International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST229242

House Pricing Prediction using ML Algorithm - A Comparative Analysis

Prof. Pushpa, Shahib Khan IR, Qamar Aziz, Tarique Anwar, Mohammed Arfath
Department of Computer Science, HKBK College of Engineering, Bangalore, India

Article Info
Volume 9, Issue 2
Page Number : 262-266
Publication Issue
March-April-2022
Article History
Accepted : 01 April 2022
Published : 09 April 2022

ABSTRACT

The real estate industry is the least transparent in our environment. Housing prices fluctuate on a daily basis and are sometimes inflated rather than based on valuation. Since the housing industry is rapidly expanding, forecasting house prices is critical not just for businesses but also for individuals. However, there are other factors that influence house price variations. The purpose of this article is to forecast a real estate property's market value. The goal here is to build a prediction model for evaluating pricing based on characteristics that influence price. We are going to predict data using Machine learning algorithm i.e. linear regression model to predict the house pricing.

Keywords - House Pricing, Market Value, Linear Regression

I. INTRODUCTION

The real estate market is one of the most price-sensitive in the world, and it is always changing. It is one of the most important domains in which machine learning techniques may be applied to improve and predict costs with great accuracy. The size of the property, its location, and its amenities are all important considerations that can influence the price. Our main goal is to create a model that estimates a customer's property cost based on his or her preferences. Our methodology examines a set of parameters chosen by the customer in order to determine the best pricing for their needs and interests. For prediction, it employs traditional techniques such as linear regression, forest regression, and boosted regression, and attempts to provide an interpretation of the data produced. After the trained model is complete, Flask (a Python framework) or Django are used to connect it with the user interface.

II. LITERATURE SURVEY

1. House Price Prediction Using Machine Learning and Neural Networks

Our dataset contains a number of critical factors, and data mining is at the heart of our system. We first cleaned up the full dataset and trimmed the numbers that were outliers. Furthermore, we weighted each parameter based on its relevance in determining the system's price, resulting in an increase in the value that each parameter retains in the system. We chose three distinct machine learning algorithms and tested our system with various combinations to provide the most reliable results possible [1].

Data is at the center of technological advancements, and predictive models may now achieve any result. This method makes considerable use of machine learning. Computer learning is supplying a reliable dataset and then making predictions based on it. The machine learns how important a given event is to the overall system based on its pre-loaded data and predicts the outcome appropriately. Anticipating stock prices, predicting the possibility of an earthquake, predicting corporate sales are only a few examples of recent applications of this approach [4].

2. House Price Forecasting Using Machine Learning

A. Varma devised a method for obtaining exact real-world assessments utilising Google maps and real-time neighbourhood data. Researchers have discovered links between a city's visual look and non-visual characteristics such as crime statistics, property costs, population density, and so on. "City Forensics: Using Visual Elements to Estimate Non-Visual City Elements," for example, uses visual attributes to predict the property's sale price. [2]. Classification and regression techniques were employed by Hujia Yu and Jiafu Wu. According to the findings, the living space square feet, roof content, and neighbourhood have the most statistical significance in determining a home's selling price. The PCA approach can also help with prediction analysis [3].

3. Housing prices prediction with deep learning: an application for the real estate market in Taiwan

The fundamental challenge in predicting house prices is that the real estate market is impacted by a variety of factors, including macroeconomic and market value. The home is both an investment and a consumer item, and the goal of the investment or consumption decides whether the property will be utilised for living or for renting [5].

Deep learning algorithms show considerable benefits in time series prediction in all of the research covered here. There are still deep learning algorithms for predicting house values that aren't widely used. As a result, this work uses deep learning algorithms (BPNN, CNN) to forecast house prices using a time series dataset.

4. Machine Learning based Predicting House Prices using Regression Techniques

Pow, Nissan, Emil Janulewicz, and L. Liu employed four regression approaches to forecast the property's pricing value: Linear Regression, Support Vector Machine, K-Nearest Neighbors (KNN), and Random Forest Regression, as well as an ensemble approach combining KNN and Random Forest Technique. The prices were predicted with the least error of 0.0985 using the ensemble technique, while PCA did not improve the prediction error [6]. Several research have also looked at how to collect characteristics and how to extract them. [8] Wu, Jiao Yang examined a variety of feature selection and extraction techniques using Support Vector Regression. To anticipate property values, several academics have constructed neural network models. To anticipate property values, Limsombunchai contrasted hedonic pricing structure with artificial neural network model [7].

III. EXPERIMENTAL SETUP AND METHODS

The simplest basic approach for prediction is linear regression. It employs two variables as variables: the predictor variable and the one that is the most important of the two, the predictor variable. These regression estimations are used to describe how one dependent variable and one or more independent variables are related. The formula is used to describe the regression equation with one dependent and one independent variable [9]. **b = y + x*a**, where b represents the estimated dependent variable score, y represents the constant, x represents the regression coefficient, and a represents the score on the independent variable.

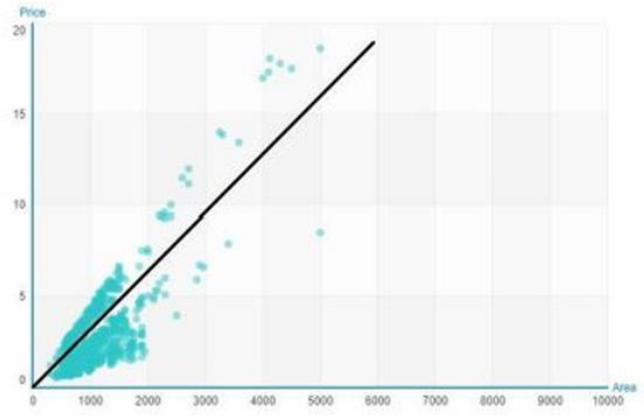

Figure 1: Linear regression scatter plot

Bagging of trees is a method used in forest regression. The main goal is to arrange the various trees in a pleasing manner. The Variance in the Trees is then reduced by averaging them. A significant number of decision trees are constructed using this method [10]. The random forest training algorithm uses the bootstrap aggregating, or bagging, strategy to train tree learners [11].

Collection of Data

There are several data processing techniques and procedures to choose from. We gathered information on real estate properties from a variety of websites. Location, carpet area, built-up area, property age, zip code, and other factors would be included in the data. We need to collect organised and classified quantitative data. Before doing any machine learning study, it is necessary to collect data. Validity of the dataset is essential; otherwise, there is no use in evaluating the data.

	MedInc	HouseAge	AveRooms	AveBedrms	Population	AveOccup	Latitude	Longitude	Target
0	8.3252	41.0	6.984127	1.023810	322.0	2.555556	37.88	-122.23	4.526
1	8.3014	21.0	6.238137	0.971880	2401.0	2.109842	37.86	-122.22	3.585
2	7.2574	52.0	8.288136	1.073446	496.0	2.802260	37.85	-122.24	3.521
3	5.6431	52.0	5.817352	1.073059	558.0	2.547945	37.85	-122.25	3.413
4	3.8462	52.0	6.281853	1.081081	565.0	2.181467	37.85	-122.25	3.422

Data Preprocessing

The process of cleansing our data collection is known as data preparation. The dataset may contain missing values or outliers. Data cleansing can take care of these issues. If a variable has a lot of missing values, we'll remove them or replace them with the average value.

Training the model

We must first train the model since the data is divided into two modules: a Training set and a Test set. The target variable is included in the training set. The training data set is subjected to the decision tree regressor method. The decision tree uses a tree structure to create a regression model.

Testing & Integrating with UI

The trained model is used to forecast property prices on a test dataset. The front end is then connected with the trained model using Python's Flask framework.

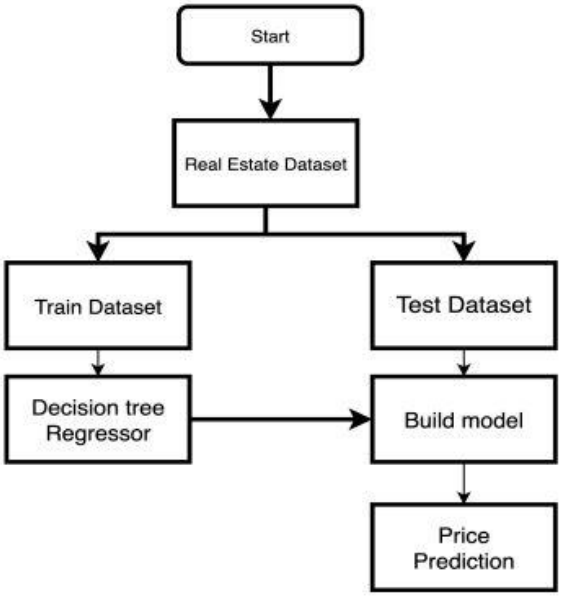

Fig 1. The generic flow of development

IV. Conclusion

A technique has been created that seeks to deliver an accurate prediction of house prices. Linear Regression, Forest Regression, and Boosted Regression are all used to their full potential by the system. The introduction of neural networks has improved the algorithm's efficiency even further. Customers will be satisfied since the method will provide precise results and eliminate the danger of buying in the wrong residence. We are also incorporating a new concept in this project. The House Bidding Management System

will be an online platform for purchasing modern apartments, houses, and offices. Buyers no longer go from property to property looking for a deal. In this age of computer science, everyone expects everything to be delivered to their doorstep.

Additional customer-beneficial features can be added to the system without interfering with its primary operation such as adding different file types to CMS to accept as a file upload from Customer (seller). The inclusion of larger cities to the database might be a key future upgrade, allowing our users to look at more residences, get greater accuracy, and so make better decisions.

V. Future Work

The system's accuracy may be enhanced. If the system's size and processing capacity grow, it will be possible to add several additional cities. Furthermore, we may use Augmented Reality to incorporate several UI/UX methodologies for a better representation of the results in a more interactive approach. In addition, a learning system may be developed that collects user feedback and history so that the system can present the most appropriate results to the user based on his preferences.

VI. Acknowledgment

A special and an earnest word of thanks to the project guide Prof. Pushpa for their constant assistance, support, patience, endurance and constructive suggestions for the advancement of the project.

VII. REFERENCES

[1]. Ayush Varma, Abhijit Sarma, Sagar Doshi, and Rohono Nair. "House Price Prediction using Machine Learning and Neural Network" 2018 International Conference on on Inventive Communication and Computational Technology (ICICCT). IEEE, 2018.

[2]. Arietta, Sean M., et al. "City forensics: Using visual elements to predict non-visual city attributes." IEEE transactions on visualization and computer graphics 20.12 (2019): 2624-2633.

[3]. Yu, H., and J. Wu. "Real estate price prediction with regression and classification CS 229 Autumn 2016 Project Final Report 1–5." (2017).

[4]. R. J. Shiller, “Understanding recent trends in house prices and home ownership,” National Bureau of Economic Research, Working Paper 13553, Oct. 2020. DOI: 10.3386/w13553.

[5]. Choujun Zhan, Zeqiong Wu, Yonglin Liu, Zefeng Xie, Wanglin Chen. “Housing prices prediction with deep learning: an application for the real estate market in Taiwan” 2020 IEEE 18th International Conference on Industrail Informatics (INDIN) IEEE 2020.

[6]. Pow, Nissan, Emil Janulewicz, and L. Liu (2014). Applied Machine Learning Project 4 Prediction of real estate property prices in Montréal.

[7]. Manasa J, Radha Gupta, Narahari N S. “Machine Learning based Predicting House Prices using Regression Techniques” 2020 Proceedings of the Second International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2020) IEEE Xplore

[8]. Wu, Jiao Yang(2017). Housing Price prediction Using Support Vector Regression.

[9]. C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006

[10]. T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, “How many trees in a random forest?” In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 7376 LNAI, 2019, pp. 154–168, ISBN: 9783642315367. DOI: 10 . 1007 / 978 - 3 - 642 - 31537-4\ 13

[11]. The elements of statistical learning, Trevor Hastie - Random Forest Generation

Cite this article as :

Prof. Pushpa, Shahib Khan IR, Qamar Aziz, Tarique Anwar, Mohammed Arfath , "House Pricing Prediction using ML Algorithm - A Comparative Analysis", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 2, pp. 262-266, March-April 2022. Available at doi : https://doi.org/10.32628/IJSRST229242
Journal URL : https://ijsrst.com/IJSRST229242