Proceedings of the International Conference on Sustainable Computing and Smart Systems (ICSCSS 2023)
IEEE Xplore Part Number: CFP23DJ3-ART; ISBN: 979-8-3503-3360-2

House and Rent Price Prediction System using Regression

Mrs. Sujata Khandaskar
Department of Computer Engineering(Vivekanand Education Society's Institute of Technology) Mumbai, India sujata.khandaskar@ves.ac.in

Chirag Panjwani
Department of Computer Engineering(Vivekanand Education Society's Institute of Technology) Mumbai, India
2020.chirag.panjwani@ves.ac.in

Vinit Patil
Department of Computer Engineering(Vivekanand Education Society's Institute of Technology) Mumbai, India
2020.vinit.patil@ves.ac.in

Deanna Fernandes
Department of Computer Engineering(Vivekanand Education Society's Institute of Technology) Mumbai, India
2020.deanna.fernandes@ves.ac.in

Prerna Bajaj
Department of Computer Engineering(Vivekanand Education Society's Institute of Technology) Mumbai, India
2020.prerna.bajaj@ves.ac.in

2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS) | 979-8-3503-3360-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSCSS57650.2023.10169290

***Abstract*—Real estate market analysis plays a crucial role in predicting house prices and rental trends. In recent years, several techniques have been developed to address this challenge, including machine learning (ML) algorithms, data mining approaches, and statistical models. There were a few issues which needed to be solved, like data quality and availability pose significant obstacles. Housing datasets often exhibit missing values, outliers, and inconsistent formats, which can hinder the performance of prediction models. Addressing these challenges requires careful data preprocessing techniques and collaboration with data providers and government agencies to ensure data accuracy and coverage. Existing systems face significant challenges that hinder their accuracy and reliability. Some of these challenges include limited availability of comprehensive and up-to-date data, complex market dynamics, and the need for efficient feature selection and model optimization. The proposed objectives includes Data collection and preprocessing, feature selection and engineering, Model selection and development along with evaluation, usability, comparison with existing systems, etc. By addressing these challenges, proposed system aims to provide more accurate and reliable predictions for real estate market analysis, aiding homeowners, tenants, and real estate professionals in making informed decisions.**

Keywords— Random forest Regression, Lasso Regression, Parameters, Machine Learning

I INTRODUCTION

Machine Learning (ML) is an artificial intelligence field that involves the creation of algorithms and statistical models that allow computers to learn from data and make predictions. In the context of house price prediction, ML algorithms can be used to analyze a wide range of variables such as location, property size, age, condition, and amenities in order to forecast the accurate and precise value of a specific property based on previous data.

Some of the ways ML algorithms are used in house price prediction model:

Linear Regression- This algorithm utilizes a linear equation to model the relationship between a dependent variable (house price) and one or more independent variables (square footage, location, age, etc.). The algorithm then utilizes historical data to train itself and based on learned relationships it can predict future house prices.
Decision Tree - It employs a tree-based model to examine the links between independent and dependent variables. It simplifies complex data and can help identify the most significant elements influencing property values.
Random Forests - Random Forests are advanced decision trees that predict by mixing several decision trees. By pooling the outcomes of numerous decision trees and selecting the most likely forecast, this algorithm

ML has become an indispensable tool for predicting house prices, and its application in the real estate market is quickly expanding. The key advantage of employing ML, regardless of the individual ML approach utilized, is its ability to automate the process of detecting correlations between variables. This is done by creating predictions. ML has the potential to alter the way real estate companies and consumers make decisions about buying, selling, or renting properties.

It can save time and reduce human error risk, making it a valuable tool on the market. To predict house prices dataset that consists of price, Area, Location, No. Of Bedrooms, Resale as the parameters and to predict rent prices dataset that consists of seller_type, bedroom, layout_type, property_type, locality, price, area, furnish_type, bathroom are used as parameters of major metropolitan cities like Hyderabad, Chennai, Delhi, Banglore, Mumbai, Kolkata.

II LITERATURE SURVEY

House price prediction has been a popular topic in machine learning due to its practical implications in the real estate market. Various machine learning algorithms have been developed to predict house prices based on different factors such as location, number of bedrooms, square footage, and nearby amenities.One common theme among the surveyed papers is the use of regression algorithms. Regression analysis is a statistical method used to estimate the relationships between dependent and independent variables. Many of the papers used multiple regression models, which involve analyzing the relationship between a dependent variable (house price) and multiple independent variables (features of the house such as number of bedrooms, square footage, and location).

For example, a research utilized a hedonic price model for regression analysis. The model assumes that the value of a property is the sum of its attributes based on the market. The authors also used Support Vector Regression (SVR), a supervised learning model trained on a subset of the training data, to predict house prices based on the given attributes.

Some studies analyzed house price prediction using the Lasso algorithm on datasets from Iowa, United States, and Malmo, Sweden. A study found that location attributes, such as proximity to common amenities like hospitals, public transportation, and restaurants, significantly affected house prices.

In addition, some papers utilized other machine learning algorithms such as random forest, K-nearest neighbor (KNN), and XGBoost.The study found that Lasso Regression provided the most accurate predictions, while XGBoost produced prediction models and checked for sequential problems in the data.

Lastly, some papers utilized robotic process automation (RPA) to reduce human efforts and improve data analysis. They divided the data into training and testing sets and scraped data from websites. The study utilized algorithms such as random forest, KNN, XGBoost, and CatBoost to predict house prices. The study found that random forest algorithms trained the model efficiently and produced more accurate predictions.

We came across several merits and demerits. We discovered several intriguing concepts,

such as temperature maps used to display the data, in which properties with high values were coloured in red and those with low prices were shaded in yellow. Also, the adoption of RPA (Robotic Process Automation) improved the accuracy and reduced the likelihood of mistakes. Some previous systems also contained the Catboost Algorithm, which takes human input. This system provides relevant things to the user using Content Based Filtering.

Existing systems, on the other hand, lacked sufficient parameter filtering, resulting in erroneous outputs. Some datasets were not sufficiently large. Some prior methods did not account for resale housing values. In conclusion, machine learning algorithms have been utilized extensively to predict house prices based on various factors such as location, square footage, number of bedrooms, and nearby amenities. Regression algorithms are commonly used, with Lasso regression providing the most accurate predictions in some studies.

III PROPOSED SYSTEM

House and rent price prediction system using ML is a system that aims to accurately forecast housing values for buyers and sellers. The system utilizes machine learning algorithms to analyze various variables such as location, property size, number of bedrooms, and more to predict the accurate and precise value of a specific property based on historical data.

The system follows the following methodology:

i) Data Collection: The first step is to gather the necessary data, including information about the house's location, area, number of bedrooms, bathrooms, etc.

ii) Data Preprocessing: The collected data is cleaned and transformed to ensure there are no duplicates or outliers. Missing values are handled by either dropping the entire tuple or replacing them with the most frequent value (for categorical variables) or the mean (for numerical variables).

iii) Feature Selection: The most significant variables influencing house prices are identified and selected as features. Irrelevant variables that could introduce noise into the system are eliminated for better accuracy. Techniques like Lasso regression, which sets coefficients of less important features to zero, are used for feature selection.

iv) Model Construction: Machine learning algorithms such as linear regression, decision trees, and neural networks are used to build the prediction models. The preprocessed data is used to train these algorithms, and evaluation metrics like mean squared error and R-squared are used to assess their performance. Random forest regression is commonly employed for its accuracy in house price prediction.

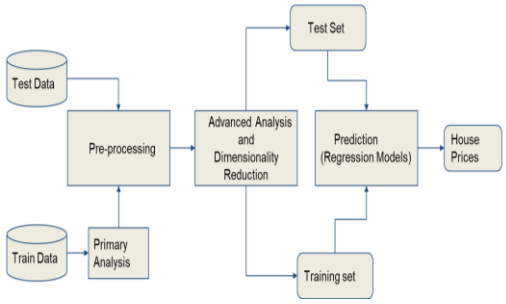
v) Website Development: The model is integrated into a user-friendly website frontend using web development tools like HTML, CSS, and JavaScript. The website allows users to browse through available houses and use the house and rent price prediction system to obtain accurate information. Features like advanced bookings and payment gateways can be added for an e-commerce experience.

vi) User Interface Design: The frontend of the website is designed to be user-friendly, with dropdowns, checkboxes, and other features to easily select parameters for property search and prediction.

The results of the house and rent price prediction system can be visualized through graphs and comparisons between predicted and actual prices. The system aims to provide accurate predictions by considering various factors and employing machine learning algorithms.

It is worth noting that the accuracy of the predictions depends on the quality and relevance of the collected data, the chosen features, and the performance of the machine learning algorithms. Continuous improvement and refinement of the system can be done by incorporating live price data from APIs and applying data mining techniques.

Overall, the house and rent price prediction system using ML aims to automate and enhance the reliability and accuracy of property price estimation, assisting both buyers and sellers in making informed decisions.



IV METHODOLOGY

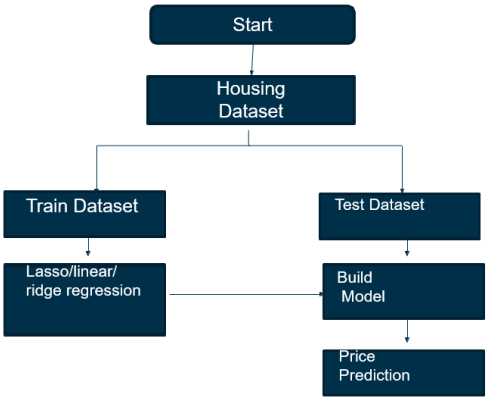
We collected data on house prices in major metropolitan cities from Kaggle. The dataset provides information on average home prices and rent prices in various neighborhoods of cities along with the flat features. Two different datasets were used for price and rent, with different parameters covering various features. The data collected was cleaned, transformed and organized before being used for predictive models. We checked for missing values which were handled by ignoring the entire tuple. Outliers impact model accuracy, so dealing with them is essential. For a better predictive model, certain parameters were established to deal with outliers. And different types of outliers were removed/handled differently to improve accuracy above the basic standard techniques utilized. Location and other categorical information was encoded using. One Hot Encoder guaranteed categorical variables were properly formatted for analysis.

The project was built considering various features like area size, number of bedrooms, locality, etc. And some parameters were used only for rent because of the different dataset. In house price prediction, resale/new was added. Several important features were considered when predicting rent prices, such as whether it was furnished, unfurnished or semi-furnished. A model to forecast house prices was created using preprocessed data and appropriate attributes. It accurately predicts prices and rentals based on those factors. Various machine learning algorithms like linear regression, random forest regression. We started off with LASSO regression for price prediction. But further we employed random forest regression to enhance our system's accuracy. The predicted prices by the system were validated by the actual prices in the area and were proved to be accurate enough. Graphs were plotted to understand the price variation with respect to the evaluation parameters chosen. By comparing the best scores across the models, one can identify which model performed the best with the given dataset and selected hyperparameters. The model with the highest R2 score indicates a better fit to the data. In our case it was LASSO regression.

The website front end was built using regular PHP, CSS, Bootstrap and JavaScript and integrated using Flask. The website consisted of drop down boxes for a friendly interface, which helped users select property features according to their choice. The evaluation parameters helped us predict prices. ML model consists of combined algorithms and multiple pre-processing techniques. We worked on parameters like furnished/unfurnished and seller type, etc. For rent, city, area, location, etc. For price after importing the dataset and libraries for graphing, Further, we divided the prices of each location into groups by calculating their mean values and grouped them into stats. Implemented a scatter matrix to simplify identification and analysis. We also calculated the price per bedroom which will help us further remove outliers from the database. Outliers were removed based on the quantitative analysis. In order to eliminate the means, StandardScalar was used to scale each feature to a unit variance. Any data point that did not fall within the range was discarded. We incorporated pipelining and column transformations. This helped us coordinate the input data into and output from the machine learning model. It covers the input of the raw data, the features, the outputs, the machine learning model and model parameters, and the outputs of predictions.

The test size for this model is 0.2, the random state is 42 and the stratification is based on the house location. We used Random Forest Regressor to predict house and rent prices. We selected random Forest Regressor as it gave the most accuracy compared to other algorithms. We received an accuracy of 90% and strive to improve it by extracting live prices from API's. Finally, we extracted the data and applied it as a model to the website. For rent price

prediction, we dropped the unnecessary parameters. Then we cleaned the data, performed quantitative analysis like getting the mean, quarters, min, max etc for removing outliers. We split the data into training and testing models, taking 20% for testing and 80% for training. We used one-hot encoding to transform categorical features into numerical dummy features useful for training models. For this experiment, we took the test size as 0.2 and looped the random state from 1-100 until we found the best level of accuracy. We used linear regression for rent price prediction. We received 89% accuracy and strive to improve accuracy through data mining.

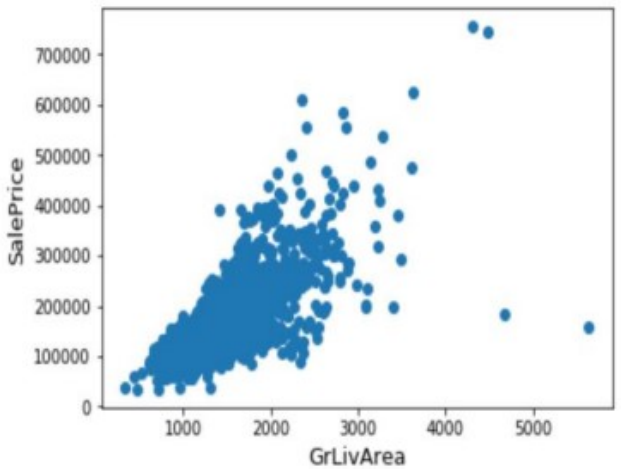


To make our website e-commerce we included other features like admin controls to add or delete listings of houses using PHP and Django apart from predicting prices.

Also, the user only needs to log in to our website to use our prediction tool. The user can then search for homes appropriately and compare pricing to expected home prices. Lastly, learn more about the residences available in the region through our e-commerce website. Our website also features rental homes and predicts rent costs. The user can compare costs easily due to this.

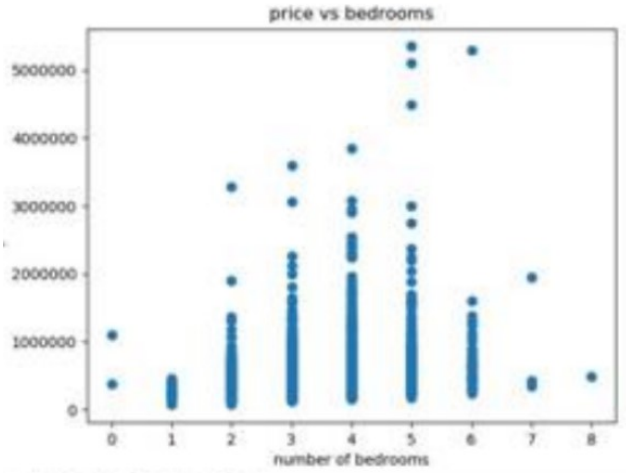
V RESULTS

After training and testing the model, we found variations between the price and the features we selected for the project.



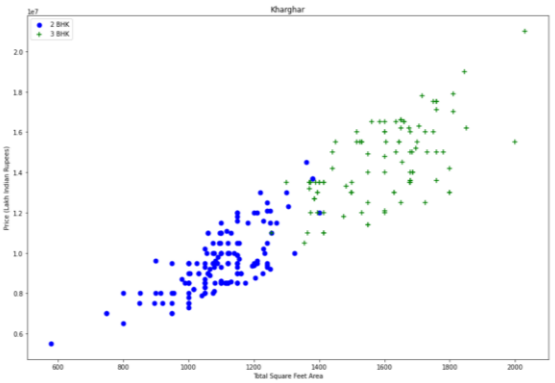
(a): Comparison between price and area.

According to the model trained, we observed that as the house/flat area increases so does the price. Hence we can see a positive correlation between the two parameters.



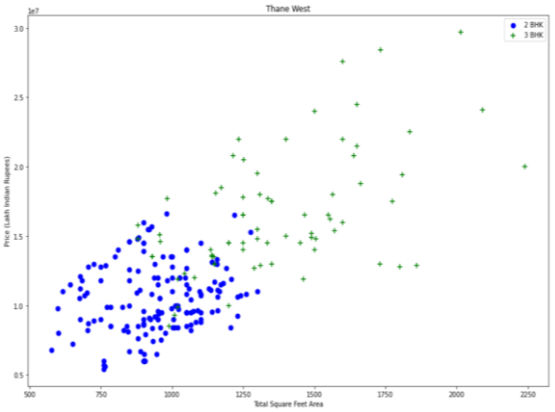
(b): Comparison between price and number of bedrooms

We also plotted graphs of different regions in metropolitan cities like Thane West and Kharghar to get an overview of the trend of prices. The graph after training is as follows:



(c) The graph shows the comparison

between the total square feet and the price in lakhs. This graph also shows a positive correlation between the two parameters in Kharghar.



(d) The graph shows the comparison between the total square feet and the price in lakhs. This graph also shows a positive correlation between the two parameters in Thane West. Hence, as the square feet increases so does the price.

Note: Blue: 2bhk

Green: 3bhk

	model	best_score	best_params
0	linear_regression	0.900945	{'normalize': True}
1	lasso	0.890001	{'alpha': 1, 'selection': 'cyclic'}
2	decision_tree	0.970647	{'criterion': 'mse', 'splitter': 'best'}

(e)Best scores and the parameters for the 3 models.

According to the findings, a property's location is the most critical element in determining its price. Homes in desirable districts, such as South Mumbai, have higher prices than properties in outlying areas. This finding emphasizes the value of location in real estate, as well as the premium placed on properties in desirable locations. Hence, if one assigned ranked attribute values to factors (values of importance), prime location would be on the upper range of values.

The data also indicate that the economy and government policies have a major impact on property and rent prices. Passing through several policies played a role in the end result.

The type of property also plays a significant role in its price. The results indicate that apartments have higher prices than independent houses, which highlights the demand for apartment living in Mumbai city. The size of the property is also a factor, with larger properties commanding higher prices than smaller ones.

Furnished houses are sometimes more expensive to rent than apartments. As a result, a powerful combination of variables, such as the one mentioned above, leads to higher costs.

According to the observations, as the size of the house/flat grows, so does the price. Furthermore, as other characteristics change, the value of the house fluctuates. A more developed location, for example, would have higher pricing than a less developed region. If the number of bedrooms grows, so does the price per square foot as well as the price increases. A house on the market for resale will have a lesser value than one that is brand new. As a result, these were the observations made.

The findings also suggest that a property's age has a significant impact on its price, with newer residences commanding more expensive prices than older properties. Pricing for properties near vital amenities is higher than properties further away. It was also discovered that the charges for places with gymnasiums, club houses, and pools were justifiably high. This was despite the fact that maintenance was expensive because it appeared to be an all-inclusive package.The study indicates that the economy and government policies play a crucial role in determining house and rent prices. The results indicate that house and rent prices are positively correlated with the economy's state and government policies that support the real estate industry.

VI CONCLUSION

In this proposed study, different machine learning methods for predicting house prices as well as rent prices in various parts of Mumbai are compared. Making the best choice when buying a home requires using price prediction. The major goal of employing this prediction and suggestion system is to streamline the process and

reduce the amount of time spent on physical calculations by humans. It helps people to buy houses on a budget and reduce loss of money. The prediction of house prices is an important and difficult topic that has recently gained a lot of attention from researchers and practitioners. There have been several systems offered for predicting housing prices, ranging from traditional statistical methods to more modern machine learning techniques. One of the most difficult difficulties in house price prediction is the selection and engineering of features. Different factors have varying roles in predicting property prices, and selecting the right combination of features is crucial for making good predictions.

VII FUTURE WORK

Future study is to add maps to the website which will give proper guidance about the location of the house. We will make sure to provide graphs comparing the prices of different houses as well as the prices of a compared with its future and past prices. Accuracy in predicting can also be increased in the later stages

VII REFERENCES

[1]Nor Hamizah Zulkifley, Shuzlina Abdul Rahman, Nor Hasbiah Ubaidullah, Ismail Ibrahim, “House Price Prediction using a Machine Learning Model: A Survey of Literature”, Dec 8, 2020

[2] Anand G.Rawool,Dattatray V. Rogue,Sainath G.Rane , Dr. Vinayak Bharadi, “House Price Prediction Using Machine Learning”, June 2020.

[3] Ahmad Abdulal Nawar Aghi
House Price Prediction, Mar 2020

[4] Prof. Pradnya Patil, Darshil Shah, Harshad Rajput, Jay Chheda, “House Price Prediction Using Machine Learning and RPA”, July 2019

[5] G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu, “House Price Prediction Using Machine Learning”, Feb 2020

[6] Arshiya Shaikh, R.Vinayaki, G. Siddhanth, Y. Phanindra Varma,
“House Price Prediction using Multivariate Analysis”, June 2021

[7] Abhishek Jadhav, Nikita Bharambe, Sharayu Saraf, Vedant Pathak, Prof. Vishal Chaudhary,
“House Price Prediction And Recommendation System”, June 2020

[8] Shashi Bhushan Jha, Radu F.Babiceanu, Vijay Pandey, Rajesh Kumar Jha, ”Housing Market Prediction Problem using Different Machine Learning Algorithms: A Case Study”, May 2020

[9] Alisha Kuvalekar ,Shivani Manchewar,Sidhika Mahadik,Shila Jawale, “House Price Forecasting Using Machine Learning”,3rd April 2020

[10] Bindu Sivasankar, Arun P. Ashok , Gouri Madhu , Fousiya S.,”
House Price Prediction” (Volume 8, Issue 7), 7th July 2020.