 813.