



PROPERTY PRICE PREDICTION SOLUTION USING MACHINE LEARNING

¹Vaibhavi Avsarmal, ²Snehanjali Karande, ³Sanhita Sawant, ⁴Kavita Bani

^{1,2,3}Student, Department of Electronics Engineering, Atharva College of Engineering, Mumbai, India

⁴Assistant Professor, Department of Electronics Engineering, Atharva College of Engineering, Mumbai, India

Abstract: When dealing with real estate, whether you buy, sell or invest in real estate, you need a physical investigation. However, in this sudden pandemic situation, government regulations prevented the entire process mentioned above to be carried out. In our project, we want to help people who are planning to buy or sell a home or a particular property by informing them about the price range and safely planning the finances of the same home. In addition to forecasting real estate prices, it is also useful for real estate investors to know the trends in real estate prices at specific locations. Predict accurate prices based on statistical analysis comparisons and comparative market analysis with other similar properties, taking into account preliminary factors such as sale price, area quality, market, average property age, population area, address, etc. increase. This model uses machine learning techniques to curate them into ML models that can serve users. The buyer's main goal is to find a dream home with all the necessary equipment. Also, they look for these homes / properties with prices in mind, and there is no guarantee that they will get a product that is reasonably priced and not overpriced. Similarly, the seller is looking for a specific number to attach to the property as a price tag. Not only is this an exaggerated guess, but a lot of research is needed to complete the home assessment. This system helps you find the starting price of a property based on geographic variables. Future costs are expected by analyzing past market patterns and range of values, as well as future advances. The first thing that comes to mind when looking for a property is to contact various realtors. The problem with this is that the agent has to pay a small portion of the amount just to find a home and set a price for you. This model avoids such dangers and predicts accurate values. The main motivation for this model is to incorporate these machine learning techniques and curate them into an ML model that can serve users.

Index Terms - Real Estate, Machine Learning, Cost prediction.

I. INTRODUCTION

Our project topic is named "Property Price Detection" and we decided to take this topic for our final year project because of the relevance this project stands in today's time which faced a sudden unpredicted pandemic where no nation was ready to fight it economically and health sector wise. This pandemic brought down prices of few businesses while also saw a splurge in the prices of products because of low production and high demand for the products. There was a huge economic imbalance faced by the nations because of lockdowns and halt created in the whole working industries. All the sectors dipped into losses and there were alternative methods for working from home that became the new normal. One such sector which like all other sectors suffered from uncertainty and downfall was "Real Estate Industries". It faced a lot of challenges and is still facing few to get back on track like before the pandemic hit us. Our group decided to make a project keeping in mind this sector of "Real Estate Industries" and to make an easy format for predicting prices of the property sitting at home and to provide a proper communication and price evaluation to the client with predicted rates. This will help the clients to visualize the prospects that property holds for them along with the profits they would earn from these investments with correct accurate predictions. All this can be done by using machine learning software without actually visiting the site of property in this lockdown situation.

The goal of this project is to create a regression model that is able to accurately estimate the price of a property given the features across different markets. The objective of this analytical program is to help us understand the relationship between property features and how these variables are used to predict price.

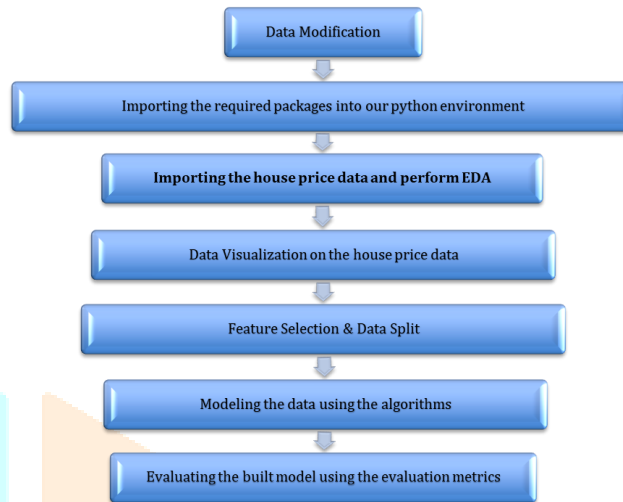
II. LITERATURE SURVEY

Sr no.	Title	Publication	Author	Description	Year
1	Real Estate Price Prediction	IJERT	Smith Dabreo , Shaleel Rodrigues , Valiant Rodrigues, Parshvi Shah	In this paper, the author has proposed solutions that demonstrate the usage of machine learning algorithms in the prediction of real estate/house prices on two real datasets.	2021
2	House price prediction	International Journal of Computer Sciences and Engineering	Bindu Sivasankar, Arun P. Ashok , Gouri Madhu , Fousiya S	This paper scrutinizes machine learning applications for the house price predictions model. The results of the work in this paper have shown how the house prices are predicted using Google Colab / Jupiter IDE. Here, different algorithms are being used and the top three of them are finalized and integrated to get the output. This paper shows how different algorithms are integrated to get a distinct and more accurate and clear answer from the given data.	2020
3	House Price Prediction Using Machine Learning	Iconic Research of Engineering Journals	ANAND G. RAWOOL , DATTATRA Y V. ROGYE , SAINATH G. RANE, DR. VINAYK A. BHARADI	This research paper highlights how house price prediction model is used for prediction of house using ML algorithm like Linear Regression , Decision Tree Regression They have compared the trends in housing cost and the current economic situations and the concerns of the buyers and the sellers.	2021
4	Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along With Joint Self-Attention Mechanism	IEEE	PEI-YING WANG, CHIAO-TING CHEN, JAIN-WUN SU, TING-YUN WANG, AND SZU-HAO HUANG	In this research paper we were enlightened about how house price prediction is one of the most relevant and popular topic and how there are numerous studies going on with the help of machine learning models or deep learning. In this paper they have explained the house buying process in depth and how the model helps to evaluate the price of the property. The research paper also highlights the machine learning proposed methodologies by explaining the important elements and features.	2021

III. PROBLEM STATEMENT

A physical examination is required when dealing with real estate, whether it is a real estate sale or investment issue. However, in this sudden pandemic situation, government regulations and security reasons prevented the entire process above from being carried out. That is the purpose of our project, to enable people planning to buy or sell a home or a particular property to know the price range and plan their financing safely at home. In addition, real estate price forecasts help real estate investors know what real estate price trends are in a particular location. An algorithm that takes into account preliminary factors such as sale price, regional quality, market, average real estate age, demographic region, address, etc. and enables accurate price predictions based on statistical analytical comparisons with other similar real estate. Create and compare one market analysis.

IV. PROPOSED SOLUTION



Usage of Python to get a visual understanding of the data.

ML application using Regression model and regression evaluation metrics.

Predict the sale price for each house.

Minimize the difference between predicted and actual rating.

To perform data visualization techniques to understand the insight of the data.

Mapping the real world problem to a Machine Learning Problem:

This problem involves predicting the prices of the houses which are continuous and real valued outputs. Thus, this is a Regression Problem.

Minimize RMSE and to provide some interpretation.

V. METHODOLOGY

Sr. No	Steps	Working
1.	Data Modification	Remove Duplicate data and check for null values
2.	Importing required packages	Pandas, Numpy, matplotlib, seaborn, scikit-learn
3.	Perform Exploratory data analysis	Get an overview of data, remove null values, get statistical analysis, convert float and object type variable to integer type
4.	Data Visualization	Used heatmap function in seaborn
5.	Feature selection and data spit	Define variables, Split the data into test and train set
6.	Modeling the data using Algorithm	Build and train different regression models using pre-built algorithm in scikit learn package
7.	Evaluating built model using evaluation metrics	Used data visualization to see the result of each candidate model, evaluate models using functions provided by scikit-learn package

First, we need to remove the duplicates from the data and check for any missing or null values. It is important to check these values as they cause errors in the machine learning model. There are many categorical variables that are labeled N / A in the absence of home functionality. These N / As can cause machine learning errors

The main packages for this project are Pandas for data processing, NumPy for manipulating arrays, Matplotlib and Seaborn for data visualization, and finally. Scikit-learn for building and evaluating ML models.

Exploratory Data Analysis (EDA) is an approach for analyzing datasets to summarize key characteristics, often using visual methods. It is important to examine the data before further analysis. This process gets an overview of the insights from the dataset with various functions and information about the house. After importing the data, we used the "head" function to check the dataset. Start the EDA process by removing all null values. Use the descriptive function to get a statistical view of the data such as mean, median, and so on. The final step in the EDA process is to check the data type of the variable that exists in the variable. If you have a variable of type float or object, you need to convert it to an integer type. This completes the EDA process. The next process is to visualize the data using the matplotlib and seaborn packages.

In our project, we used heatmaps to visualize the data. Heatmaps are very helpful in finding relationships between two variables in a dataset. Heatmaps can be easily created using the 'heatmap' function provided by Python's Seaborn package.

This process defines independent and dependent variables. After defining the variables, use them to divide the data into train sets and test sets. Data splitting can be done in Python using the train_test_split function provided by Scikit-learn. Use the training data to "teach" the model. The test dataset consists of as many features as the training data. The test dataset is trying to predict the selling price, so it excludes the selling price.

This process requires you to create and train six types of regression models: XGBOOST regression model, ridge regression model, lasso regression model, random forest regression model, linear regression model, and support vector regression model. We used off-the-shelf algorithms provided by Python's Scikit-learn package. The process is the same for all models. First define the variables that store the model algorithm, then fit the train set variables to the model, and finally make some predictions in the test set. You need to create five different linear regression models using the algorithms provided by scikit-learn. To find out which models are better suited for your data, you can use evaluation metrics to evaluate each model and draw conclusions. In machine learning, training refers to the process of teaching a model using an example training dataset. This is the final step in the process. We used data visualization to see the results for each candidate model. Our performance metric is mean square error. R^2 is a measure of how well a dependent variable can explain the variance of an independent variable. This is the most common scoring metric for regression models. To evaluate the model, you need to use the function provided by Python's Scikit-learn package. We then compared the metric values for each model and selected the one that was more suitable for a particular dataset.

VI. CONCLUSION

In our project, we want to contribute in the form of helping people survive the rapid changes that pandemics have brought to the functioning of the real estate industry. Based on various problems related to the real estate industry, we thoroughly investigated and tried to solve the problems that the real estate industry had during the pandemic. We have helped users in the field with the Real Estate Price Forecast project to predict and estimate real estate prices by planning the real estate finances and future plans they are considering. We developed algorithms and used machine learning mechanisms to realize our projects and help our users. Real estate prices were determined by calculating the selling price in the area. Therefore, real estate price forecasting models are very important for bridging information gaps and improving real estate efficiency. Home Price Forecast helps people planning to buy a home by estimating future price ranges and helping to raise money. In addition, real estate price forecasts help real estate investors know how real estate prices are moving in a particular location.

VII. ACKNOWLEDGMENT

We owe sincere thanks to our college Atharva College of Engineering for giving us a platform to prepare a project on the topic "Property Price Prediction" and providing us with the necessary resources for the project. We would also like to thank our Principal Dr. Shrikant Kallurkar for instigating within us the need for this research and giving us the opportunities and time to conduct and present research on the topic. We are sincerely grateful for having Dr. Bhavin Shah as our guide and Head of Electronics Engineering Department, and our project coordinators Prof. Kavita Bani. Research would have seemed difficult without their motivation, constant support and valuable suggestions. Working on this project also helped us to do lots of research and we came to know about many new aspects in the field of Data Science. We are immensely grateful to all involved in this project, who helped us throughout as without their inspiration and valuable suggestions it would not have been possible to develop this project within the prescribed time.

VIII. REFERENCES

- [1] 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), 1–24. <https://doi.org/10.3390/app8112321> [Crossref], [Google Scholar]
- [2] Mark, A.S., & John, W.B. Estimating Price Paths for Residential Real Estate. *Journal of Real Estate Research*; 2003: 25, 277–300.
- [3] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324> [Crossref], [Web of Science ®], [Google Scholar]
- [4] Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd international conference on Machine learning*. ACM. [Crossref], [Google Scholar]