Forecasting Real Estate Prices: Regression Models and Accuracy Assessment

Abstract : Despite the financial markets' well-known volatility, machine learning techniques for predicting housing prices have gained significant adoption in recent years by identifying intricate patterns within the real estate market. This is because these algorithms can handle large volumes of data and detect complex patterns that are difficult for humans to identify. To forecast house prices, the proposed approach focuses on using various machine learning models such as linear regression, support vector regression, ridge regression, lasso regression, and random forest. This application of machine learning provides a promising opportunity for stakeholders such as homeowners, housing agents, brokers, and investors to make well-informed decisions. The analysis compares the models' performance in predicting house prices using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R2, and includes graph visualization. In the proposed model, support vector regression and random forest outperform other models.

Keywords : Housing Price, Linear Regression, Random Forest, Support Vector Regression (SVR), Prediction, performance metrics

1. Introduction : Housing plays a crucial role in determining the success of a country's economy. As the economy grows, people tend to move from rural to urban areas which leads to an increase in the urban population. This surge in urban population causes a corresponding rise in demand for accommodation, which in turn increases the price of housing. Moreover, the development of infrastructure in a particular area can cause a sudden increase in the price of houses in that area. According to research on the impact of the Bangalore Metro, there was a 5-20% increase in residential property values after the announcement of the metro project, and a 10-30% increase after the completion of construction. This was due to the enhanced accessibility, connectivity, and overall attractiveness of the areas served by the new metro. [1]. Soaring house prices have a significant impact on household wealth as housing is a major asset for most families. As a result, this wealth effect can greatly influence consumer spending and economic activity. [2] House prices depend on various factors, including the size, age, and condition of the property. Generally, bigger, newer, and well-maintained homes tend to cost more than smaller, older, or poorly maintained ones. Moreover, the price of a home can also be impacted by the location and characteristics of the neighborhood. Properties in high-end areas with better facilities, such as schools, amenities, and low crime rates, usually have higher prices. [3]. Physical aspects that can be easily observed by most people when buying a house include the size of the property, number of rooms, the size of the kitchen and garage, the availability of a garden, the area of land and structures, and the age of the property. The physical attributes of a house, such as its size, year of construction, number of bedrooms and bathrooms, and other interior features, may also have an impact on its price. [4]. Developers often use marketing strategies to attract investors by emphasizing factors like proximity to hospitals, markets, schools, airports, and main roads. The location of a property is a significant factor in determining its price, as the value of land is dependent on the area it is situated in. Therefore, understanding house price trends and the factors that influence them is crucial not only for tenants, but also for homeowners, industry analysts, policymakers in the real estate sector, and urban and regional planning authorities. [5].

In the real estate market, the use of computer-based prediction systems, such as machine learning, has become increasingly popular due to its ability to improve decision-making, reduce risk, and increase efficiency in property valuation, management, and investment . machine learning algorithms can automatically categorize

properties based on various features, such as number of bedrooms, number of bathrooms, and square footage, among others. This categorization helps to rank search results and suggest comparable properties, making it easier for buyers and sellers to find what they're looking for. [6] Machine learning algorithms can analyze historical sales data and other relevant factors such as demographics, location, size, and amenities to accurately predict the value of a property. machine learning has the potential to simplify real estate transactions by providing valuable insights and data-driven recommendations that can help real estate professionals and individuals make more informed decisions. [7]. Machine learning algorithms can predict properties that have the potential to appreciate in value or yield high rental income by analyzing historical data and recent market patterns. Furthermore, they can evaluate market trends, property data, and economic indicators to determine investment risks associated with a specific property or market. Machine learning algorithms can identify potential fraud, such as mortgage fraud, by analyzing data. Energy consumption data from buildings can be analyzed by machine learning algorithms to identify patterns and optimize energy use, thus reducing costs.. Fourth, By analyzing data on occupancy rates, rental rates, and tenant behavior, machine learning algorithms can optimize property management operations, such as lease renewal, rent collection, and maintenance scheduling.. Fifth, Machine learning algorithms can be used by real estate websites and apps to suggest properties to consumers based on their interests, search histories, and activity [8]

This study focuses on the real estate market to demonstrate how machine learning can surpass traditional statistical methods in predicting prices using the Housing Dataset [9]. The paper follows this structure: In Section 2, it reviews five machine learning algorithms (Linear Regression, Lasso Regression, Ridge Regression, Support Vector Regression, and Random Forest). Then, it delves deeper into these methods, explaining their inner workings, formulas, how they learn. Next, it describes the data used in the study, explaining the meaning of each piece of data and its source. The subsequent section presents the results obtained using the three most promising algorithms and compares their performance. The Kaggle housing dataset [9] was employed to evaluate the performance of the proposed model, with Mean Absolute Error (MAE), R2 (Coefficient of Determination), and Root Mean Square Error (RMSE) serving as performance evaluation metrics. Finally, all performance metrics were depicted on a bar graph to determine which method appeared to be the best fit for the Housing Dataset. The paper concludes by summarizing the findings.

---

2. Literature Review : Real estate appraisal is a matter of great importance to numerous stakeholders worldwide within this industry. There is a diverse range of real estate assessments that serve multiple objectives. [10]. Due to inflationary challenges caused by the COVID-19 pandemic, Turkey is currently experiencing an increased interest in the real estate industry, which is in line with global trends. This interest is driven by a desire to protect existing investments and take advantage of profit opportunities. However, analyzing the real estate industry is a complex task that involves multiple factors, and requires a comprehensive evaluation of fair market values using multivariate approaches..[11], Research related to real estate appraisal is on the rise due to the growing demand. Advancements in this field, such as the creation of new models, and the enhancement of data pre-processing procedures, are contributing to the improvement of real estate appraisal accuracy..

In their research on the Spanish real estate market, Authors[12] conducted tests on different machine learning algorithms to compare their advantages and limitations using various methods. The study findings indicated that ensemble models comprising of regression trees consistently performed better than other models. The development of an intelligent system that predicts real estate prices using data from real estate sales was done by Author [13]. The study involved 14 different methods and 176 features were used for price prediction. The Random Forest algorithm gave the most accurate price predictions for 852 houses in Istanbul's Ataşehir district.. Reference [15] employed the Random Forest (RF) algorithm to make predictions on the house prices in London. Despite the small dataset size, the study demonstrated that RF outperformed the traditional regression approach based on Generalized Linear Models (GLMs) in terms of prediction improvement. According to the study's findings, RF was better at capturing complex relationships and patterns in the data than GLM. This could explain why RF performed better in this study. As part of their work, researchers[14] designed a system that

leverages machine learning algorithms such as Linear Regression, Random Forest, and Boosted Regression to accurately predict real estate prices in Mumbai. The primary objective of this system is to mitigate investment risks for users by providing accurate output and preventing them from investing in the wrong real estate.

Study	Method	Dataset	Strengths	Weaknesses	Main Contribution	Results
Kang, Y., Zhang [4]	Systematic Review	Open Source	Focused on RF Regressor	Less specific dataset focus	Highlighting effective ML algorithms	Random Forest
Abigail Bola Adetunjia [5]	Single ML Algorithm.	Boston Housing Dataset	High efficiency of Random Forest	Linear Regression less efficient	k-fold cross-validation	High Accuracy in Random Forest
Lennon [6] (2023)	Multiple ML algorithms	Hong Kong Dataset	Performance metrics Evaluation	Lower accuracies in R-Squared metrics	RF better than KNN, decision tree	Accuracies above 95%
Baldominos,. [7] (2014)	Multiple ML algorithms	Madrid House Dataset, Spain	High accuracy in KNN and Neural network	Limited comparison with other methods	Improved Regression/ Neural Network method	KNN, SVR performed differently in metrics
Zhihang Liu [10]	Multiple methods	Dataset with ten features	Comprehensive methodology	No specific weaknesses noted	Detailed prediction on Lasso, ridge , svm	Lasso regression with accuracy 90.15%
Jiao Yang Wu [17] (2023)	Support Vector regression	King County, USA.	Superior performance of SVR	Other techniques less effective	Feature extraction with SVR	SVR Performed with high accuracy

3. Methodology : The methodology section detailed the process from initial Data collection ,Feature engineering, model selection, training, and evaluation, including data cleaning, pre-processing, and performance metric assessment, as depicted in Figure 1(System Architecture).

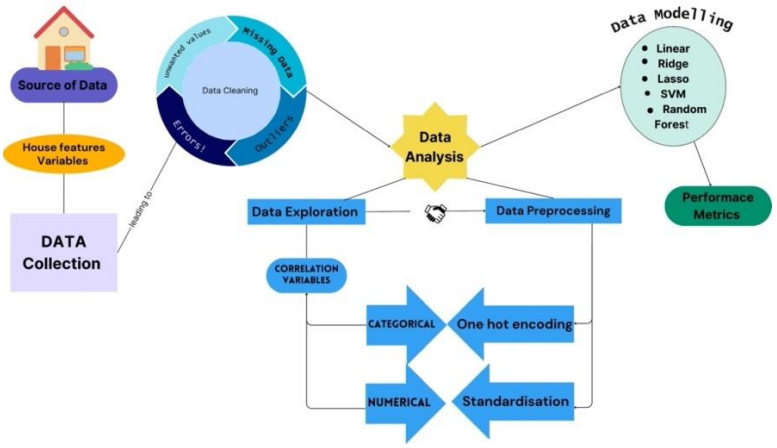

Figure : 1(System Architecture)

3.1. Data Collection :

The dataset used for predicting house prices was obtained from Kaggle and contains 80 attributes and 1460 data samples. The dependent variable used for prediction is "Sales price" and it is derived from 80 feature variables such as Lot area, street, bedrooms, year built, total area, and more. Prior to building a regression model,

exploratory data analysis is a necessary step. This process helps researchers discover the underlying trends in the data, which assists in selecting appropriate machine learning techniques. Therefore, data exploration was conducted to understand the characteristics of the dataset and their significance. Some of the key features in the dataset include LotArea, which is the lot size in square feet, OverallQual, which represents the overall material and finish quality, YearBuilt, which is primarily the year in which the house was built, GarageArea, which is the size of the garage in square feet, and more. Since the model uses a supervised learning method, the dataset must be divided into the training dataset and the testing dataset. For the training dataset, 80% of the dataset was used to train the model while the remaining 20% was used for testing..

3.2 Exploratory Data Analysis

The process of Exploratory Data Analysis (EDA) involved a careful examination of the dataset to uncover trends and patterns that affect real estate sales prices. Visualizations were used prior to using a correlation matrix to assess the relationship between dependent and independent variables and to identify trends. Joint plots were also utilized to discover additional patterns. Additionally, the distribution of sales price status was analyzed to explore variations and understand the impact of different variables, as illustrated in Figure 2.

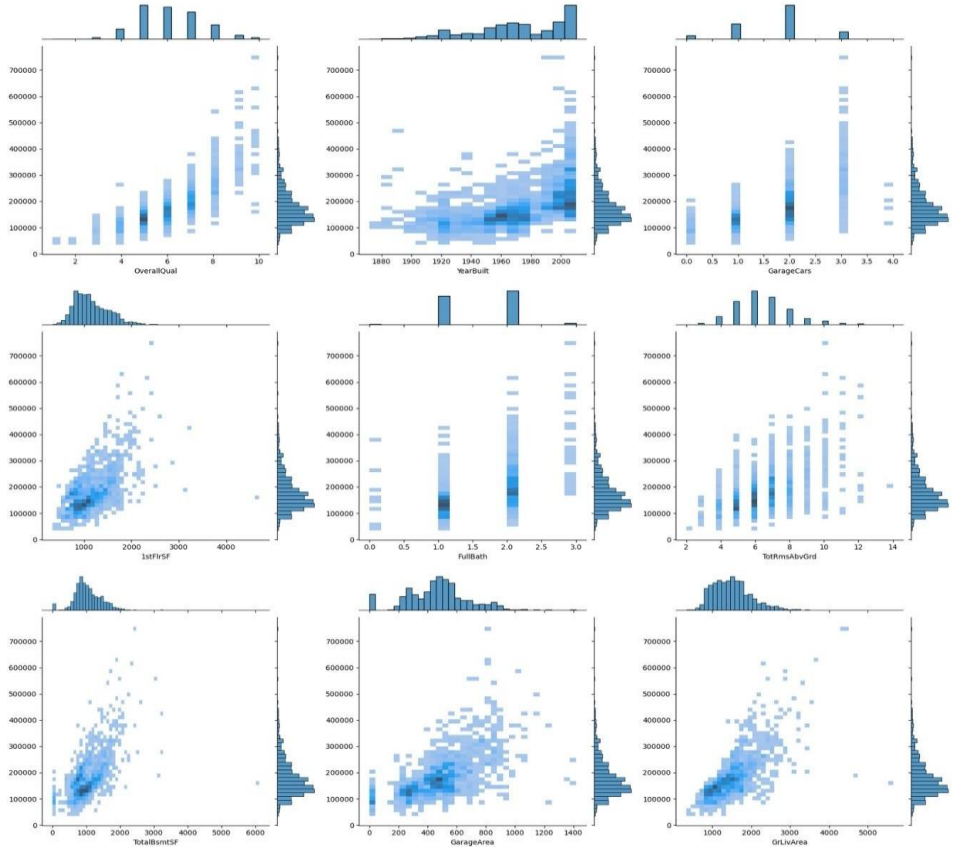

Figure : 2 (Parameters distribution)

3.3. Feature selection /Extraction

Even though our dataset has a lot of variables (80 in this case), not all of them are helpful for building our machine learning model. Some might even hurt its ability to make predictions. Research has shown that simply removing features can be risky, so choosing the right ones is crucial. To do this, we can first create a correlation matrix. This helps us see how closely related each variable is to the others. If two features are highly correlated then one of them might be redundant and can be removed. Figure 3 shows this correlation matrix for our specific dataset

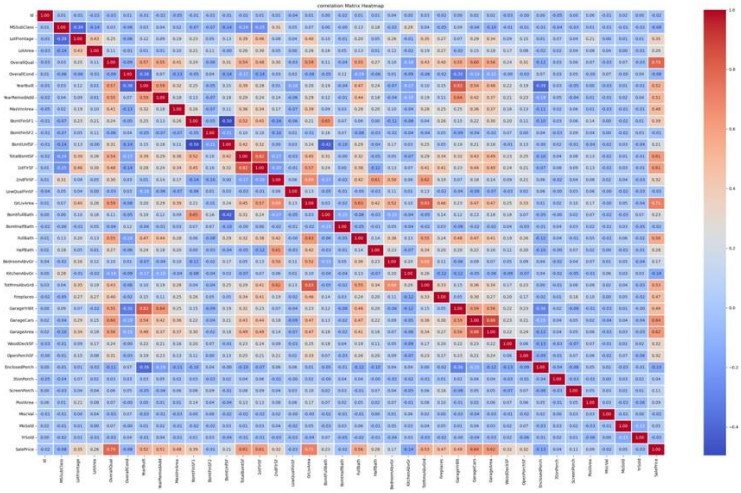

Figure : 3 (HeatMap)

(I have chosen numerical variables that show a correlation rate of over 0.50 or below -0.50, using the Pearson Correlation Method which is set as the default value of the "method" parameter in the corr() function. When it comes to selecting categorical variables, I have picked those that I believe have a significant impact on the target variable. As can be seen in Figure : 4

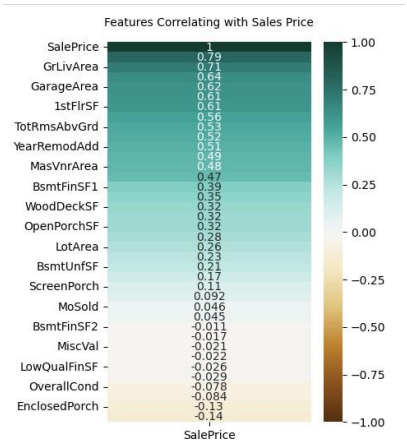

Figure : 4

(*We can see in figure 3 & 4 that the most 6 relevantparameters are: households and Overallquay =0.79 GrLivArea=0.71, GarageCar = 0.64,GarageArea = 0.62 and TatalBsmt=0.61, 1stFlrSF = 0.61, FullBath = 0.56)*

3.4. Data Transformation :

Before creating models, it is important to analyze the acquired data for model training and testing appropriately. This helps the models to learn patterns more quickly and efficiently. Standardization was applied to numerical values, while categorical values were encoded one-at-a-time. Once the data exploration is done and the most suitable feature is selected using the heatmap, the next stage is the pre-processing of the data of the selected features. Typically, the datasets used for training and testing have multiple features. It is highly probable that various features may have different scales, which can lower the model's performance. Therefore, scaling was carried out to ensure that the features are on a relatively similar scale. For this task, the Standard Scaler function available in the Python Skitlearn module was used. The Standard Scaler function assumes that the data is naturally distributed within each function, scales it so that it is now clustered about 0 with a standard deviation

of 1. The feature's mean and standard deviation are measured and then the feature is scaled based on this information.

3.5 Machine Learning Algorithms

3.5.1. Linear Regression

Linear Regression is a type of machine learning algorithm that establishes a linear connection between independent variables (X) and dependent variables (Y). When a model is evaluated, the difference between the actual value of the target (Y) and its predicted value is known as residuals..

The Linear regression is given by Equation : $Y = mX + C$

(Here, X is the explanatory variable/ Independent Variable and Y is the dependent variable, M is the slope of the line and c is the Y-intercept)

3.5.2. Ridge Regression

Ridge regression is a regularization technique to mitigate multicollinearity and overfitting. RG tends to shrink the coefficients towards zero but doesn't usually set them exactly to zero. This penalty term is proportional to the square of the magnitude of the coefficients, hence encouraging smaller coefficient values. The strength of regularization is controlled by a hyperparameter λ (lambda), which is usually determined via cross-validation. It help improve the model's generalization performance, especially in situations where there are many correlated features. In our,Housing Dataset its useful when the independent variables (features) are highly correlated, which can lead to unstable estimates of the regression coefficients

The objective function of ridge regression can be written as:

$$\text{minimize} \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

Where:

- y_i is the observed value for the ith observation.
- $\beta_0, \beta_1, ..., \beta_p$ are the regression coefficients to be estimated.
- x_{ij} is the value of the jth predictor variable for the ith observation.
- λ is the regularization parameter.

3.5.3. Lasso Regression :

Lasso Regression is a modified version of linear regression, where the model is penalized for the sum of absolute values of the weights. Lasso penalizes the sum of absolute values of coefficients. As the λ (lambda)value increases, coefficients decrease and eventually become zero. This way, lasso regression eliminates insignificant variables from our model. In our Housing Data, the Lasso regression will eliminate the variable which are least important to the predict target variable.

The objective function of lasso regression can be written as:

$$\text{minimize} \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Where:

- y_i is the observed value for the ith observation.
- $\beta_0, \beta_1, ..., \beta_p$ are the regression coefficients to be estimated.
- x_{ij} is the value of the jth predictor variable for the ith observation.
- λ is the regularization parameter.

*3.5.4. Support vector Regression (SVR)**:***

SVR is a powerful regression tool based on Support Vector Machines (SVM). Unlike regular regression methods that focus on fitting a straight line, SVR is adept at handling curvy, non-linear relationships between data. It achieves this by finding a flat, multi-dimensional surface (called a hyperplane) that best captures the connection between the input data and the predicted outcome. This makes SVR particularly useful for tasks where the data doesn't follow a simple straight line. Compared to classification problems where the output falls into distinct categories, regression problems deal with continuous outputs. See Figure: 5

i) Hyperplane : In SVM this is basically the separation line between the data classes. In SVR, it is defined as the line that will help us predict the continuous value or target value.

ii) Boundary Line : The Boundary Line is made up of two parallel lines on each side of the support vector, with an error threshold value (epsilon) for each. These lines create a margin between data points.

iii) Support Vectors : Support Vectors are the data points that are closest to the boundary. They are the vectors that influence the position of the hyperplane.

iv) Kernel : The Kernel is a function that maps lower dimensional data into higher dimensional data. This function is important because SVR performs linear regression in a higher dimension. There are several types of kernel functions, such as polynomial kernel, Gaussian kernel, and sigmoid kernel.

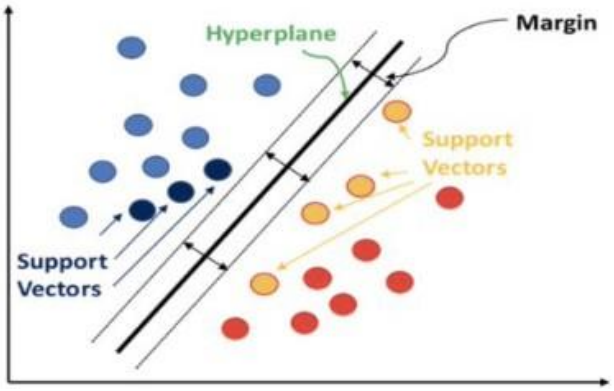

Figure : 5 (SVR)

3.5.5. Random Forest Regression **:**

Random Forest is an ML algorithm that is mainly used to solve classification and regression problems by combining decision trees. The algorithm combines multiple decision trees to form a forest. Each tree is trained using different subsets of data and a randomly selected set of features. This approach results in the creation of many trees that offer different perspectives to capture additional features and relationships in the data. The primary goal of Random Forest is to obtain stable and reliable predictions by combining the forecasts of each tree. It can also effectively reduce overfitting and is not recommended for high-dimensional data. Random Forest is useful for evaluating features, such as feature selection and importance ranking, and is widely used in both industry and academic fields. Figure : 6 (Random Forest Regression)

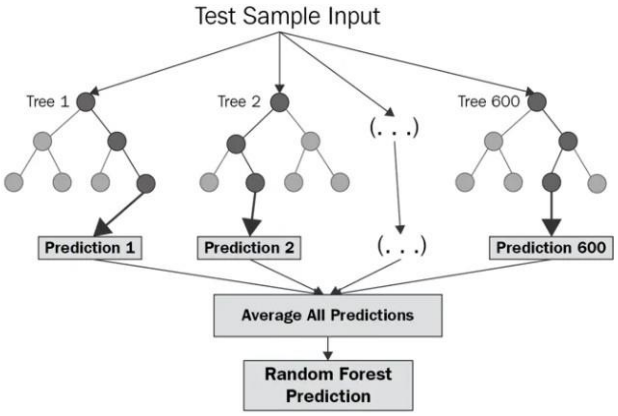

Figure : 6 (Random Forest Regression)

3.6 Evaluation Measures

3.6.1 Mean Absolute Error

Mean Absolute Error (MAE) measures the difference between two values.. It calculates the average error magnitude between predicted and actual values, regardless of their direction.(i.e., whether the errors are positive or negative) Its formula is Shown as Equation 1

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t|,$$

Equation : 1

Where, A_t is the true value , F_t is the prediction value, The formula sums the absolute difference between forecasted values and actual values, divided by the number of samples.

3.6.2 Mean Squared Error

The Mean Squared Error (MSE) is a metric used to measure the performance of predictors. It assesses how well the predictions match the actual observations. A lower MSE value indicates better performance. The MSE value is calculated by averaging the squared errors. The calculation considers the spread of the predictions across data samples (variance) and how closely the average predicted value matches observations (bias). Its formula is Shown as Equation 2.

$$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2,$$

Equation : 2

Where, A_t is the observed value , F_t is the prediction value, n is total number of Samples.

3.6.3 Root Mean Square Error

The Root Mean Square Error (RMSE) is a measurement used in regression analysis to determine the amount of variance between predictions made by a model and the actual values of the data. This value provides a ruler for the spread of the differences between the predicted and actual values around the model. RMSE helps us understand how closely the data points are clustered around the best-fit line or curve. The formula for RMSE involves calculating the average of the squared differences between predicted and actual values, and then taking the square root of that average.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{A_t - F_t}{A_t}\right)^2},$$

Equation : 3

Where, A_t is the observed value , F_t is the prediction value, n is total number of Samples.

3.6.4 R Squared Error

R-squared (R2) is a significant statistical measure that shows how much of the dependent variable's variance can be predicted from the independent variables in a regression model. In simpler terms, R-squared measures the amount of variation in the dependent variable that can be explained by the independent variables in a model.

R-squared = 1 - (SSres / SStot)

Equation : 4

Where, SSres (The residual sum of squares) is the sum of squared differences between the predicted values and the actual values

SStot (The total sum of squares) is the sum of squared differences between the actual values and the mean of the dependent variable.

4. Experiments and Results

Five different models, Linear Regression, Ridge Regression, Lasso Regression, Support Vector Machine, and Random Forest, were utilized to predict Sales price. Their effectiveness was evaluated by testing them against the data. The three regression techniques, Linear Regression, Ridge regression and Lasso regression, share a lot of similarities with each other. The difference between these techniques lies only in terms of regularization norms. Similar to Lasso regression, Ridge regression also requires an alpha parameter. The comparison between the predicted and actual values is performed on the test data and the results are visualized on a graph Figure

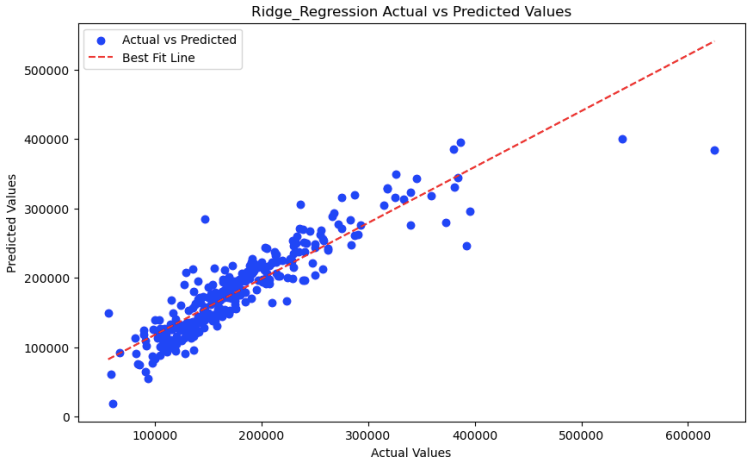

Figure :7 Predicted Values vs Actual Values (Ridge Regression)

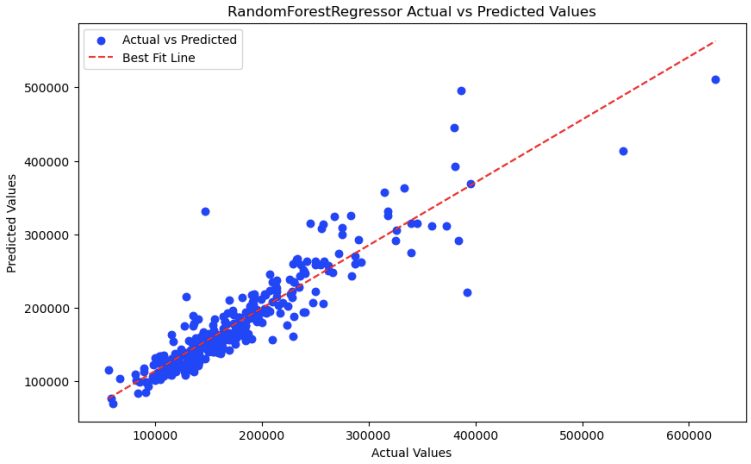

Figure :8 Predicted Values vs Actual Values (Random Forest Regression)

To distinguish between Ridge regression (Figure 7), Random Forest regression (Figure 8) and Support Vector Regression (Figure 9) it is best practice to visually compare the data distribution. Based on the graphs, it can be seen that SVR and RFR performed well in distributing the data. When all data points are in close proximity to the best-fit line, there is less error, which should align with the distribution of the data points. By examining the graphs, it is evident that there is minimal deviation of the data points from the best-fit line in RFR and SVR regression compared to Ridge regression. This implies that RFR and SVR have less error than Ridge regression.

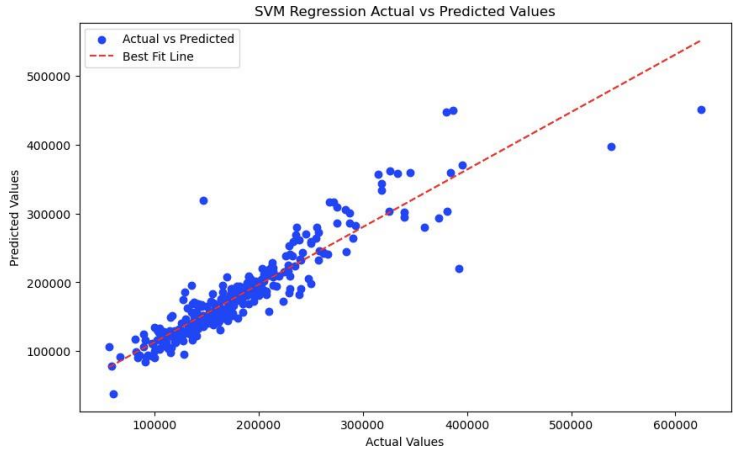

Figure :9 Predicted Values vs Actual Values (SVM Regression)

Based on the evaluation metrics(TABLE I), it can be observed that the traditional linear, Lasso, and ridge models exhibit a strong R2 score, which indicates a potentially good fit for the data. However, despite their high R2 scores, the models have high values for MAE and RMSE, which implies that there are significant prediction errors that are not captured by the R2 score. On the other hand, the SVR regression and Random Forest models show high R2 scores and low MAE and RMSE values, suggesting a more generalized model that fits the dataset well.

Model	MAE	MSE	RMSE	R2 Score
SVR	17602.738811	8.270466e+08	28758.417423	0.848461
RandomForestRegressor	18828.430862	8.702714e+08	29500.362813	0.840540
Ridge	20552.386254	9.960398e+08	31560.097994	0.817496
Lasso	20585.312111	9.982372e+08	31594.891941	0.817093
LinearRegression	20591.032999	9.986892e+08	31602.044571	0.817010

TABLE I: Evaluation metrics

5. Conclusion and After Work

The present study discusses a comprehensive methodology for predicting real estate prices using advanced machine learning techniques. The study analysed various regression models, and the Support Vector Machine (SVM) model emerged as the most effective model with an accuracy rate of 84.8%. The SVM model outperformed all other regression models in terms of accuracy and stability, followed closely by the Random Forest Regression model with an accuracy rate of 84.0%. Other regression models, such as Ridge, Lasso, and Linear Regression, also achieved remarkable accuracy rates of 81.0%.

The study's findings provide valuable insights for predicting house prices, which can help investors and common people make informed decisions. Accurate predictions can lead to better decision-making scenarios in the industry. The Support Vector Regression (SVR) model also excelled in all performance metrics, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R-squared).

The Future study recommends expanding the dataset to include a wider range of demographic and house variables to improve the models' accuracy. Additionally, applying advanced techniques such as deep learning and ensemble methods could enhance the models' accuracy and efficiency further. The study also suggests implementing real-time updates to ensure the models stay relevant. The application of these methods to other house decision-making processes could markedly enhance accuracy and efficiency across the industry.

6. References :

1. Venkatesh Panchapagesan,Madalasa Venkataraman, Chinmoy Ghosh “On the Impact of Infrastructure Improvement on Real Estate Property Values: Evidence from a Quasi-natural Experiment in an Emerging Market ” IIMB-WP No. 646/2021

2. Alam I. Asadov , Mansor H. Ibrahim , Mansor H. Ibrahim “Impact of House Price on Economic Stability: Some Lessons from OECD Countries” , (21 February 2023)

3. “4 Primary Factors that Influence Home Values” : link : www.malenohomes.com/blog/factors-that-influence-price-of-home.

4. Kang, Y., Zhang, F., Peng, W., Gao, S., Rao, J., Duarte, F., & Ratti, C. (2020). “Land Use Policy Understanding house price appreciation using multi-source big geo-data and machine learning”. Land Use Policy, July, 104919. https://doi.org/10.1016/j.landusepol.2020.104919

5. Greenaway-mcgrevy, R., & Sorensen, K. (2021). A Time-Varying Hedonic Approach to quantifying the effects of loss aversion on house prices. Economic Modelling, 99(March), 105491. https://doi.org/10.1016/j.econmod.2021.03.010\

6. Lennon H. T. Choy ,Winky K. O. Ho , Department of Real Estate and Construction, University of Hong Kong ,Independent Researcher, Hong Kong SAR, China “The Use of Machine Learning in Real Estate Research”

7. Baldominos, A.; Blanco, I.; Moreno, A.J.; Iturrarte, R.; Bernárdez, Ó.; Afonso, C. Identifying real estate opportunities using machine learning. Appl. Sci. 2018, 8, 2321. [CrossRef]

9. Sun, D.; Du, Y.; Xu, W.; Zuo, M.; Zhang, C.; Zhou, J. Combining online news articles and web search to predict the fluctuation of real estate market in big data context. Pac. Asia J. Assoc. Inf. Syst. 2015, 6, 19–37.

10. Zhihang Liu(2023) Real Estate Price Prediction based on Supervised Machine

Learning Scenarios 11. Janssen, C., Söderberg, B., & Zhou, J. (2001). Robust estimation of hedonic models of price and income for investment property. Journal of Property Investment & Finance, 19(4), 342-360

12. Afonso, B., Melo, L., Oliveira, W., Sousa, S., & Berton, L. (2019). Housing Prices Prediction with a Deep Learning and Random Forest Ensemble. In Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional,389-400.

13. Baldominos, A., Blanco, I., Moreno, A. J., Iturrarte, R., Bernárdez, Ó., & Afonso, C. (2018). Identifying Real Estate Opportunities Using Machine Learning, Applied sciences, 8(11), 2321

14. Aydemir, E., Aktürk, C., & Yalçınkaya, M. A. (2020). Estimation of Housing Prices with Artificial Intelligence, Turkish Studies, 15(2), 183-194.

15. Varma, A., Sarma, A., Doshi, S., & Nair, R. (2018). "House Price Prediction Using Machine Learning and Neural Networks". In 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018, 1936-1939. IEEE.

16. H. Peng et al., "Lifelong Property Price Prediction: A Case Study for the Toronto Real Estate Market," IEEE Trans Knowl Data Eng, vol. 35, no. 3, pp. 2765– 2780, 2023, doi: 10.1109/TKDE.2021.3112749.

17. Jiao Yang Wu (2023) Housing Price pr Housing Price prediction Using Suppor ediction Using Support Vector Regr or Regress