# Recommendation of Regression Models for Real Estate Price Prediction using Multi-Criteria Decision Making

Ajay Kumar

Original scientific article

***Abstract*—Accurate prediction of real estate prices is an essential task for establishing real estate policies. Even though various regression models for real estate price prediction have been developed so far, selecting the most suitable regression model is a challenging task since the performance of different regression models varies for different accuracy measures. This paper aims to recommend the most suitable regression model for real estate price prediction, considering various performance measures altogether using multi-criteria decision making (MCDM). The evaluation of regression models involves a number of competing accuracy measures; hence, choosing the best regression model for predicting real estate price is modeled as the MCDM problem in the proposed approach. An experimental study is designed using 22 regression models, three MCDM methods, six performance measures, and three real estate price datasets to validate the proposed approach. Experimental outcomes show that Gradient Boosting, Random Forest, and Ridge Regression are recommended as the best regression models based on MCDM ranking. The results of the experimental study show that the proposed MCDM-based strategy can be utilized effectively in real estate industries to choose the best regression model for predicting real estate prices by optimizing several competing accuracy measures.**

***Index terms*—Real Estate Price Prediction, Regression Models, MCDM, Weighted Sum Model (WSM), Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), and Evaluation based on Distance from Average Solution (EDAS).**

## I. Introduction

In day-to-day life, real estate price prediction plays a significant role. Real estate agencies and people sell or buy houses; the agencies buy to run a business, and people buy to live in or as an investment. Either way, everyone should get exactly what they pay for. Undervaluation and overvaluation in the real estate market have always been an issue. Hence there is always a need for an accurate model for predicting real estate prices. An exact model for real estate market prediction can benefit real estate sellers, buyers, and economic experts.

Price prediction of real estate is one of the most often researched areas in which the capabilities of machine learning models are investigated [1]. Even though various regression models for real estate price prediction have been developed so far, selecting the most suitable regression model is a challenging task since the performance of different regression models varies for different accuracy measures. Furthermore, the widely accepted No Free Lunch (NFL) theory in computational intelligence disproves the existence of a single prediction strategy that will outperform other techniques across all competing model performance measures for a specific application domain [2]. Thus, it becomes difficult to decide which regression technique should be used to build accurate real estate price prediction models. This study provides an MCDM-based framework for evaluating different regression models for real estate price prediction by optimizing several performance measures simultaneously. To the best of the author's knowledge, no one has attempted to evaluate regression models for real estate price prediction considering the optimization of various performance measures taken into account altogether. However, some researchers (described in detail in the related work section) have considered more than one performance measure to evaluate the regression models for real estate price prediction, but it can be observed that in their study, they have proposed the most suitable regression model taking into account only a single performance measure at a time.

The contribution of this paper is summarized as follows:

- This paper proposes a novel approach based on MCDM to recommend the most suitable regression model among the different available regression models for real estate price prediction considering several conflicting accuracy measures altogether.
- The proposed approach uses three MCDM methods- WSM, TOPSIS, and EDAS since the ranking of regression models assigned by more than one MCDM method will be more trustworthy.
- For the validation of the proposed approach, twenty-two regression models were evaluated for predicting real estate prices on three real estate price datasets (as described in detail in section IV.*A*) considering six performance measures taken into account altogether.
- The regression model that gets the highest rank by all the three MCDM methods (WSM, TOPSIS, and EDAS) is recommended as the most suitable regression model for real estate price prediction.

Manuscript received June 14, 2023; revised July 4, 2023. Date of publication September 11, 2023. Date of current version September 11, 2023.

Ajay Kumar is with the Department of Information Technology, KIET Group of Institutions Delhi-NCR, Ghaziabad-201206, India (email: ajaygarg100@gmail.com).

Digital Object Identifier (DOI): 10.24138/jcomss-2023-0102

The remaining part of the paper is laid out as follows. Section II highlights the related work. The proposed approach, selected regression models, performance metrics, and MCDM method employed in this study are described in section III. The datasets used in this study, as well as the experimental strategy used to validate the suggested approach, are explained in Section IV. Section V highlights the results, discussion, and findings along with the MCDM ranking of regression models. At last, section VI concludes the study.

## II. Related Work

Several regression models for real estate price prediction developed by different researchers in previous studies are presented in this section. In [3] authors conducted a comparative study for house price prediction using seven regression models: Ridge, Lasso, Multiple Linear Regression (MLR), Gradient Boosting, Ada Boost, and elastic net. In [4] authors proposed a model for predicting real estate prices using four machine learning models- backpropagation neural network (BPNN), general regression neural network (GRNN), least-square support vector regression (LSSVR), and regression tree (CART). The authors examined these four machine learning models for predicting real estate prices based on absolute percentage error and found that LSSVR is more efficient than the other three models.

In [5] authors use three machine learning techniques- MLR, decision tree regression, and decision tree classification for house price prediction modeling. In [6] author used multiple linear regression to estimate house prices using California house price data as a case study in his research. In [7] authors proposed a method for predicting the prices of houses in real estate by applying three regression models, namely, Gradient Boosting, random forest, and linear regression. Based on experimental results, they showed that the Gradient Boosting model produced efficient results in real estate price prediction compared to the other two models, random forest and linear regression.

In [8] authors conducted an empirical study comparing three machine learning regression models, namely, generalized regression neural network, feed-forward neural network, and support vector machine for house price prediction in Turkey. In [9] authors conducted a survey study to demonstrate the importance of machine learning regression models in house price prediction. They analyzed and verified support vector regression and artificial neural networks as the most appropriate models for house price prediction.

In [10] authors developed a real estate price prediction model using machine learning techniques such as Gradient Boosting, random forest, and AdaBoost. In [11] author developed a house price prediction model using XGBoost regression. A joint self-attention mechanism-based deep learning model for predicting house prices was proposed by authors in [12]. In [13] authors developed long short-term memory (LSTM) based model for predicting real estate prices. In [14] author propose a hybrid model for predicting real estate prices using PSO and MLR.

In [15] authors demonstrate the use of machine learning regression models for real estate price prediction. The authors provide an overview of the use of existing machine learning techniques for predicting real estate prices on two different real estate price datasets. In [16] authors have highlighted the usefulness of artificial neural networks in developing real estate price-prediction models. In [17] authors applied three machine learning regression models, namely, random forest, linear regression, and decision tree for real estate price prediction. They compared the results of these three machine learning models using the evaluation metrics MAE and MSE, considering each metric separately. In [18] authors proposed a house price prediction model using extreme gradient boosting. In [19] authors propose a hybrid method that classifies the houses for which the cluster is unknown, identifies different housing clusters from the available data, and predicts house prices by developing unique prediction models for each class. The performance of the suggested hybrid model was compared with eight machine learning techniques- XGBoost regression, random forest, decision tree, AdaBoost, support vector regression, ridge regression, Lasso, and multiple linear regression. Two performance measures- RMSE and MAPE were used for the comparative study. However, they used one performance measure at a time.

In [20] authors developed house price prediction models using four machine learning techniques- random forest regressor, Histogram gradient boosting regressor, gradient boosting regressor, and linear regression. They have used only one performance measure to evaluate the performance of their proposed models for house price prediction. In [21] authors evaluate the performance of two machine learning models- random forest regressor and decision tree regressor for predicting house prices in Mumbai. The authors used four performance measures- root mean square error (RMSE), mean absolute error (MAE), mean squared error (MSE), and R-squared. However, they have not considered these four performance measures altogether.

After reviewing the research in this area, it has been observed that most researchers have focused on developing various regression models for real estate price prediction. However, no one has focused on the issue of choosing the best regression model among the various available regression models by taking multiple accuracy measures into account altogether. This study proposes an MCDM-based method to recommend the best regression model among various available regression models, considering many competing accuracy measures. To the author's knowledge, no one has made an effort to tackle this problem utilizing an MCDM-based strategy.

## III. Research Methods

### *A. Proposed Method*

This research suggests an MCDM-based methodology to create a ranking index of regression models for real estate price prediction, considering various performance measures taken into account altogether. MCDM is a well-known technique for selecting the best alternative among the different available alternatives considering various criteria [22]. The literature contains a number of MCDM methods. However, all the MCDM methods use a decision matrix as the input for producing the ranking of alternatives. A decision matrix is a matrix that represents the performance of alternatives with respect to evaluation criteria.

In the proposed approach, because of the involvement of various performance measures (evaluation criteria), the issue of selecting the most suitable regression model (alternative) for real estate price prediction can be modeled as an MCDM problem. A detailed description of the proposed approach is given in the following steps, followed by the graphical representation in Fig. 1.

Step 1: Train the different available regression models, let's say *m*, over the real estate price prediction dataset.

Step 2: Measure the performance *m* regression models in terms of the various performance measures, let's say *n*.

Step 3: Store the performance results of *m* regression models with respect to *n* performance measures obtained in step 2 in a decision matrix of order $m \times n$.

Step 4: Apply the three MCDM methods- WSM, TOPSIS, and EDAS (thoroughly discussed in section III.*D*) on the decision matrix, to obtain the ranks of the regression models for real estate price prediction.

Step 5: Based on the rank produced by the three MCDM methods, the regression model with the highest rank is recommended for predicting real estate prices.

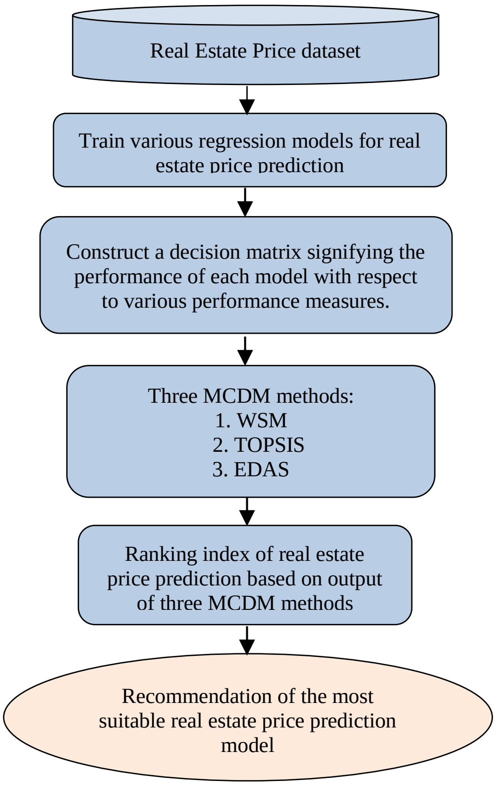


Fig. 1. Graphical representation of the proposed approach

## *B. Regression Models*

As many regression models are available, it is impossible to take all the regression models to validate the proposed approach. This study chooses twenty-two regression models for predicting real estate prices used in the previous studies, as described in the related work section. Table I contains a listing of all twenty-two regression models.

TABLE I
REGRESSION MODELS

| Sr. No. | Regression Models |
|---|---|
| 1 | Multiple linear Regression (MLR) |
| 2 | Isotonic Regression (IR) |
| 3 | Multilayer Perceptron (MLP) |
| 4 | Ridge Regression (RIDGE) |
| 5 | Elastic Net (EN) |
| 6 | Pace Regression (PR) |
| 7 | Support Vector Regression (SVR) |
| 8 | KSTAR |
| 9 | Locally Weighted Learning (LWL) |
| 10 | Additive Regression (AR) |
| 11 | Bagging |
| 12 | Ensemble Selection (ES) |
| 13 | Regression by Discretization (REGBYDISC) |
| 14 | Conjunctive Rule learner (CR) |
| 15 | Decision Table (DT) |
| 16 | M5Rules |
| 17 | M5P |
| 18 | Random Forests (RF) |
| 19 | Reduced Error Pruned Tree (REPTREE) |
| 20 | ADABOOST |
| 21 | Gradient Boosting (GRADBOOST) |
| 22 | K-Nearest Neighbors (KNN) |

## *C. Performance Measures*

This study chooses six performance measures as evaluation criteria for evaluating regression models for real estate price prediction. Among these six criteria, four are cost criteria (Mean Absolute Error, Standard Deviation of Absolute Errors, Normalized Root Mean Square Error (NRMSE), and Mean Balanced Relative Error), and two are beneficial criteria (Pearson's Correlation Coefficient and Pred (0.25)). The minimum value is expected in the case of the cost criterion. On the other hand, the maximum value is desired for the beneficial criterion. The following paragraph describes a brief description of six performance measures.

Given *n* is the total number of observations. For $i^{th}$ observation, $u_i$ represents the actual price, and $v_i$ represents the predicted price. Below is the description of all six performance measures is given.

- Mean Balance Relative Error (MBRE) is the mean of balanced relative errors [23] and can be calculated using "(1)".

$$MBRE = \frac{1}{n}\sum_{i=1}^{n}\frac{|u_i - v_i|}{\min(u_i, v_i)} \quad (1)$$

- NRMSE can be calculated as follows:

$$NRMSE = \sqrt[2]{\frac{\sum_{i=1}^{n}(u_i - v_i)^2}{\sum_{i=1}^{n}u_i^2}} \quad (2)$$

- MAE is the mean value of absolute errors and can be calculated using “(3)”.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|u_i - v_i| \quad (3)$$

- SdARE is the Standard deviation of absolute residual errors, where the absolute residual error is the difference between actual and estimated values.
- The correlation coefficient (r) [24] shows the relationship's strength between two variables (in this paper, predicted price value and actual price value). The formula to calculate Pearson’s correlation coefficient used in this paper is given below.

$$r = \frac{n(\sum vu) - (\sum v)(\sum u)}{\sqrt{\left[n\sum v^2 - (\sum v)^2\right]\left[n\sum u^2 - (\sum u)^2\right]}} \quad (4)$$

- Pred (0.25) [23] is the number of predicted values for which the magnitude of relative error (MRE) ≤ 0.25 is divided by the number of observations, and MRE is given in “(5)”.

$$MRE = \|u_i - v_i\| / u_i \quad (5)$$

### *D. MCDM Methods*

For making judgments when there are competing criteria, there are numerous MCDM techniques accessible. Since there is currently no methodology that enables the selection of a particular MCDM method, we have taken into consideration three MCDM approaches- WSM [25], TOPSIS [26], and EDAS [27], to assess regression models for real estate price prediction. The ranking of regression models generated by many MCDM approaches will be more reliable than a single MCDM approach's ranking. Detailed procedure of three MCDM methods- WSM, TOPSIS, and EDAS is given in this section.

These MCDM methods take as input a decision matrix, let's say $D_{a\times c}$, where $a$ and $c$ are the number of real estate price prediction regression models as the alternatives and performance measures as criteria, respectively. In this paper, values of $a$ and $c$ are twenty-two and six, respectively. In the matrix $D_{a\times c}$, each entry $D_{ij}$ denotes the value of the performance measure $j$ for the corresponding $i^{th}$ real estate price prediction regression model. Further detailed procedure of three MCDM methods is explained as follows:

#### D.1 WSM (Weighted Sum Model)

Step1: Calculate the normalized decision matrix $ND_{a\times c}$ as follows:

$$ND_{a\times c} = \frac{D_{ij}}{\sqrt{\sum_{i=1}^{a} D_{ij}^{2}}} \quad (6)$$

Step2: Calculate the total benefit and cost of twenty-two real estate price prediction regression models (alternatives) using the following equations.

$$A_i^{\text{total benefit}} = \sum_{j=1}^{m} w_j ND_{ij} \quad (7)$$

$$A_i^{\text{total cost}} = \sum_{j=1}^{n} w_j ND_{ij} \quad (8)$$

where $w_j$, represent the weight of criteria $j$, $m$ represents the count of beneficial criteria, and $n$ represents the count of cost criteria. In this paper, the values of $m$ and $n$ are two and four, respectively. The performance measures- Correlation Coefficient (r) and Pred (0.25) are beneficial criteria. Performance measures MBRE, NRMSE, MAE and SdARE are cost criteria as described in section III.*C*.

Step3: Use the following equation for calculating the WSM score.

$$A_i^{\text{wsm-score}} = A_i^{\text{total benefit}} - A_i^{\text{total cost}} \quad (9)$$

Step4: Rank the real estate price prediction regression models on the basis of the WSM score. The model with the highest score is considered the best.

#### D.2 TOPSIS

Step1: Calculate the weighted normalized decision matrix by multiplying all the values in each column of the normalized decision matrix $ND_{a\times c}$ (obtained from “(6)” as described in section III. D.1) by the corresponding weight of each criterion.

Step2: Find the positive ideal value and negative ideal value for each performance measure(criterion). In the case of negative criteria (also called cost criteria) minimum value will be the positive ideal value, and the maximum value will be the negative ideal value. In the case of positive criteria (also called beneficial criteria) maximum value will be the positive ideal value, and the minimum value will be the negative ideal value.

Step3: Calculate the Euclidean distance of each alternative from the positive ideal and from the negative ideal solution.

Step4: In this step, a score of performance is calculated for each real estate price prediction regression model (alternative). The score is expressed as a ratio of the distance of each alternative from the negative ideal solution to the difference of distance from the negative ideal solution and distance from the positive ideal solution.

Step5: Finally, rank each real estate price prediction regression model (alternative) according to its performance score, where the maximum score is ranked as the topmost.

#### D.3 EDAS

In this method, the best alternative is selected on the basis of distance from the average solution. Two measures, namely positive distance from average solution (*PDAVG*) and negative distance from average solution (*NDAVG*), play a key role in selecting the better alternative. An alternative with a higher value of *PDAVG* and a lower value of *NDAVG* is considered a superior solution than the average solution. The detailed procedure is described as follows:

Step1: Calculate the average solution (*AVG*).

[Description] The average solution for performance measure *j* can be calculated by using the following equation:

$$AVG_j = \frac{\sum_{i=1}^{a} D_{ij}}{a} \quad (10)$$

where $D_{ij}$ signify the value of performance measure *j* for the corresponding real estate price prediction regression model *i*, and *a* represents the number of real estate price prediction regression models. In this study the value of *a* is 22.

Step2: Calculate the positive distance from the average solution.

[Description] Positive distance from the average solution (obtained from "(2)") can be measured by using the following equations:

$$PDAVG_{ij} = \frac{\max(0, D_{ij} - AVG_j)}{AVG_j} \text{ for benificial criteria}\} \quad (11)$$

$$PDAVG_{ij} = \frac{\max(0, AVG_j - D_{ij})}{AVG_j} \text{ for cost criteria}\} \quad (12)$$

Here $PDAVG_{ij}$ represents the positive distance for the real estate price prediction regression model (alternative) *i* from the average solution for $j^{th}$ performance measure.

Step3: Calculate the weighted sum of the positive distance from the average solution by using the following equation:

$$WSP_i = \sum_{j=1}^{c} w_j * PDAVG_{ij} \quad (13)$$

Here $WSP_i$ represents the weighted sum of the positive distance from an average solution of $i^{th}$ alternative (real estate price prediction regression model), and *c* represents the number of performance measures. In this paper value of *c* is six and $w_{ij}$=1/6.

Step4: Calculate the negative distance from the *AVG*.

[Description] Negative distance from *AVG* (obtained from (10)) can be measured by using the following equations:

$$NDAVG_{ij} = \frac{\max(0, AVG_j - D_{ij})}{AVG_j} \text{ for benificial criteria}\} \quad (14)$$

$$NDAVG_{ij} = \frac{\max(0, D_{ij} - AVG_j)}{AVG_j} \text{ for cost criteria}\} \quad (15)$$

Here $NDAVG_{ij}$ represents the negative distance for the real estate price prediction regression model (alternative) *i* from the average solution for $j^{th}$ performance measure.

Step5: Calculate the weighted sum of the negative distance from the *AVG* as follows:

$$WSN_i = \sum_{j=1}^{c} w_j * NDAVG_{ij} \quad (16)$$

where $WSN_i$ represents the weighted sum of the negative distance from an average solution of $i^{th}$ alternative (real estate price prediction regression model), and *c* has the same meaning as in step 3.

Step6: Normalize the weighted sum of the positive and negative distance from the average solution using the following equations.

$$NWSP_i = \frac{WSP_i}{\max(WSP_i)} \quad (17)$$

$$NWSN_i = 1 - \frac{WSN_i}{\max(WSN_i)} \quad (18)$$

For the $i^{th}$ alternative (real estate price prediction regression model) $NWSP_i$, and $NWSN_i$ represents the normalized value of the weighted sum of positive distance and the weighted sum of the negative distance from the average solution, respectively.

Step7: calculate the evaluation factor ($EF_i$) for the $i^{th}$ alternative (real estate price prediction regression model) as follows:

$$EF_i = \frac{(NWSP_i + NWSN_i)}{2} \quad (19)$$

Step8: Rank the real estate price prediction regression models (alternatives) on the basis of the value of the evaluation factor, where the highest value gets the first rank.

## IV. Experimental Setup

This section is divided into two subsections IV. *A* and IV. *B*. A brief description of the datasets used in this study is presented in subsection IV.*A*. Subsection IV.*B* describes the procedure of experimental design of the proposed MCDM-based approach for the recommendation of the most suitable regression model for real estate price prediction modeling considering various competing evaluation metrics.

### A. Datasets

Three real estate valuation datasets, DS1, DS2, and DS3, are used in the experiment. DS1 and DS2 datasets are collected from Boston, USA, and Taipei City, Taiwan, respectively, and downloaded from the UCI machine learning repository [28]. Another real estate valuation dataset DS3 downloaded from Kaggle [29]. All three datasets are summarized in Table II.

TABLE II
Dataset Structures

| Datasets | Number of attributes | Number of instances |
|---|---|---|
| DS1 | 14 | 506 |
| DS2 | 7 | 414 |
| DS3 | 12 | 545 |

### B. Experimental Design

The following procedure is used for experimental design.

Step1: Develop twenty-two regression models for predicting real estate prices for each dataset. Open-source Weka version 3.8.3 was used for developing twenty-two real estate price prediction regression models.

Step2: Obtain the results of six performance measures as described in section III.*C* for twenty-two regression models used for predicting real estate prices. The result is a 22×6 matrix for each dataset.

Step3: To produce an index indicating the rank of regression models for predicting real estate prices for each dataset, utilize the 22×6 matrix for each dataset created in step 2 as input for the three MCDM methods- WSM, TOPSIS, and EDAS.

Step4: Finally, 22×1 matrix is obtained as the output of each MCDM method representing ranks of twenty-two real estate price prediction regression models for each dataset. Open-source package R version 4.0.2 was used for MCDM ranking.

The experimental study can be best understood using Fig. 2.

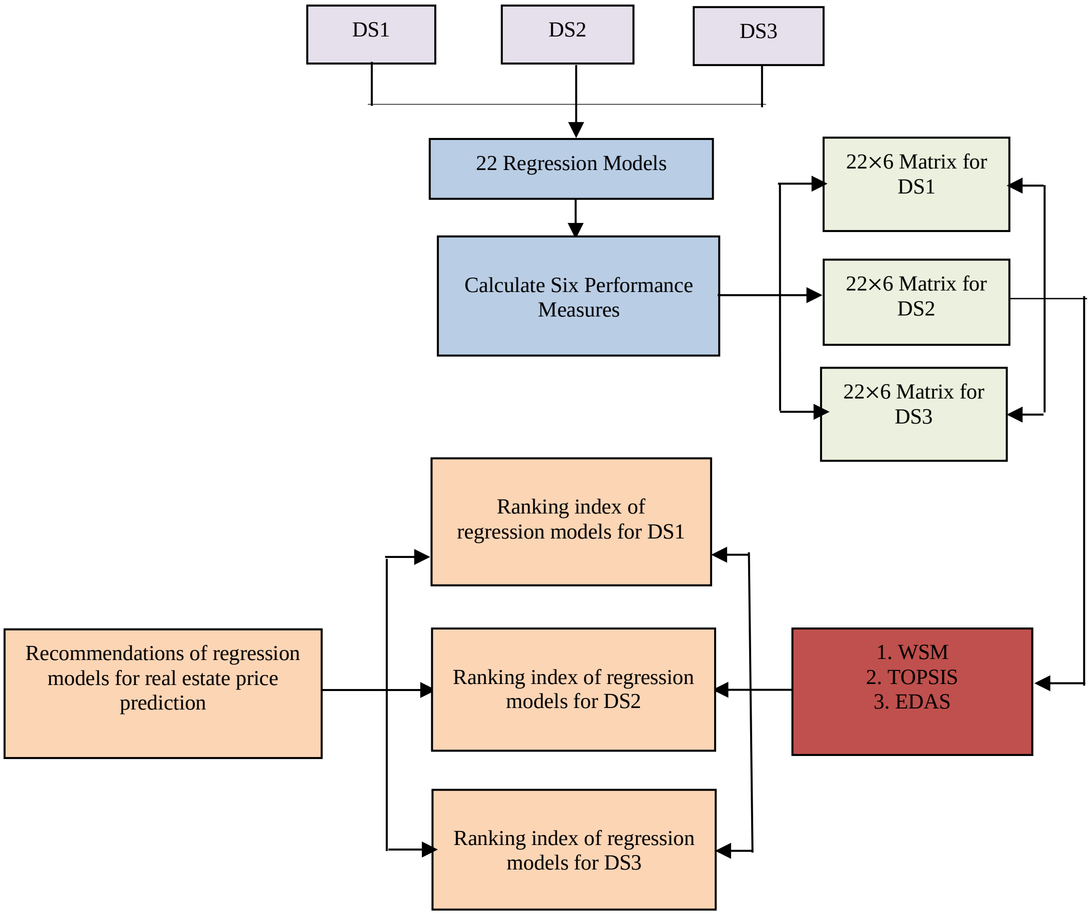


Fig. 2. Graphical representation of the experimental study

## V. Results and Discussion

For better understanding, this section is divided into two subsections V.*A* and V.*B*. Subsection V.*A* describe the real estate price prediction results in terms of six performance measures used in this study for three datasets, DS1, DS2, and DS3, by applying selected regression models. Subsection V.*B* presents the MCDM ranking of regression models.

### *A. Real estate Price Prediction Results*

This section presents the real estate price prediction results. Twenty-two machine learning techniques were trained over the three open-source datasets for real estate price prediction modeling. Each real estate price prediction regression model is measured by the six performance measures (MBRE, NRMSE, MAE, SdARE, correlation coefficient (r), Pred (0.25)) as described in section III.*C*. Real estate price prediction results of three datasets, DS1, DS2, and DS3, using selected regression models are listed in Table III-V.

- Table III shows that GRADBOOST performs best for MBRE, NRMSE, MAE, and Pred (0.25) for the DS1 dataset. RF performs best for performance measures SdARE. Regression model KSTAR is best for Correlation Coefficient (r).
- According to Table IV, the regression model RF performs the best for the performance measures MBRE, NRMSE, and correlation coefficient (r) for the DS2 dataset. Regression model KSTAR is best for MAE and Pred (0.25). Regression model ES is best for performance measures SdARE.
- According to Table V, SVR performs the best for the performance measures MBRE, MAE, Correlation Coefficient (r), and Pred (0.25). RIDGE is best for performance measure NRMSE. PR is best for performance measure SdARE.

TABLE III
RESULTS OF DATASET DS1

| Regression Model | MBRE | NRMSE | MAE | SdARE | r | Pred (0.25) |
|---|---|---|---|---|---|---|
| MLR | 0.2728 | 0.2176 | 3.6782 | 3.8090 | 0.8173 | 0.7648 |
| IR | 0.2079 | 0.2186 | 3.8086 | 3.7147 | 0.8159 | 0.7747 |
| MLP | 0.1888 | 0.1927 | 3.1660 | 3.4595 | 0.8624 | 0.8241 |
| RIDGE | 0.2864 | 0.2148 | 3.6509 | 3.7391 | 0.8225 | 0.7747 |
| EN | 0.4722 | 0.2293 | 3.9490 | 3.9425 | 0.7954 | 0.7470 |
| PR | 0.2728 | 0.2176 | 3.6782 | 3.8090 | 0.8173 | 0.7648 |
| SVR | 0.2400 | 0.2173 | 3.5463 | 3.9232 | 0.8222 | 0.7964 |
| KSTAR | 0.1617 | 0.1736 | 2.8112 | 3.1524 | ***0.9239*** | 0.8636 |
| LWL | 0.2796 | 0.2577 | 4.7310 | 4.1154 | 0.7375 | 0.6858 |
| AR | 0.1940 | 0.2003 | 3.3770 | 3.5160 | 0.8512 | 0.7964 |
| BAGGING | 0.1600 | 0.1770 | 2.8035 | 3.2699 | 0.8833 | 0.8656 |
| ES | 0.1690 | 0.1854 | 2.9895 | 3.3805 | 0.8714 | 0.8557 |
| REGBYDISC | 0.1830 | 0.1961 | 3.1736 | 3.5643 | 0.8582 | 0.8063 |
| CR | 0.3240 | 0.2915 | 5.4393 | 4.5513 | 0.6403 | 0.6126 |
| DT | 0.2085 | 0.2093 | 3.3700 | 3.8173 | 0.8332 | 0.8182 |
| M5Rules | 0.1734 | 0.1866 | 3.0061 | 3.4033 | 0.8711 | 0.8439 |
| M5P | 0.1679 | 0.1809 | 2.9145 | 3.2989 | 0.8779 | 0.8557 |
| RF | 0.1578 | 0.1713 | 2.7396 | ***2.8428*** | 0.8912 | 0.8518 |
| REPTREE | 0.1837 | 0.1984 | 3.2050 | 3.6108 | 0.8516 | 0.8142 |
| ADABOOST | 0.1860 | 0.1913 | 3.3470 | 3.2351 | 0.8659 | 0.7905 |
| GRADBOOST | ***0.1356*** | ***0.1604*** | ***2.4340*** | 3.2016 | 0.8929 | ***0.8786*** |
| KNN | 0.1734 | 0.1831 | 3.1266 | 3.1761 | 0.8758 | 0.8379 |

TABLE IV
RESULTS OF DATASET DS2

| Regression Model | MBRE | NRMSE | MAE | SdARE | r | Pred (0.25) |
|---|---|---|---|---|---|---|
| MLR | 0.2642 | 0.2306 | 6.5714 | 6.5838 | 0.7291 | 0.7512 |
| IR | 0.1950 | 0.2130 | 5.9365 | 6.2094 | 0.7750 | 0.7899 |
| MLP | 0.8622 | 0.3619 | 10.1031 | 10.5401 | 0.5051 | 0.6184 |
| RIDGE | 0.2618 | 0.2321 | 6.6008 | 6.6421 | 0.7248 | 0.7512 |
| EN | 0.2900 | 0.2324 | 6.5992 | 6.6562 | 0.7241 | 0.7536 |
| PR | 0.2642 | 0.2306 | 6.5714 | 6.5838 | 0.7291 | 0.7512 |
| SVR | 0.4232 | 0.2350 | 6.5514 | 6.8499 | 0.7264 | 0.7488 |
| KSTAR | 0.1636 | 0.1977 | ***4.6528*** | 6.2451 | 0.8107 | ***0.8551*** |
| LWL | 0.2312 | 0.2333 | 6.7819 | 6.5272 | 0.7223 | 0.7633 |
| AR | 0.1902 | 0.2083 | 5.7095 | 6.1662 | 0.7866 | 0.7947 |
| BAGGING | 0.1727 | 0.1976 | 5.2961 | 5.9587 | 0.8099 | 0.8357 |
| ES | 0.1811 | 0.2000 | 5.4881 | ***5.7137*** | 0.8048 | 0.8140 |
| REGBYDISC | 0.1857 | 0.2136 | 5.5356 | 6.6004 | 0.7740 | 0.8043 |
| CR | 0.2493 | 0.2457 | 7.2458 | 6.7595 | 0.6845 | 0.7343 |
| DT | 0.2314 | 0.2438 | 6.8467 | 7.0571 | 0.6927 | 0.7512 |
| M5Rules | 0.1823 | 0.2051 | 5.5380 | 6.1475 | 0.7933 | 0.8116 |
| M5P | 0.1798 | 0.2042 | 5.4733 | 6.1531 | 0.7955 | 0.8237 |
| RF | ***0.1530*** | ***0.1848*** | 4.9657 | 6.0896 | ***0.8480*** | 0.8333 |
| REPTREE | 0.1794 | 0.2062 | 5.5049 | 6.2381 | 0.7941 | 0.8213 |
| ADABOOST | 0.2227 | 0.2190 | 6.3410 | 6.1521 | 0.7705 | 0.7415 |
| GRADBOOST | 0.1720 | 0.2039 | 5.2710 | 6.3141 | 0.7992 | 0.8285 |
| KNN | 0.1851 | 0.2104 | 5.6803 | 6.3069 | 0.7849 | 0.8092 |

TABLE V
RESULTS OF DATASET DS3

| Regression Model | MBRE | NRMSE | MAE | SdARE | r | Pred (0.25) |
|---|---|---|---|---|---|---|
| MLR | 0.1987 | 0.2152 | 0.8041 | 0.7531 | 0.8078 | 0.7596 |
| IR | 0.2901 | 0.2979 | 1.1228 | 1.0320 | 0.5792 | 0.6018 |
| MLP | 0.1950 | 0.3404 | 1.0501 | 1.3907 | 0.6324 | 0.6991 |
| RIDGE | 0.1971 | ***0.1943*** | 0.7998 | 0.7508 | 0.8096 | 0.7725 |
| EN | 0.2952 | 0.3032 | 1.1466 | 1.0468 | 0.5568 | 0.5853 |
| PR | 0.1996 | 0.2151 | 0.8073 | ***0.6893*** | 0.8078 | 0.7725 |
| SVR | ***0.1832*** | 0.2161 | ***0.6878*** | 0.7771 | ***0.8502*** | ***0.8743*** |
| KSTAR | 0.2170 | 0.2452 | 0.8580 | 0.9168 | 0.7489 | 0.7284 |
| LWL | 0.2866 | 0.2937 | 1.1017 | 1.0231 | 0.5967 | 0.6257 |
| AR | 0.2334 | 0.2408 | 0.9263 | 0.8139 | 0.7533 | 0.7083 |
| BAGGING | 0.2080 | 0.2313 | 0.8434 | 0.8315 | 0.7738 | 0.7450 |
| ES | 0.2394 | 0.2549 | 0.9557 | 0.8885 | 0.7163 | 0.6972 |
| REGBYDISC | 0.2537 | 0.2808 | 1.0018 | 1.0312 | 0.6822 | 0.6881 |
| CR | 0.3024 | 0.3066 | 1.1580 | 1.0597 | 0.5446 | 0.6092 |
| DT | 0.2491 | 0.2654 | 0.9728 | 0.9491 | 0.6895 | 0.6899 |
| M5Rules | 0.1987 | 0.2152 | 0.8041 | 0.7531 | 0.8078 | 0.7596 |
| M5P | 0.1987 | 0.2152 | 0.8041 | 0.7531 | 0.8078 | 0.7596 |
| RF | 0.1962 | 0.2159 | 0.7888 | 0.7741 | 0.8066 | 0.7670 |
| REPTREE | 0.2397 | 0.2667 | 0.9726 | 0.9584 | 0.6978 | 0.6991 |
| ADABOOST | 0.2525 | 0.2559 | 0.9995 | 0.8472 | 0.7423 | 0.6440 |
| GRADBOOST | 0.1981 | 0.2240 | 0.8104 | 0.8118 | 0.7900 | 0.7615 |
| KNN | 0.2915 | 0.3003 | 1.1060 | 1.0680 | 0.5837 | 0.6239 |

According to the discussion above, no single real estate price prediction regression model can be deemed the best regression model when six performance measures are taken into account for each dataset. Therefore, it also encourages us to assess real estate price prediction regression models using the MCDM approach when there are many performance measures involved.

### *B. MCDM Ranking*

Tables VI-VIII show the ranking index for 22 real estate price prediction regression models for datasets DS1, DS2, and DS3, respectively.

TABLE VI
RANKING INDEX FOR DS1 DATASET

| Regression Model | Ranking Generated by MCDM Methods | | | | |
|---|---|---|---|---|---|
| | WSM Score | TOPSIS Score | EDAS Score | RPM Score | Final Rank |
| MLR | 0.7781 | 0.6162 | 0.3831 | 5.7736 | 17 |
| IR | 0.8071 | 0.7168 | 0.4624 | 5.0000 | 15 |
| MLP | 0.8886 | 0.8487 | 0.6849 | 3.1034 | 9 |
| RIDGE | 0.7818 | 0.5995 | 0.3826 | 6.0943 | 19 |
| EN | 0.7125 | 0.2867 | 0.1306 | 6.9894 | 21 |
| PR | 0.7781 | 0.6162 | 0.3831 | 5.7736 | 17 |
| SVR | 0.7982 | 0.6835 | 0.4283 | 5.3333 | 16 |
| KSTAR | 0.9727 | 0.9589 | 0.9415 | 1.0000 | 3 |
| LWL | 0.6887 | 0.4698 | 0.2015 | 6.7742 | 20 |
| AR | 0.8610 | 0.8075 | 0.5896 | 4.3333 | 13 |
| BAGGING | 0.9641 | 0.9546 | 0.9157 | 1.3333 | 4 |
| ES | 0.9293 | 0.9082 | 0.8146 | 2.0000 | 6 |
| REGBYDISC | 0.8813 | 0.8443 | 0.6595 | 3.5484 | 11 |
| CR | 0.6073 | 0.3169 | 0.0000 | 7.2188 | 22 |
| DT | 0.8353 | 0.7636 | 0.5206 | 4.6667 | 14 |
| M5Rules | 0.9204 | 0.8973 | 0.7876 | 2.5455 | 8 |
| M5P | 0.9430 | 0.9266 | 0.8562 | 1.6667 | 5 |
| RF | 0.9805 | 0.9720 | 0.9593 | 0.6667 | 2 |
| REPTREE | 0.8761 | 0.8362 | 0.6439 | 4.0000 | 12 |
| ADABOOST | 0.8889 | 0.8388 | 0.6841 | 3.3110 | 10 |
| GRADBOOST | ***0.9968*** | ***0.9910*** | ***1.0000*** | ***0.3333*** | ***1*** |
| KNN | 0.9284 | 0.8958 | 0.8101 | 2.4348 | 7 |

TABLE VII
RANKING INDEX FOR DS2 DATASET

| Regression Model | Ranking Generated by MCDM Methods | | | | |
|---|---|---|---|---|---|
| | WSM Score | TOPSIS Score | EDAS Score | RPM Score | Final Rank |
| MLR | 0.8141 | 0.8216 | 0.4759 | 5.2174 | 15 |
| IR | 0.9016 | 0.9176 | 0.7206 | 4.0000 | 12 |
| MLP | 0.5200 | 0.0000 | 0.0000 | 7.3333 | 22 |
| RIDGE | 0.8114 | 0.8211 | 0.4743 | 5.8846 | 18 |
| EN | 0.8012 | 0.7951 | 0.4625 | 6.3216 | 19 |
| PR | 0.8141 | 0.8216 | 0.4759 | 5.2174 | 15 |
| SVR | 0.7664 | 0.6518 | 0.4037 | 7.0000 | 21 |
| KSTAR | 0.9866 | 0.9815 | 0.9889 | 0.6667 | 2 |
| LWL | 0.8255 | 0.8418 | 0.5204 | 4.6667 | 14 |
| AR | 0.9184 | 0.9332 | 0.7794 | 3.3333 | 10 |
| BAGGING | 0.9710 | 0.9751 | 0.9479 | 1.0000 | 3 |
| ES | 0.9523 | 0.9573 | 0.8899 | 1.6667 | 5 |
| REGBYDISC | 0.9114 | 0.9300 | 0.7545 | 3.6667 | 11 |
| CR | 0.7836 | 0.7984 | 0.4530 | 6.5517 | 20 |
| DT | 0.7985 | 0.8217 | 0.4967 | 5.3774 | 17 |
| M5Rules | 0.9368 | 0.9495 | 0.8397 | 2.6667 | 8 |
| M5P | 0.9440 | 0.9558 | 0.8635 | 2.0000 | 6 |
| RF | ***0.9908*** | ***0.9865*** | ***1.0000*** | ***0.3333*** | ***1*** |
| REPTREE | 0.9389 | 0.9518 | 0.8461 | 2.3333 | 7 |
| ADABOOST | 0.8623 | 0.8745 | 0.6094 | 4.3333 | 13 |
| GRADBOOST | 0.9544 | 0.9639 | 0.8926 | 1.3333 | 4 |
| KNN | 0.9205 | 0.9364 | 0.7856 | 3.0000 | 9 |

TABLE VIII
RANKING INDEX FOR DS3 DATASET

| Regression Model | Ranking Generated by MCDM Methods | | | | |
|---|---|---|---|---|---|
| | WSM Score | TOPSIS Score | EDAS Score | RPM Score | Final Rank |
| MLR | 0.9868 | 0.9685 | 0.9777 | 1.6667 | 5 |
| IR | 0.7175 | 0.3475 | 0.0758 | 6.5517 | 20 |
| MLP | 0.7655 | 0.3532 | 0.1597 | 5.7736 | 17 |
| RIDGE | ***0.9934*** | ***0.9818*** | ***1.0000*** | ***0.3333*** | ***1*** |
| EN | 0.7012 | 0.3172 | 0.0210 | 7.0000 | 21 |
| PR | 0.9891 | 0.9703 | 0.9868 | 1.0909 | 3 |
| SVR | 0.9926 | 0.9730 | 0.9980 | 0.6667 | 2 |
| KSTAR | 0.8941 | 0.7607 | 0.6557 | 3.3333 | 10 |
| LWL | 0.7326 | 0.3746 | 0.1238 | 5.8846 | 18 |
| AR | 0.8889 | 0.7589 | 0.6337 | 3.6667 | 11 |
| BAGGING | 0.9346 | 0.8642 | 0.8047 | 3.0000 | 9 |
| ES | 0.8500 | 0.6707 | 0.4926 | 4.0000 | 12 |
| REGBYDISC | 0.7948 | 0.5038 | 0.3193 | 5.3333 | 16 |
| CR | 0.6974 | 0.3006 | 0.0000 | 7.3333 | 22 |
| DT | 0.8207 | 0.5911 | 0.4032 | 5.0000 | 15 |
| M5Rules | 0.9868 | 0.9685 | 0.9777 | 1.6667 | 5 |
| M5P | 0.9868 | 0.9685 | 0.9777 | 1.6667 | 5 |
| RF | 0.9884 | 0.9717 | 0.9836 | 1.2000 | 4 |
| REPTREE | 0.8275 | 0.6050 | 0.4228 | 4.6667 | 14 |
| ADABOOST | 0.8372 | 0.6398 | 0.4677 | 4.3333 | 13 |
| GRADBOOST | 0.9642 | 0.9210 | 0.9051 | 2.6667 | 8 |
| KNN | 0.7194 | 0.3283 | 0.0759 | 6.4407 | 19 |

Three MCDM methods, WSM, TOPSIS, and EDAS (as described in section III.*D*), were applied to the 22 × 6 decision matrix (real estate price prediction results of 22 regression models for six evaluation metrics obtained in section V.*A*) to calculate the WSM score, TOPSIS score, and EDAS score (as described in detail in section III.*D*). Next, the rank of each regression model is determined for each MCDM method. According to the WSM, TOPSIS, and EDAS procedures, the higher the score of the alternative (real estate price prediction model) higher will be the rank. The individual ranks of the regression models obtained by three MCDM methods are then combined to get the final rank list using the rank position method. The rank position method (RPM) (also called the reciprocal rank method) [30] takes into account the position of each alternative according to each subordinate ranking technique. In this study, alternatives are the regression models for real estate price prediction modeling, and subordinate ranking techniques are three MCDM methods. The rank position method is based on the RPM score (regression model with a lower RPM score will be assigned a higher rank). RPM score for each regression model employed to generate the final ranking can be calculated using "(20)".

$$\text{RPM Score} = \frac{1}{(1/\text{rank(WSM)} + 1/\text{rank(TOPSIS)} + 1/\text{rank(EDAS)})} \quad (20)$$

where rank (WSM), rank (TOPSIS), and rank (EDAS) are the rankings of regression models produced by WSM, TOPSIS, and EDAS respectively.

The following inferences are drawn from Table VI–VIII based on the application of three MCDM techniques, namely WSM, TOPSIS, and EDAS.

- From Table VI, it can be observed that the regression model GRADBOOST is placed at rank one. Hence, upon optimization of all competing performance measures, it is found that the regression model Gradient Boosting (GRADBOOST) is recommended as the most acceptable regression model for real estate price prediction modeling for the DS1 dataset.
- From Table VII, it can be observed that the regression model RF is placed at rank one. Thus, on the optimization of all competing performance metrics, it is also seen that the regression model Random Forest (RF) is suggested as the most appropriate regression model for real estate price prediction modeling for the DS2 dataset.
- From Table VIII, it can be observed that the regression model RIDGE is placed at rank one. As a result of the optimization of all competing performance metrics, it is further noted that the regression model Ridge Regression (RIDGE) is suggested as the most appropriate regression model for real estate price prediction modeling for the DS3 dataset.

These regression models are recommended due to the fact that all three MCDM methods—WSM, TOPSIS, and EDAS ranked them first. Table IX provides a concise summary of these recommendations.

TABLE IX
BEST REGRESSION MODEL FOR REAL ESTATE PRICE PREDICTION

| Best regression model based on individual performance measures for three datasets | | | | | | | *Best regression model based on optimization of all six performance measures* |
|---|---|---|---|---|---|---|---|
| Dataset | Performance Measures | | | | | | |
| | MBRE | NRMSE | MAE | SdARE | r | Pred (0.25) | |
| DS1 | GRADBOOST | GRADBOOST | GRADBOOST | RF | KSTAR | GRADBOOST | *GRADBOOST* |
| DS2 | RF | RF | KSTAR | ES | RF | KSTAR | *RF* |
| DS3 | SVR | RIDGE | SVR | PR | SVR | SVR | *RIDGE* |

### *C. Theoretical and Practical Implications*

Undervaluation and overvaluation in the real estate market have always been an issue. Hence there is always a need for an accurate model for predicting real estate prices. The proposed method in this study can benefit real estate sellers, buyers, and economic experts. Various regression models are available for real estate price prediction. Different regression models produce inconsistent findings for various performance indicators, making selecting the appropriate regression model problematic. This study provides a unique framework for evaluating different regression models for real estate price prediction using MCDM by taking into account various conflicting performance measures altogether. This research will also help future research scholars to solve various decision-making problems using MCDM in other domains, such as in the healthcare industry for selecting the most suitable diabetes prediction model, and in software industries for selecting the most appropriate software testing techniques, etc.

## VI. CONCLUSION AND FUTURE WORK

The decision to purchase real estate is definitely essential in most individuals' lives. As a result, real estate price prediction can provide helpful information to aid with real estate transactions. Various regression models are available for real estate price prediction. Different regression models produce inconsistent findings for various performance indicators, making selecting the appropriate regression model problematic. To address this issue, we present a novel strategy for evaluating regression models based on MCDM methods in the domain of real estate price prediction. In this method, we first calculate the values of all the performance measures for twenty-two applied regression models for each dataset. The findings of six performance metrics for the applied regression models, as stated in section V.*A*, reveal that no model can be proclaimed the best based on all performance indicators for any dataset. As a result, the regression models must be evaluated based on the optimization of all six performance measures.

Next, using three MCDM techniques, an index is created to rank regression models based on six performance indicators. For validation of the proposed approach, three real estate price datasets were used for analysis. Experimental results show that the regression model Gradient Boosting (GRADBOOST) is best suitable for the DS1 dataset, regression model Random Forest (RF) is best suitable for the DS2 dataset, and regression model Ridge Regression (RIDGE) is best suitable for DS3 dataset as they are ranked at first position by all three MCDM method WSM, TOPSIS, and EDAS. As a limitation, the proposed approach uses a smaller number of datasets. However, applying the proposed approach to a large number of datasets may be the direction for future work to generalize the results. The proposed method in this study may be extended to evaluate various machine learning models in other domains such as in the healthcare industry for selecting the most suitable diabetes prediction model, and in software industries for selecting the most appropriate software testing techniques, etc.

## REFERENCES

[1] J. Kalliola, J. Kapociute-Dzikiene, and R. Damasevicius, "Neural network hyperparameter optimization for prediction of real estate prices in Helsinki," *PeerJ Computer Science*, vol. 7, 2021. https://doi.org/10.7717/peerj-cs.444.

[2] D.H. Wolepart and W.G. Macready, "No Free Lunch Theorems for Optimization," IEEE Transactions on Evolutionary Computation, vol. 1, no. 1, pp. 67-82, 1997.

[3] CH. R. Madhuri, G. Anuradha, and M. V. Pujitha, “House price prediction using regression techniques: A comparative study,” in *International Conference on Smart Structures and Systems (ICSSS)*, 2019, pp. 1-5. https://doi.org/10.1109/ICSSS.2019.8882834.

[4] P.-F. Pai and W.-C. Wang, “Using machine learning models and actual transaction data for predicting real estate prices,” *Applied Sciences*, vol. 10, no. 17, p. 5832, 2020. https://doi.org/10.3390/app10175832.

[5] M. Thamarai, Sri Vasavi Engineering College, Andhra Pradesh, India, and S. P. Malarvizhi, “House price prediction modeling using machine learning,” *International Journal of Information Engineering & Electronic Business*, vol. 12, no. 2, pp. 15–20, 2020. https://doi.org/10.5815/ijieeb.2020.02.03.

[6] Z. Wu, “Prediction of California House Price Based on Multiple Linear Regression,” *Academic Journal of Engineering and Technology Science*, vol. 3, no. 7, pp. 11-15, 2020. https://doi.org/10.25236/AJETS.2020.030702.

[7] A. Singh, A. Sharma, and G. Dubey, “Big data analytics predicting real estate prices,” *International Journal of System Assurance Engineering and Management*, vol. 11, no. 2, pp. 208–219, 2020. https://doi.org/10.1007/s13198-020-00946-3.

[8] M. Erkek, K. Çayırlı, and A. Hepsen, “Predicting House Prices in Turkey by Using Machine Learning Algorithms,” *Journal of Statistical and Econometric Methods*, vol. 9, no. 4, pp. 31–38, 2020.

[9] N. H. Zulkifley, S. A. Rahman, N. H. Ubaidullah, and I. Ibrahim, “House Price Prediction using a Machine Learning Model: A Survey of Literature,” *International Journal of Modern Education & Computer Science*, vol. 12, no. 6, pp. 46-54, 2020. https://doi.org/10.5815/ijmecs.2020.06.04.

[10] L. Mrsic, H. Jerkovic, and M. Balkovic, “Real estate market price prediction framework based on public data sources with case study from Croatia,” in *Communications in Computer and Information Science*, Singapore: Springer Singapore, 2020, pp. 13–24. https://doi.org/10.1007/978-981-15-3380-8_2.

[11] J. Avanijaa, “Prediction of house price using xgboost regression algorithm,” *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 2, pp. 2151-2155, 2021.

[12] P.-Y. Wang, C.-T. Chen, J.-W. Su, T.-Y. Wang, and S.-H. Huang, “Deep learning model for house price prediction using heterogeneous data analysis along with joint self-attention mechanism,” *IEEE Access*, vol. 9, pp. 55244–55259, 2021. https://doi.org/ 10.1109/ACCESS.2021.3071306

[13] T. Zhang and B. Fang, “Prediction of the real estate industry economics based on LSTM model,” *Journal of Physics: Conference Series*, vol. 1802, no. 4, p. 042095, 2021. https://doi.org/ 10.1088/1742-6596/1802/4/042095.

[14] C. Zhou, “House price prediction using polynomial regression with Particle Swarm Optimization,” *Journal of Physics: Conference Series*, vol. 1802, no. 3, p. 032034, 2021. https://doi.org/ 10.1088/1742-6596/1802/3/032034.

[15] S. Dabreo, S. Rodrigues, V. Rodrigues, and P. Shah, “Real Estate Price Prediction,” *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 4, pp. 644–649, 2021.

[16] A. Deaconu, A. Buiga, and H. Tothăzan, “Real estate valuation models performance in price prediction,” *International Journal of Strategic Property Management*, vol. 26, no. 2, pp. 86–105, 2022. https://doi.org/10.3846/ijspm.2022.15962.

[17] A. Louati, R. Lahyani, A. Aldaej, A. Aldumaykhi, and S. Otai, “Price forecasting for real estate using machine learning: A case study on Riyadh city,” *Concurrency and Computation*, vol. 34, no. 6, 2022. https://doi.org/10.1002/cpe.6748.

[18] J. Zaki, A. Nayyar, S. Dalal, and Z. H. Ali, “House price prediction using hedonic pricing model and machine learning techniques,” Concurrency and Computation, vol. 34, no. 27, 2022. https://doi.org/10.1002/cpe.7342

[19] S. O. Akyuz, B. E. Erdogan, and O. Yıldız, “A Novel Hybrid House Price Prediction Model”, Computational Economics, 2022. https://doi.org/10.1007/s10614-022-10298-8

[20] S. Sharma, D. Arora, G. Shankar, P. Sharma, V. Motwani, “House Price Prediction using Machine Learning Algorithm”, Proceeding of the 7th International Conference on Computing Methodologies and Communication (ICCMC), IEEE, 2023. https://doi.org/10.1109/ICCMC56507.2023.10084197

[21] P. S. M. Reddy, “Decision Tree Regressor Compared with Random Forest Regressor for House Price Prediction in Mumbai”, Journal of Survey in Fisheries Sciences, 10(1S):2323-2332, 2023.

[22] E. Ballestero, B. P. Gladish, A. G. Bernabeu, (Eds.) “Socially Reponsible Investment. A Multi-Criteria Decision-Making Approach,” International Series in Operations Research and Management Science. Springer, Cham, 2015.

[23] T. Menzies, E. Kocaguneli, B. Turhan, L. Minku, and F. Peters: *Sharing Data and Models in Software Engineering,* Oxford, England, Morgan Kaufmann, 2014.

[24] D. J. Sheskin, “The pearson product-moment correlation coefficient,” in Handbook of parametric and nonparametric statistical procedures, 5th ed., London, New York: Chapman and Hall/CRC, 2020.

[25] P. C. Fishburn, “Additive Utilities with Incomplete Product Sets: Application to Priorities and Assignments,” *Operations Research*, vol. 15, no. 3, pp. 537–542, 1967.

[26] C. L. Hwang and K. Yoon, “Methods for Multiple Attribute Decision Making,” in *Multiple Attribute Decision Making*, Berlin Heidelberg: Springer, pp. 58–191, 1981.

[27] M. Keshavarz Ghorabaee, E. K. Zavadskas, L. Olfat, and Z. Turskis, “Multi-criteria inventory classification using a new method of evaluation based on distance from average solution (EDAS),” *Informatica (Vilnius)*, vol. 26, no. 3, pp. 435–451, 2015. http://dx.doi.org/10.15388/Informatica.2015.57.

[28] D. Dua and C. Graff, “UCI Machine Learning Repository [http://archive.ics.uci.edu/ml],” Irvine, CA: The University of California, School of Information and Computer Science, 2019.

[29] M. Yasser, “Housing Prices Dataset”, https://www.kaggle.com/yasserh/housing-prices-dataset.

[30] S. Altuntas, T. Dereli, and M.K. Yilmaz, "Evaluation of Excavator Technologies: Application of Data Fusion Based MULTIMOORA Methods," Journal of Civil Engineering and Management, vol. 21, no. 8, pp. 977–997, 2015.


**Ajay Kumar** is assistant professor with Department of Information Technology, KIET Group of Institutions, Delhi-NCR, Ghaziabad, India. He received his master degree in CSE from National Institute of Technical Teachers Training and Research, Chandigarh. He has done his Bachelor of Engineering degree in CSE from Dr. B. R. A University Agra, Uttar Pradesh, India. His research fields include software quality, artificial intelligence, soft computing, and multi-criteria decision-making.