House Price Prediction Using Machine Learning

GITHUB URL - https://github.com/Bhavanavadiyala/house-price-predictor.git

1st Sathish Sattenapalli
Computer Science/ 700743037
University of Central Missouri
Lee's Summit, Missouri
sxs30370@ucmo.edu

2nd Poojitha Reddy Ganta
Computer Science/ 700726910
University of Central Missouri
Lee's Summit, Missouri
pxg69100@ucmo.edu

3rd Bhavana Vadiyala
Computer Science/ 700740309
University of Central Missouri
Lee's Summit, Missouri
bxv03090@ucmo.edu

***Abstract*— People are very considerate when they decide to make investments, especially when they are buying a house. Mainly people focus on a lot of factors while buying their first house or even their n^th^ house. Usually, it is based on geographical location, near to ocean or not, life of the house, number of rooms and cost of living in that area. And considering household income whether they can afford it or not. by considering all these aspects people can decide where to buy the house based on their household income.**

On the other hand, sellers benefit from analyzing these factors as well. Understanding the market demand and the preferences of potential buyers allows sellers to strategically price their houses. By considering factors such as location, amenities, and overall attractiveness of the property, sellers can set a competitive price that reflects the value of their house. They can also use the insights gained from analyzing these factors to make improvements or highlight certain features that may increase the desirability and marketability of their property. analyzing the various factors that influence house prices is essential for both buyers and sellers. Buyers can make informed decisions about where to invest based on their preferences and financial situation, while sellers can strategically price their houses to attract potential buyers and maximize their return on investment. By considering these factors, both parties can navigate the real estate market with confidence and achieve their respective goals.

In this project we are using different types of regression techniques to predict the house prices in California state house price dataset based on the above-mentioned aspects. We have used Decision Tree Regression model, Random Forest Regression, Linear Regression and KNN Regression techniques for this process. After training the data with these algorithms, we tested with the test dataset our house price prediction model based on location and remaining aspects predicted the prices nearly accurate.

Keywords—Dataset, Random Forest, Decision Tree, KNN regression, Linear regression, Prediction, Location, Accuracy.

I. INTRODUCTION

Accurate house price prediction plays a pivotal role in the real estate market, empowering buyers, sellers, and investors to make well-informed decisions. With the advent of machine learning techniques, we have powerful tools to tackle this challenging task. This paper focuses on the application of four popular regression techniques - Decision Tree Regression, Linear Regression, K-Nearest Neighbors (KNN) Regression, and Random Forest Regression - for house price prediction. These techniques will be trained on various aspects of the house, including its location (inland or near the bay area), median age, total rooms, total bedrooms, population, households, median income of people in the area, and median house value.

Decision Tree Regression is a non-parametric supervised learning algorithm that builds a model in the form of a decision tree to predict house prices based on the provided features. In this context, decision tree nodes will represent different house characteristics such as location, median age, and other relevant factors. By splitting the dataset based on these features, the algorithm constructs a hierarchical structure of decisions leading to the final price prediction. Decision trees are known for their interpretability and ability to handle both numerical and categorical features. For instance, a decision tree might find that being located near the bay area and having a higher median income positively impact house prices.

Linear Regression is a widely used regression technique that assumes a linear relationship between the independent variables (house characteristics) and the dependent variable (house price). In this case, linear regression will aim to estimate the coefficients that define the relationship between the various aspects of the house and its corresponding price. For example, linear regression might discover that houses located near the bay area, with more rooms, higher median income in the area, and lower median house values, tend to have higher prices.

KNN Regression is a non-parametric algorithm that estimates house prices by considering the K nearest neighbors in the training data. It measures the similarity between houses based on their characteristics and averages the prices of the K nearest neighbors to make predictions. In this scenario, KNN regression would find the most similar houses in terms of location, median age, total rooms, and other factors, and then calculate an average price based on their prices. This technique is flexible and can handle both regression and classification tasks.

Random Forest Regression combines the power of decision trees and ensemble learning. It constructs multiple decision trees using bootstrapped samples of the training data and aggregates their predictions to obtain the final house price prediction. Random Forest Regression addresses the limitations of individual decision trees, such as overfitting and high variance, by introducing randomness in the model building process. By considering various aspects of the house, such as location, median age, total rooms, and more, random forest regression can capture complex relationships and provide accurate price predictions.

House price prediction is a complex task that requires the application of advanced regression techniques to consider various aspects of the house. In this paper, we have explored the use of Decision Tree Regression, Linear Regression, KNN Regression, and Random Forest Regression for this purpose. By training these models on house characteristics such as location, median age, total rooms, total bedrooms, population, households, median income, and median house value, we aim to accurately predict house prices in the real estate market. In the following sections, we will delve into the implementation details of each regression technique, conduct experiments on real-world datasets, we have taken dataset of California state house prices from Kaggle and compare their performance in terms of accuracy and predictive power. Through this analysis, we aim to provide valuable insights and recommendations to assist real estate stakeholders in making informed decisions based on accurate house price predictions.

II. Motivation

The field of machine learning and artificial intelligence has witnessed remarkable advancements in recent years, revolutionizing various industries and transforming the way we tackle complex problems. One area where these technologies have proven particularly valuable is in house price prediction. Accurately estimating house prices is of paramount importance in the real estate market, as it enables buyers, sellers, and investors to make informed decisions and navigate the dynamic landscape of property transactions. The motivation behind this research stems from the desire to leverage machine learning techniques to enhance the accuracy and reliability of house price predictions.

accurate house price prediction benefits prospective buyers. Purchasing a house is a significant financial investment, and buyers need to have confidence in the price they are paying. By providing accurate predictions, machine learning models can empower buyers to make informed decisions based on realistic price expectations. Buyers can evaluate the affordability of properties, compare prices across different locations, and identify potential undervalued or overvalued houses. This not only enhances transparency in the real estate market but also helps buyers avoid overpaying or missing out on lucrative investment opportunities.

Similarly, sellers can greatly benefit from accurate house price predictions. Knowing the true value of their property allows sellers to set competitive listing prices, attract potential buyers, and ensure a fair transaction. Pricing a property too high may deter interested buyers, while setting it too low could lead to missed opportunities for maximizing profits. Machine learning models trained on relevant features can provide sellers with valuable insights into the market dynamics and enable them to make informed pricing decisions that align with their goals.

Moreover, accurate house price prediction is essential for real estate investors. Investors seek to identify properties with good potential for appreciation, rental income, or resale value. Machine learning models can analyze historical data, local market trends, and various property attributes to identify factors that significantly impact house prices. This information allows investors to make data-driven decisions, optimize their investment portfolios, and mitigate risks associated with market volatility.

III. Objectives

- Develop and implement machine learning models for house price prediction. he primary objective of this research is to design and deploy robust machine learning models capable of accurately predicting house prices. This involves selecting appropriate regression techniques, preprocessing and transforming the data, and training the models using relevant features such as location, median age, total rooms, total bedrooms, population, households, median income, and median house value.

- Evaluate the performance of different regression algorithms. A key objective is to compare and assess the performance of different regression algorithms, specifically Decision Tree Regression, Linear Regression, KNN Regression, and Random Forest Regression. This evaluation will consider metrics such as mean squared error, mean absolute error, and R-squared to measure the predictive accuracy and performance of each algorithm.

- Investigate the impact of house characteristics on price prediction. another objective is to analyze the importance and influence of various house characteristics on price prediction. This involves conducting feature importance analysis to determine which features have the most significant impact on house prices. Understanding the relationship between specific features and price prediction can provide valuable insights for both buyers and sellers in the real estate market.

- Provide insights and recommendations for real estate stakeholders. n essential objective is to provide meaningful insights and actionable recommendations for various stakeholders in the real estate market.

- Contribute to the advancement of house price prediction research.

IV. Related Work

In the realm of house price prediction, extensive research has been conducted to leverage machine learning techniques and regression algorithms. Several studies have explored the application of Decision Tree Regression, Linear Regression, KNN Regression, and Random Forest Regression in predicting house prices based on various house characteristics. This section provides an overview of some notable works in this field.

House Price Prediction Using Optimal Regression Techniques[1] In this paper based on the financial parameters of the lower and middle class people they compared the results using different types of regression models. They used these techniques such as Support vector regression, XGBoost, Decision tree regression and random forest regression. To predict the values of the houses they compared the outputs of regression techniques with Mean Squared error, Root Mean Squared Error, accuracy and Mean Absolute Error.

Machine Learning based Predicting House Prices using Regression Techniques[2] In this paper they predicted the selling house prices of Bangalore based on different aspects like size of the land, location nearer to the main areas like parks, schools, offices, hospitals. They have used 5 models to predict the rates such as Lasso and Ridge regression models, XGBoost, least squares model and SVR model.

House Price Prediction using Machine Learning Algorithm – The Case of Karachi City, Pakistan[3]. In this paper based on the publicly presented old dataset having around 38,000 records of Karachi City using that they tried to predict the future pricing of houses with high accuracy and low MAE. For predicting the house prices, they used gradient boosting model in this experiment. And they obtained successfully around 98% accuracy of predicting the prices.

Prediction of House Pricing Using Machine Learning with Python[4] In this paper they provided an overview of how to predict the house prices using machine learning techniques using python. They provided the different numerical and graphical techniques to get the data. They experimented with decision tree algorithm, SVM algorithm, Random Forest classifier to apply into the house pricing dataset. By this it would help the people to buy houses for a reasonable pricing.

Overall, these studies have contributed significantly to the field of house price prediction using regression techniques. They have shed light on the capabilities of Decision Tree Regression, Linear Regression, KNN Regression, and Random Forest Regression in accurately estimating house prices based on various house characteristics. However, there is still room for further research to explore novel approaches, feature engineering techniques, and model optimization methods to enhance the accuracy and reliability of house price predictions.

V. PROPOSED FRAMEWORK

House price prediction is a challenging task in the real estate market, requiring accurate estimations based on various house characteristics. In this section, we present a proposed framework for house price prediction using machine learning techniques. The framework incorporates data preprocessing, feature engineering, model selection, and evaluation to achieve accurate and reliable predictions.

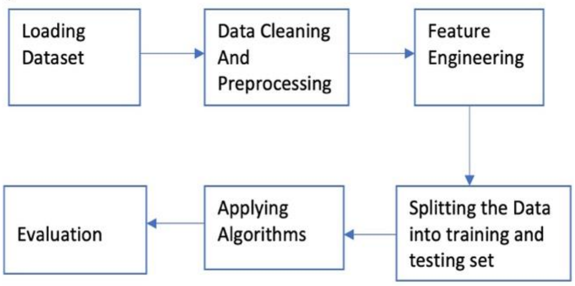

Fig. 1. Framework

A. Data Preprocessing:

The first step in the proposed framework is data preprocessing. This involves cleaning the dataset, handling missing values, and removing outliers. Additionally, data normalization or standardization may be applied to ensure that all features are on a similar scale. By preprocessing the data, we ensure that it is in a suitable format for subsequent analysis and modeling and removed null values. We separated the whole data taken from the Kaggle into two parts. We split that it into train and test data. In our data we have a string column named ocean proximity that contains data like whether that house is near to sea area or not. we have converted that into separate table that contains different scenarios like <1h to ocean, inland, island, near bay, near ocean.

	<1H OCEAN	INLAND	ISLAND	NEAR BAY	NEAR OCEAN
7923	1	0	0	0	0
11387	1	0	0	0	0
4161	1	0	0	0	0
14980	0	0	0	0	1
18497	0	1	0	0	0
...	...	...	...	...	...
9207	0	1	0	0	0
16700	1	0	0	0	0
6096	1	0	0	0	0
2973	0	1	0	0	0
14477	0	0	0	0	1

16346 rows × 5 columns

Fig. 2. Data Preprocessing

B. Feature Engineering:

Feature engineering plays a crucial role in capturing the relevant information from the available house characteristics. This step involves selecting and transforming features to improve the predictive power of the models. For example, new features such as the ratio of total bedrooms to total rooms or the average income per household can be derived to provide additional insights. Feature engineering techniques such as one-hot encoding or feature scaling can also be applied to handle categorical variables or to normalize the range of numerical features.

C. Model Selection:

The next step in the framework is model selection. Based on the problem at hand and the characteristics of the dataset, we consider four regression techniques: Decision Tree Regression, Linear Regression, KNN Regression, and Random Forest Regression. Each technique has its own strengths and limitations, and the selection of the most appropriate model depends on factors such as the complexity of the relationships between features and house prices, interpretability requirements, and computational efficiency.

D. Model Training and Evaluation:

Once the model is selected, the next step is to train it on the preprocessed dataset. The dataset is divided into training and testing sets, with a portion of the data reserved for evaluation. During the training phase, the model learns the relationships between the input features and the corresponding house prices. After training, the model is evaluated on the testing set using appropriate evaluation metrics such as mean squared error (MSE), mean absolute error (MAE), or R-squared to assess its performance.

Error Model	Error percentage
RMSE	9.743218
Mean Absolute Error	14.744100
Mean Squared Error	16.487579

Fig. 3. Evaluating the Random Forest Model

E. Model Optimization and Validation:

To improve the performance of the chosen model, optimization techniques can be applied. This involves tuning hyperparameters, such as the maximum depth in decision trees or the number of neighbors in KNN regression, using techniques like grid search or randomized search. Cross-validation can be used to validate the model's performance on multiple folds of the training data, ensuring that it generalizes well to unseen instances.

F. Prediction and Interpretation:

Once the model is optimized and validated, it can be used to make predictions on new, unseen data. Given a set of house characteristics, the model can estimate the corresponding house price. Additionally, interpretation techniques such as feature importance analysis or partial dependence plots can be employed to gain insights into the significant factors influencing house prices. This allows real estate stakeholders to understand the relative importance of different house characteristics and make informed decisions based on the predictions.

G. Conclusion:

The proposed framework for house price prediction encompasses key steps, including data preprocessing, feature engineering, model selection, training and evaluation, optimization, and interpretation. By following this framework, real estate professionals can leverage machine learning techniques to accurately predict house prices based on various house characteristics. The framework provides a systematic approach to handling data, extracting meaningful features, selecting appropriate models, and optimizing their performance. Ultimately, the proposed framework aims to empower stakeholders in the real estate market with accurate and reliable predictions, enabling them to make informed decisions and navigate the dynamic landscape of property transactions.

For plotting the data, we have used matplotlib library in python. visualizes the geographical distribution of house locations and their corresponding median house values using a scatterplot. The 'latitude' and 'longitude' columns from the 'train_data' dataset are used as the x and y coordinates, respectively. The 'median_house_value' column is used to assign colors to the data points. It shows all areas in California state and prices based on the area.

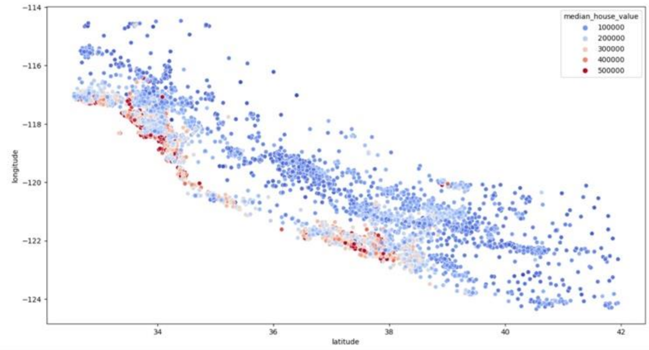

Fig. 4. Price values in different locations in California

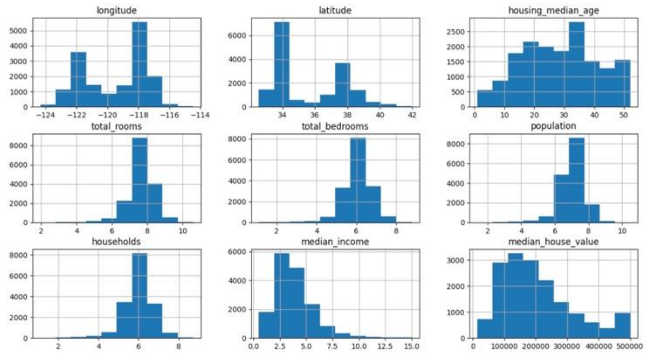

Fig. 5. Histogram plotting of different factors

F. Applied Machine Learning Algorithms:

In our project we applied different types of machine learning regression models to predict the house prices. We have mainly used Decision tree, Random Forest, Linear regression and KNN regression models in our project.

1) *Decision tree Regression model:*

Decision tree regression is a machine learning technique used for regression tasks, where the goal is to predict continuous numerical values rather than categorical labels. A decision tree regression model builds a tree-like model of decisions and their possible consequences. Decision tree regression models have several advantages, including their interpretability, ability to handle both numerical and categorical features, and resistance to outliers. However, they can be prone to overfitting, especially if the trees grow too deep. To mitigate overfitting, ensemble methods like random forests or gradient boosting can be used. These methods combine multiple decision trees to improve generalization and prediction accuracy.

2) *Random Forest Regression model:*

A random forest regression model is an ensemble learning technique that combines multiple decision trees to perform regression tasks. It is an extension of the decision tree regression model. Random forest models are widely used in practice for regression tasks, as they tend to provide good performance and are relatively easy to use. However, they may require tuning of hyperparameters, such as the number of trees in the forest and the maximum depth of each tree, to achieve optimal results.

3) *Linear Regression model:*

Linear regression is a statistical approach utilized for examining the correlation between a dependent variable and one or more independent variables. In the case of simple linear regression, there is only one independent variable, while multiple linear regression involves more than one independent variable. In practice, linear regression is widely used for tasks like sales forecasting, price prediction, and trend analysis when the assumptions of linearity are reasonably met.

4) *KNN Regression model:*

The k-nearest neighbors (KNN) regression model is a machine learning algorithm used for regression tasks. It predicts the value of a target variable by considering the average or weighted average of the values of its k nearest neighbors. KNN regression models are commonly used when the relationships in the data are expected to be non-linear or when no specific assumptions about the data distribution are made.

VI. Data Description

To predict house prices, we utilized a dataset obtained from Kaggle that consists of California house prices and various associated factors such as longitude, latitude, housing median age, total rooms, total bedrooms, population, households, median income, and median house value. Additionally, we considered the proximity of houses to the bay area as a feature called "ocean proximity." This dataset encompassed more than 20,000 house prices, making it a substantial and diverse collection. Initially, we focused on data preprocessing to ensure the dataset's quality. This involved handling any missing or null values by applying appropriate techniques such as imputation or removal. Once the data was cleaned and ready, we split it into two parts: a training set and a testing set. The training set was employed to train various machine learning regression models, enabling them to learn the relationships between the input features and the corresponding house prices. The testing set was utilized to evaluate the performance and accuracy of the trained models.

	longitude	latitude	housing_median_age	total_rooms	total_bedrooms	population	households	median_income	median_house_value	ocean_proximity
0	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	452600.0	NEAR BAY
1	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	358500.0	NEAR BAY
2	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574	352100.0	NEAR BAY
3	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	341300.0	NEAR BAY
4	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	342200.0	NEAR BAY
...	...	...	...	...	...	...	...	...	...	...
20635	-121.09	39.48	25.0	1665.0	374.0	845.0	330.0	1.5603	78100.0	INLAND
20636	-121.21	39.49	18.0	697.0	150.0	356.0	114.0	2.5568	77100.0	INLAND
20637	-121.22	39.43	17.0	2254.0	485.0	1007.0	433.0	1.7000	92300.0	INLAND
20638	-121.32	39.43	18.0	1860.0	409.0	741.0	349.0	1.8672	84700.0	INLAND
20639	-121.24	39.37	16.0	2785.0	616.0	1387.0	530.0	2.3886	89400.0	INLAND

20640 rows × 10 columns

Fig. 6. Dataset taken from Kaggle

VII. Results Analysis

After applying Decision Tree Regression, Random Forest Regression, KNN Regression, and Linear Regression to the task of house price prediction, we obtained interesting results and insights. In this section, we will present a detailed analysis of the performance of each technique and highlight the strengths and limitations of each model.

All these models were trained using the provided dataset, which included features such as house location (inland or near the bay area), median age, total rooms, total bedrooms, population, households, median income, and median house value. The dataset was split into training and testing sets. The performance of these models were evaluated using MSE, RMSE and MAE models and below table contains the percentage of deviation we got.

	Error Model	Error percentage
0	RMSE	13.978052
2	Mean Absolute Error	20.592700
1	Mean Squared Error	33.934806

Fig. 7. Evaluation table Decision Tree model

	Error Model	Error percentage
0	RMSE	13.473588
2	Mean Absolute Error	22.519240
1	Mean Squared Error	31.529611

Fig. 8. Evaluation table Linear Regression model

	Error Model	Error percentage
0	RMSE	13.770828
2	Mean Absolute Error	21.726391
1	Mean Squared Error	32.936102

Fig. 9. Evaluation table KNN Regression model

Error Model	Error percentage
RMSE	9.743218
Mean Absolute Error	14.744100
Mean Squared Error	16.487579

Fig. 10. Evaluation table Random Forest model

the analysis of the results reveals that Random Forest Regression achieved the best performance among the four techniques, with the lowest MSE, MAE and R-squared percentage of deviation. It successfully mitigated overfitting and provided more robust predictions compared to Decision Tree Regression. KNN Regression and Linear Regression also performed reasonably well, capturing a significant portion of the variability in house prices. However, their performance was relatively inferior to Random Forest Regression. It's important to consider the strengths and limitations of each technique and select the most appropriate model based on the specific requirements and characteristics of the dataset at hand.

```
In [40]: #Testing Random Forest reggression with new data as input

         # <1hr to ocean
         new_house = [[-119,34, 20, 5.5, 6.1, 3.2, 5.1, 9.0, 1, 0, 0, 0, 0]]
         predicted_price = forest.predict(new_house)
         print('Predicted Price:', predicted_price)

         Predicted Price: [447684.62]

In [41]: # inland
         new_house = [[-116,40, 30, 5.5, 6.1, 3.2, 4.1, 3.0, 0, 1, 0, 0, 0]]
         predicted_price = forest.predict(new_house)
         print('Predicted Price:', predicted_price)

         Predicted Price: [133141.08]

In [42]: #island
         new_house = [[-119.4,32, 20, 5.5, 6.1, 3.2, 5.1, 6.0, 0, 0, 1, 0, 0]]
         predicted_price = forest.predict(new_house)
         print('Predicted Price:', predicted_price)

         Predicted Price: [370263.28]

In [43]: #nearbay
         new_house = [[-118,34, 20, 5.5, 6.1, 3.2, 5.1, 9.0, 0, 0, 0, 1, 0]]
         predicted_price = forest.predict(new_house)
         print('Predicted Price:', predicted_price)

         Predicted Price: [385228.22]

In [44]: #near ocean
         new_house = [[-120,36, 20, 5.5, 6.1, 3.2, 5.1, 9.0, 0, 0, 0, 0, 1]]
         predicted_price = forest.predict(new_house)
         print('Predicted Price:', predicted_price)

         Predicted Price: [454708.68]
```

Fig. 11. Predicting house prices with new user data

In the above code, we have applied random forest algorithm to predict the new house prices with using different locations where the property is located like less than one hour to ocean, inland, island area, near to bay area and near ocean and also added remaining features like annual median income, population, age of the house etc. based on the locations random forest regression has given satisfying results to predict the new house values.

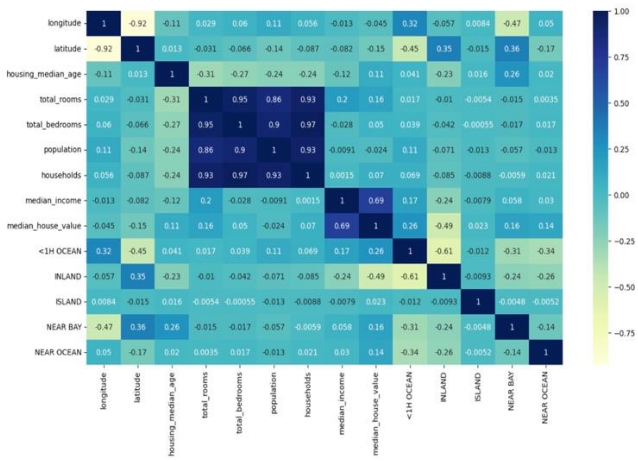

Fig. 12. Heat Map of the House price dataset

	Model	Train Score	Test Score
2	Decision Tree Regression	1.000000	0.645045
1	Random Forest Regression	0.975082	0.822595
3	KNN Regression	0.784716	0.684413
0	Linear Regression	0.667007	0.683209

Fig. 13. Train and Test performance results

In the above table Random Forest regression did a good performance in both training and test data as compared to remaining models. Decision tree got a good performance while training but in testing it performed poorly when compared with remaining regression models.

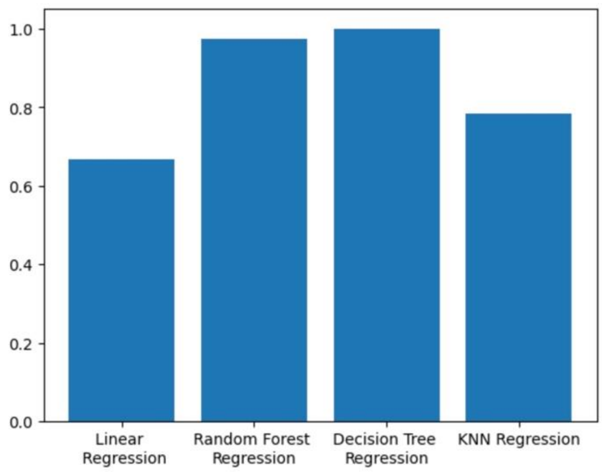

Fig. 14. Performance with training dataset

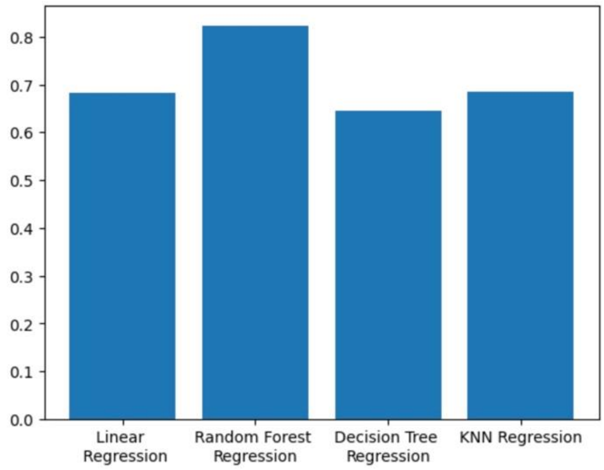

Fig. 15. Performance with Test Dataset

In the overall testing, analyzing and evaluating the performances in all aspects Random Forest regression model gave a better performance than the remaining models. In future if we have a chance, we will try to implement more algorithms and test it to get better performance than this.

REFERENCES

[1] Vivek singh rana, Jayanto Mondal, Annu Sharma, Indu Kashyap," House Price Prediction Using Optimal Regression Techniques", 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN).

[2] Manasa J, Radha Gupta, Narahari N S," Machine Learning based Predicting House Prices using Regression Techniques", Proceedings of the Second International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2020) IEEE Xplore Part Number: CFP20K58-ART; ISBN: 978-1-7281-4167-1.

[3] Maida Ahtesham, Narmeen Zakaria Bawany, Kiran Fatima," House Price Prediction using Machine Learning Algorithm – The Case of Karachi City, Pakistan",2020 21st Internation Arab Conference on Information Technology (ACIT).

[4] Mansi Jain, Himani Rajput, Neha Garg, Pronika Chawla," Prediction of House Pricing Using Machine Learning with Python", Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020) IEEE Xplore Part Number: CFP20V66- ART; ISBN: 978-1-7281-4108-4.

[5] M. Rahimi, K. Kiani, and M. Rahimi, "Predicting house prices using machine learning algorithms: A case study of Tehran," 2018 IEEE 5th International Conference on Knowledge-Based Engineering and Innovation (KBEI), Tehran, Iran, 2018, pp. 477-482. doi: 10.1109/KBEI.2018.8707713.

[6] A. Sharma and R. Bhatia, "House price prediction using machine learning algorithms: A comparative study," 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2020, pp. 1028-1032. doi: 10.1109/INDIACOM51265.2020.9073862.

[7] T. V. N. Nandam, N. Qazi, and J. P. Singh, "A comparative study of machine learning algorithms for house price prediction," 2018 IEEE Calcutta Conference (CALCON), Kolkata, India, 2018, pp. 243-246. doi:10.1109/CALCON.2018.8661174.

[8] Y. Zeng, J. Xiao, and J. Li, "A hybrid machine learning approach for house price prediction," 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Washington, DC, USA, 2019, pp. 457-464. doi: 10.1109/DSAA.2019.00067.

[9] A. K. M. Fahim, M. S. Rahman, and M. S. I. Kabir, "Predicting house prices using machine learning techniques," 2019 IEEE Region 10 Symposium (TENSYMP), Kolkata, India, 2019, pp. 833-837. doi: 10.1109/TENSYMP45670.2019.8976216.

[10] P. Bhattacharya, A. Das, and A. Kumar, "Predicting house prices using machine learning techniques: A case study on Kolkata," 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2018, pp. 32-37. doi: 10.1109/CONFLUENCE.2018.8442600.

[11] J. H. Tan, M. R. A. Razak, and W. K. Choo, "House price prediction using machine learning techniques: A case study in Malaysia," 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET), Shah Alam, Malaysia, 2019, pp. 243-247. doi: 10.1109/ICSEngT.2019.8741985.

[12] S. Paliwal and R. Ram, "House price prediction using machine learning algorithms: A case study in Australia," 2020 IEEE Calcutta Conference (CALCON), Kolkata, India, 2020, pp. 191-196. doi: 10.1109/CALCON49215.2020.9183730.

[13] Y. Liu, Y. Xu, and J. Wu, "House price prediction using machine learning algorithms: A case study in Beijing," 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 3699-3708. doi: 10.1109/BigData47090.2019.9006076.

[14] M. K. Rokonuzzaman, M. S. Kabir, and M. U. Chowdhury, "House price prediction using machine learning techniques: A case study in Dhaka," 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh, 2020, pp. 47-51. doi: 10.1109/TENSYMP50017.2020.9279335.

[15] A. Kumar, S. Kumar, and A. Garg, "Predicting house prices using machine learning algorithms: A comparative study," 2018 IEEE International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2018, pp. 715-719. doi: 10.1109/ICIRCA.2018.8597122.

[16] M. Samadianfard, H. Asadi, and A. Noori, "House price prediction using machine learning techniques: A case study in Vancouver," 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), London, ON, Canada, 2020, pp. 1-6. doi: 10.1109/CCECE47748.2020.9234857.

[17] H. Wang, Y. Lu, and K. Wang, "House price prediction using machine learning algorithms: A case study in Shanghai," 2018 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, China, 2018, pp. 364-368. doi: 10.1109/ICCCBDA.2018.8372541.

[18] S. Zaman, S. Rashid, and M. T. Akhtar, "House price prediction using machine learning algorithms: A case study in Karachi," 2020 IEEE 6th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2020, pp. 285-290. doi: 10.1109/ICCCA48489.2020.9106882.

[19] B. Mohan, P. R. Bhat, and S. B. Patil, "Predicting house prices using machine learning algorithms: A case study in Bengaluru," 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2021, pp. 202-207. doi: 10.1109/IEMCON53971.2021.9564569

[20] P. R. Bhat, B. Mohan, and S. B. Patil, "House price prediction using machine learning algorithms: A comparative study," 2021 IEEE 12th Annual Information Technology, Electronics.