House Price Prediction System Using Machine Learning

[1]Prof. Gayatri Naik,[2]Master.Dhruv Patil, [3]Master.Rohit Kore, [4]Miss.Priti Ingale

[1]Asst.Professor,[2,3,4]UG Student,[1,2,3,4]Computer Engg. Dept. Shivajirao S. Jondhle College of Engineering & Technology, Asangaon, Maharshatra, India. [1]krishngita123@gmail.com, [2]inbox.dhruvpatil@gmail.com,[3]rohitkore258@gmail.com, [4]pritiingale400@gmail.com

Abstract- With the help of Python modules, this paper gives an overview of how to predict house expenses using various regression models. The proposed technique took into accountthe more refined components of the property price calculation and offered a more accurate projection. It also goes over the many graphical and numerical techniques that will be used to predict the price of a home. This paper explains what a machine learning-based house price model is and how it works, as well as which dataset is used in our proposed mode It is one of the most important areas where machine learning techniques may be used to better and accurately anticipate expenses. The purpose of this article is to forecast a real estate property's market value. Based on geographical criteria, this approach assists in determining a beginning price for a property.

Keywords- Regression

I. INTRODUCTION

Houses/homes are a basic human need, and their prices vary according to the amenities provided, such as parking space, location, and so on. House pricing is a source of concern for many inhabitants, whether wealthy or low-income, because the value ofa property can never be determined by its size or closeness to offices. Purchasing a home is one of a family's most important and essential decisions, as it consumes all of their investment earnings and, in some cases, covers them with loans. Predicting accurate house pricing values is a difficult task. The proposed model would allow for the precise prediction of housing values. Medical care issues have become a major topic in today's culture, with issues such as an unbalanced and insufficient allocation of medical resources becoming increasingly obvious. In this context, the use of machine learning has become an unavoidable trend inthe current evolution of medical care of the a The model can help us to find the price of an intelligence (ANN) algorithms were first used to measure abdominal discomfort by scientists at the University of Leeds at United Kingdom in 1972. ML-based pathological diagnosis of tumors, lung cancer, and other diseases has gradually gained traction. Alibaba, Amazon, and Baidu, for example, have developed their own research teams to work on it. The use of ML in medical care has significantly reduced medical costs while also providing a new way for citizens to see a doctor and making people's lives easier. People's demand, However, it provides a new impetus for ML research and development, as well as support for its continuous improvement. Buying a house is one of a family's most important and essential purchases because it consumes all their investment funds and, on sometimes, covers them with loans. Predicting accurate house pricing values is a difficult task. The project helps us with accurately predicting the price.

II. AIMS AND OBJECTIVE

a) Aim

The project's goal is to produce an effective house price prediction model, validate the model's accuracy, and identify key home price parameters that feed the model's predictive potential.

b) Objective

The purpose of the project is to use Python machine learning to make proper and accurate house price projections. As a result, this paradigm can aid customer openness while also facilitating comparisons. If a customer discovers that a house's price on a certain website is higher than the model's prediction, he may reject the residence .

III. LITERATURE SURVEY

Paper 1: **Analyzing the factors affecting mortality rates making use of Decision tree in R Language**

This is a study for a social cause, the Mortality Rate, which was rising at a critical rate and raising concern among the world's citizens. This situation was investigated by looking at various factors such as birth rate, literacy rate, number of health centers, and so on, using the R tool's decision tree technique, which displayed trees from two decades and analyzed the factors affecting mortality rates and their contribution to driving the rate. As a result, the governing authorities will be able to learn more about the concerns and work to minimize infant mortality rates.

 DOI : 10.35291/2454-9150.2022.0160

Paper 2 : International journal of housing and markets.

On a local and international basis, housing remains the most popular land use. In recent years, housing has grown in popularity as a means of investment. for both private and institutional investors in a number of countries, needing a higher level of study. The purpose of the Worldwide Journal of Housing Markets and Analysis is to provide a forum for the sharing of information and ideas about housing, housing markets, and their interactions on a global scale.

Paper 3 : House Price Prediction using Hedonic model vs Artificial NeuralNetwork

The goal of this research is to see how well a hedonic model and an artificial neural network model predict housing prices. The size, age, and kind of the house, also the number of bedrooms, bathrooms, garages, neighboring amenities, and geographic position, are all considered. The model helpsto plan financially as well.

Paper 4 : Lung cancer prediction using ML and advanced imaging techniques

To assist doctors in the management of ambiguous pulmonary nodules identified by accident or on a screen, machine learning- based lung cancer prediction algorithms have been developed. These technologies may be able to reduce variability in nodule categorization, enhance decision-making, and thereby decrease the number of benign nodules that are monitored or worked up unnecessarily. This page summarizes the major lung cancer prediction algorithms thathave been proposed so far, as well as some of their merits and weaknesses. It helps people to identify and expect the disease and provide help in the medical industry.

Paper 5 : A Tool for CME Arrival Time Prediction Using ML Algorithms: CAT-PUMA

CMEs (coronal mass ejections) are the most powerful eruptions in the Solar System. CMEs can cause major disruptions in interplanetary space and potentially affect human activities in a variety of ways, including infrastructure damage and economic losses. We present a new method for predicting partial/full-halo CME arrival times using ML algorithms in this paper (CAT-PUMA).

IV. EXISTINGSYSTEM

For housing price prediction, the existing system employed xgboost. Based on housing price and characteristics data in Beijing, China, the objective of this work is to analyses the essential explanatory features and design an accurate meodel to execute Housing price forecasting by location in Beijing. The findings show that, although being more time consuming, machine learning methods improve estimation accuracy significantly when compared to classic hedonic methods. Furthermore, the xgboost model is found to be the least accurate in describing and predicting the geographical dynamics of housing price in Beijing. The system revolves around theexpense and sale .

V. COMPARTIVE STUDY

SR NO.	PAPER TITLE	AUTHOR NAME	METHOD	ADVANTAGE	DISADVANTAGE
1.	Analyzing the FactorsAffecting Infant Mortality Rate UsingDecision Tree in R language	Parul Kalra,Deepti Mehrotra	Decision Tree inR Language	Good ApproachExplained	Time Consuming
2.	International Journal ofHousing Markets and Analysis (Int J House Market Anal)	RJ Bolton,DJ Hand	Deep Learning, Stereo Geometry	Good ApproachExplained	Difficult to understand
3.	House Price Prediction: Hedonic Price Model vs Artificial Neural network	Zhang,Xinwei;abHan,Yaocia,WeiXu,WangQilia	XGBoost	Good ApproachExplained	Time Consuming
4.	Lung cancer predictionusing ML and advanced imaging	Timor Kadir,Fergus Gleeson	Region-orientedstrategy	Best ApproachExplained	Little Bit time Consuming
5.	A new tool for CME arrival time predictionusing ML algorithms: CAT- PUMA	Jiajia Liu, Yudong ye, Yuming wang, Shenlong shen	CME , CAT-PUMA	Good ApproachExplained	Difficult to understand

 DOI : 10.35291/2454-9150.2022.0160

VI. PROBLEM STATEMENT

This statistical analysis' purpose is to assist us understand the relationship between house attributes and how the factors are utilized to forecast property price. You must manually build dummy variable/label encoding for categorical features before feeding them into the model when using two separate models to increase prediction rate and enhance accuracy in xgboost. You only need to define category feature names or indexes for Catboost to do it on its own. For larger datasets, the training period is fairly long. Furthermore, xgboost is determined to be the least accurate model in explaining the peculiar dynamics of property prices in Beijing.

VII. PROPOSED SYSTEM

The method provided is based on linear regression. The purpose of this research is to better predict property prices and deliver more precise results. The information for the house prediction is gathered from public ally accessible sources. Validation training is performed on 50% of the dataset, with the remaining 50% being used for testing purposes. This method divides the datasetinto several subsets. It had been attempted toprepare on all of the subsets at that moment; nevertheless, one (k-1) subset was left for the assessment of the prepared model. This technique emphasizes k times, with a different subset rotated each time for the preparation reason. Furthermore, xgboost is determined to be the least accurate model in explaining the peculiar dynamics of property prices in Beijing.The findings show that, although being more time consuming, machine learning methods improveestimation accuracy significantly when compared to classic hedonic methods. For larger datasets, the training period is somewhat long. Furthermore, the xgboost model is determined to be the least accurate in explaining the peculiar dynamic project.

VIII. ALGORITHM

The general idea of working of proposedsystem algorithm is given as follow:

Step.1: Start

Step.2: Get valid input from user

Step.3: Collect the dataset

```
class csvdatamodel(models.Model):longitude =
models.CharField(max_length=50) latitude = models.EmailField()
housing_median_age = models.CharField(max_length=40) total_rooms
= models.CharField(max_length=40) total_bedrooms =
models.CharField(max_length=50, default="", editable=True)
population = models.CharField(max_length=40, default="",
editable=True) households = models.CharField(max_length=40,
default="", editable=True) median_income =
models.CharField(max_length=40, default="", editable=True)
median_house_value = models.CharField(max_length=40, default="",
editable=True) ocean_proximity = models.CharField(max_length=40,
default="", editable=True)
```

Step.4: Perform Data Cleaning

Step.5: Train the model using XGBoost from sklearn.impute import SimpleImputertrain_ft = df.drop(['ocean_proximity', 'median_house_value'], axis=1)
imputer = SimpleImputer(strategy='median')imputer.fit(train_ft)

```
train_ft.median().values
x = imputer.transform(train_ft)train_new_set
    pd.DataFrame(x,columns=train_ft.colums
    )
train_new_set.head()
train_new_set.isna().sum()train_new_set.head() train_new_set.shape
train_new_set.info()

X = train_new_set.values
```

Y = df['median_house_value'] **Step 6**:Test and Run the model from sklearn.linear_model importLinearRegression
lr = LinearRegression()lr.fit(x_train, y_train)
predictions = lr.predict(x_test[:10])print("predictions:", predictions)
y_train[:10]
data = {'predicted': predictions, 'Actual':y_test[:10].values, 'Diff': (predictions - y_test[:10].values)}
error_df = pd.DataFrame(data=data)print("error diff:", error_df)
return render(request,"user/houseprediction.html",{ "errordiff":error_df})

Step.7: Display Stored Result

Step.8: Exit

IX. MATHEMATICAL MODEL

1. Simple Linear Regression

Simple linear regression employs a slope- intercept form, with a and b being the coefficients that we strive to "learn" in order to make the most accurate predictions. Our input data is represented by X, and our forecast is represented by Y.-

$$Y = bX + a$$

2. Multi variable Regression

The model will attempt to learn a more complex multi-variable linear equation, suchas this, where w stands for coefficients or weights. Our model will attempt to learn a more complex multi-variable linear equation, such as this, where w stands for coefficients weight and the height of it

$$Y(x_1, x_2, x_3) = w_1x_1 + w_2x_2 + w_3x_3 + w_0$$

The variables represent the attributes or distinct pieces of information, we have abouteach observation.

3. Mean Squared Error (MSE) CostFunction

The MSE is defined as:

$$MSE = J(W) = \frac{1}{m}\sum_{i=1}^{m}(y^{(i)} - h_w(x^{(i)}))^2$$

Where,

$$h_w(x) = g(w^T x)$$

The MSE is a metric that indicates how far the average model predictions differ from theactual values. When the model performs "badly" on our training data, the number is higher.

MSE's first derivative is calculated as follows:

$$MSE' = J'(W) = \frac{2}{m}\sum_{i=1}^{m}(h_w(x^{(i)}) - y^{(i)})$$

4. One Half Mean Squared Error (OHMSE)

We'll make a modest change to the MSE by multiplying it by 1/2 so that the 2s balance out when we compute the derivative:

$$OHMSE = J(W) = \frac{1}{2m}\sum_{i=1}^{m}(y^{(i)} - h_w(x^{(i)}))^2$$

X. ADVANTAGES

1) House price predictions are anticipated to assist customers who are planning to buy a home by allowing them to know the price range in the future so that they

 DOI : 10.35291/2454-9150.2022.0160

may properlyarrange their finances.

2) House price projections are also useful for property investors who want to know the trajectory of housing prices in a specific area.

3) House price predictions are anticipated to assist customers who are planning to buy a home by allowing them to know the price.

4) House price projections are also useful for property investors who want to know the trajectory of housing prices in a specific area

XI. SYSTEM ARCHITECTURE

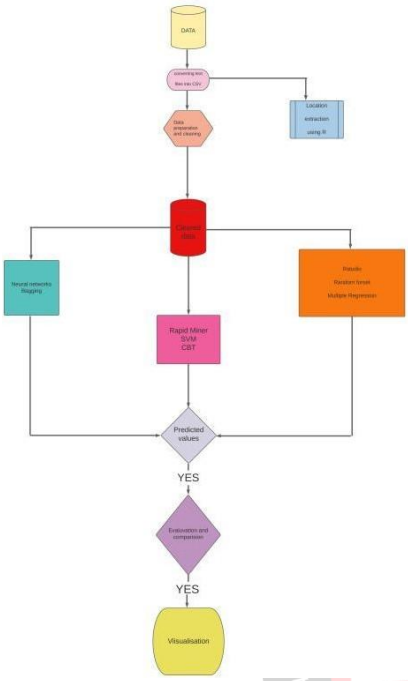

Fig.1: System Architecture

XII. DESIGN DETAILS

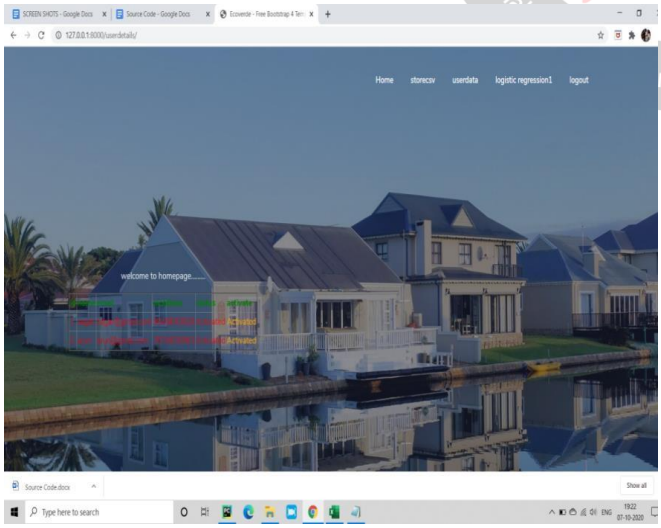

Fig 2: Result

XIII. CONCLUSION

Thus, we have tried to implement the paper "V. Limsombunchai, Christopher Gan, Minsoo Lee""House price prediction: Hedonic price model vs. artificial neural network" American Journal of Applied Sciences, 2004. Hence based on the customer demand, we have concluded that the model has forecasted the price of a house. The main objective of the project has been fulfilled that it will aid those who are looking to purchase a home in correctly planning their theorem finance. Furthermore, property investors who wish to know the trend of housing costs in a certain location might benefit from house price projections.

REFERENCES

[1] V. Limsombunchai, Christopher Gan,Minsoo Lee —House price prediction: Hedonic price model vs. artificial neural network,‖ American Journal of applied Sciences, 2004

[2] [2] R. A. Rahadi, S. K. Wiryono, D. P. Koesrindartotoor, and I.B. Syamwil,—Factors influencing the price of housing in Indonesia,‖ Int. J. Hous. Mark. Anal., vol. 8, no. 2, pp. 169–188, 2015

[3] Jain, N., Kalra, P., &Mehrotra, D. (2019). Analysis of Factors Affecting Infant Mortality Rate Using Decision Tree in RLanguage. In Soft Computing: Theories and Applications (pp. 639-646). Springer, Singapore.

[4] Kadir, T., & Gleeson, F. (2018). Lung cancer prediction using machine learning and advanced imaging techniques. Translational Lung Cancer Research, 7(3), 304-312.

[5] Liu, J., Ye, Y., Shen, C., Wang, Y., &Erdélyi, R. (2018). A New Tool for CME Arrival Time Prediction using Machine Learning Algorithms: CATPUMA. The Astrophysical Journal, 855(2), 109.

[6] Velankar, S., Valecha, S., &Maji, S. (2018, February). Bitcoin price prediction using machine learning.In Advanced Communication Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020) IEEE Xplore Part Number: CFP20V66-ART; ISBN: 978-1-7281-4108- 4 978-1-7281-4108-4/20/$31.00 ©2020

[7] Malhotra, R., & Sharma, A. (2018). Analyzing Machine Learning Techniques for Fault Prediction Using Web Applications. Journal of Information Processing Systems, 14(3).

[8] Choo, M. S., Uhmn, S., Kim, J. K., Han,J. H., Kim, D. H., Kim, J., & Lee, S. H. (2018). A Prediction Model Using Machine Learning Algorithm for Assessing Stone- Free Status after Single Session Shock Wave Lithotripsy to Treat Ureteral Stones. The Journal of urology.

[9] Nilashi, M., Ibrahim, O., Ahmadi, H., Shahmoradi, L., &Farahmand, M. (2018). A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocyberneticsand Biomedical Engineering, 38(1), 1-15.

[10] Fan, C., Cui, Z., &Zhong, X. (2018, February). House Prices Prediction with Machine Learning Algorithms.In Proceedings of the 2018 10th International Conference on Machine Learning and Computing (pp. 6-10).ACM.

[11] Zhou, J., Zhang, H., Gu, Y., &Pantelous, A. A. (2018). Affordable levels of house prices using fuzzy linear regressionanalysis: the case of Shanghai. Soft Computing, 1-12.

[12] Jang, H., Ahn, K., Kim, D., & Song, Y. (2018, June). Detection and Prediction of House Price Bubbles: Evidence from a New City. In International Conference on Computational Science(pp. 782-795). Springer, Cham.

[13] Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern recognition, 30(7), 1145- 1159.

[14] Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data.Expert Systems with Applications, 42(6), 2928-2934.

[15] Harrison, D., and D. L. Rubinfeld. 1978. "Hedonic Housing Prices and the Demand for Clean Air."J. Environ. Econ.Manag.5 (1): 81– 102.

 DOI : 10.35291/2454-9150.2022.0160