iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35831

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

House Price Prediction using Machine Learning Techniques

G. Gayathri Priya[1], Ch. Naga Akhil[2], E. Sai Mahija[3], G. Teja[4], Mrs. D. RagaVamsi[5]
[1, 2, 3, 4]*Student, CSE Department & GEC, Gudlavalleru, Krishna Dt.*
[5]*M.Tech, Assistant Professor, CSE Department & GEC, Gudlavalleru, Krishna Dt.*

Abstract: ***The real estate market is one of the most price-driven, but it is still affected by volatility. This is one of the main uses of machine learning ideas to improve and predict costs with high precision. As housing prices are fluctuating, People are cautious when trying to buy a new house based on their budget and marketing strategy. The purpose of the paper is to forecast consistent home prices for non-owners based on their financial dispositions and aspirations. The paper involves predictions using various Regression techniques like linear regression, random forest regression, polynomial regression, robust regression, lasso regression, elastic net regression, stochastic gradient descent, svm regression, artificial neural network. On a data set, house price prediction has been done by combining all of the above-mentioned strategies to determine which is the most effective. The purpose of the paper is to assist the seller in accurately estimating the selling price of a house. Physical circumstances, and location, among other things, were all taken into account while determining the cost.***

I. INTRODUCTION

Data is at the core of technological innovation, and can now use the predictive model to achieve any result. Machine learning is widely used in this method. Machine learning means that it provides a valid data set. In addition, the prediction is based on. The machine itself understands the importance of a specific event to the entire system based on its pre-loaded data and predicts the result accordingly. Develop machine learning algorithms, build models based on your data, and use them to predict new data. The main difference from existing algorithms is that the model is built on the input data as well as executing a series of commands. Map learning uses data with labelled results, while unsupervised learning uses unlabelled data. There are some common machine learning algorithms like regression, classification, neural networks, and deep running. Reinforcement learning and role-playing learning are widely used in deep learning. How to predict house prices using machine learning algorithms? It's hard to get as many results as possible depending on the model built. For a particular house value, it is determined by location, size, type of house, city, national tax cycle, population movement, interest rate and many other factors that can affect the supply and demand of.

Regression is a machine learning device that urges to set expectations by capturing current measurable information, associations between target parameters and various independent variables. According to this definition, the cost of a house depends on variables such as number of rooms, area of living area, area, etc. If you are unlikely to apply a forgery to understand how these parameters, you can calculate the home valuation of a specific land area of. Various modern applications of this technology have endless possibilities for stock price forecasting, earthquake potential forecasting, corporate sales forecasting, listing are limitless

.The main motivation for the project was to use the appropriate algorithm of to find the best for predicting prices with a low error rate, thus making the best possible prediction of house prices. The evaluation that has been performed on this paper is especially primarily based totally at the datasets of usa_housing. In this paper, we attempt to illustrate all of the possible Regression strategies that are appropriate to our problem.

II. RELATED WORK

The value of a particular property depends on the surrounding infrastructure. Recently, a range of writers have appeared looking for the best attributes for their clients along with various techniques Raghunandhan mentioned the basics of how data mining works and the supporting algorithms used for prediction. The most important part is which machine learning algorithm is best for predicting housing prices. Usually, the environmental conditions of the location determine our expected price types for different types of houses. Manjula [12] showed us several important characteristics when using regression models to predict real estate prices with high accuracy. Rapach In addition, Strauss (2007) uses the automatic decentralized regression (ARDL) model framework, and the conjecture with 25 determinants develops the real accommodation cost of to the unique state of the eighth reserved area selected by. They found that the ARDL model is often better than the reference AR model.

Based at the administrative information from the Netherlands, Joep et al locate that better earnings and wealth client ends in better buy price, whilst better earnings and wealth dealer ends in decrease promoting price.

When comparing the 79 variables provided by the competition Kaggle, we find that the information is incomplete, and some characteristics do not need to be predicted, and are implicit conditions for house prices. We try to find those hidden features between these features, then we transform these features into a normal distribution and then transform to increase linearity. We also explored the parameter fit regression algorithm.

III. SYSTEM DESIGN AND ARCHITECTURE

1) Phase 1: Collection of data

Data processing strategies and techniques are numerous. We collected data for kaggle website i.e. usa_housing data set. In this data set we have attributes like area income, area house age, number of rooms, no of beds room, area population, price, address. In this data set we have 7 columns and 5000 entries.

2) Phase 2: Data Preprocessing

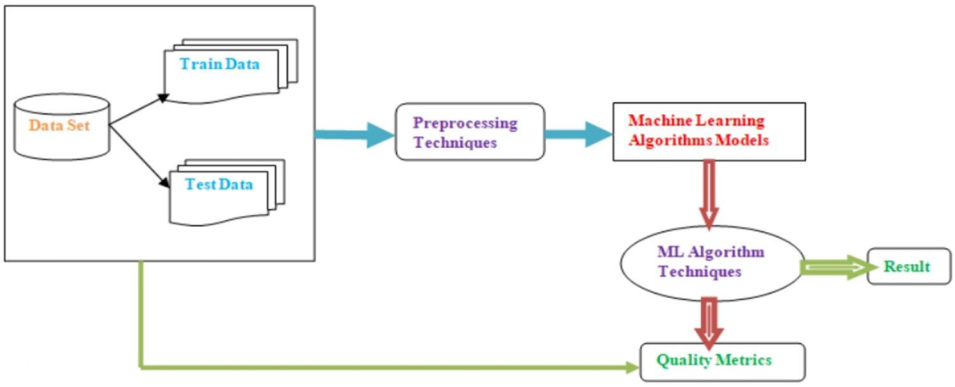
Data Preprocessing in Machine Learning is a data mining technique that transforms raw data into an understandable and readable format. The dataset may contain missing values or outliers. Data cleaning can take care of these issues. If a variable has a lot of missing values, we'll remove them or replace them with the average value.

3) Phase 3: 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.

4) Phase 4:Testing and Integrating with UI

The trained model is used to test the data set and predict housing prices. The educated version is then integrated with the front to cease the usage of Flask in python.



IV. METHODOLOGY

A. Random forest Regression

The random forest algorithm is made up of different decision trees. Each decision tree has the same node, but using different data results in different leaves. Merge the decisions from multiple decision trees to find an answer that represents the average of all these decision trees. The random forest algorithm has a supervised learning model; uses classification information to "learn" a method of classifying unlabeled information.

B. Robust Regression

Robust regression is a form of regression analysis designed to overcome some of the limitations of traditional parametric and nonparametric methods. Robust regression methods are designed not to be unduly affected by assumptions that the underlying data generation process violates. One instance in which robust estimation should be considered is when there is a strong suspicion of heteroscedasticity.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

C. Ridge Regression

Ridge regression is a way to create a parsimonious model when the number of predictor variables in a set exceeds the number of observations, or when a data set has multicollinearity (correlations between predictor variables).]Ridge Regression is a tool for analysis of [MR]multiple regression on the data that have multicollinearity (mcl). Multicollinearity(mcl) is the existence of near–linear relationships among the variables which are independent. [RR]Ridge regression applies a special type of condition on parameters as in Eq. 1 and Eq. 2 (ββ's): ^βridgeβ^ridge was chosen to reduce the error of sum of the squares.

∑a=1(ya−p∑b=1xabβb)2+λp∑b=1β2 (1) which is equivalent minimization as

∑ni=1(yi−∑pj=1xijβj)2∑i=1n(yi−∑j=1pxijβj)2 (2) subject to, for some c>0c>0,

∑pj=1β2j<c∑j=1pβj2<c, i.e. constraining the sum of the squared coefficients.

D. Lasso Regression

Lasso Regression which is similar to minimizing the sum of squares with constraint Σ |Bj≤ s Some of the βs are gotten smaller to precisely 0, ensuing in a regression version that's less difficult to interpret.

A tuning parameter, λ controls the energy of the L1 penalty. λ is largely the quantity of shrinkage:

When λ = 0, no parameters are eliminated. The estimate is the same as the one discovered with linear regression.

As λ increases, increasingly coefficients are set to 0 and eliminated (theoretically, while λ = ∞, all coefficients are eliminated).

As λ increases, bias increases.

As λ decreases, variance increases.

E. Elastic net Regression

This technique solves the problem of regularization. For α1 , which is strictly between the values 0 and 1, and the non-negative λ1, the elastic network [ER] solves the problem as Equation 3:

β ^ = argmin || qPβ|| 2 + λii || β || 2 + λi || β || 1 (3)

F. Stochastic gradient descent

Gradient descent is a very general optimization algorithm that can find the best solution for various problems. The general idea of Gradient Descent is to iteratively adjust the parameters to minimize the cost function. Gradient descent measures the local gradient of the error function with respect to the parameter vector and progresses in the direction of the gradient descent. Once the gradient is zero, it has reached the minimum.

G. Polynomial Regression

Polynomial regression is a special case of linear regression, in which we fit a polynomial equation to data that has a curvilinear relationship between the target variable and the independent variable. In polynomial regression, we have a polynomial equation of degree *n* represented as:

$$Y = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \ldots + \theta_n x^n \quad (4)$$

Here: $\theta 0$ is the bias, $\theta 1$, $\theta 2$, …, θn are the weights in the equation of the polynomial regression, and *n* is the degree of the polynomial.

H. Artificial Neural Network

The neural network is an artificial intelligence model of originally designed for to replicate the learning process of the human brain. Model is made up of three main layers: input data layer (attributes from example), hidden layer (usually called the "black box"), and output layer (estimated house price). The neural network is a network of interconnected artificial neurons. Its rule is to adjust the strength or weight of the connection between units according to externally provided data. ANN is like a nonlinear regression or multiple regression model, with unobservable linked variables. Once the network topology and parameters are specified, can be expressed as a common statistical or econometric model. Neural networks are used for different purposes, such as model estimation, classification, prediction, etc. It is used here as a modelling tool and a practical alternative to the famous hedonic econometric model. In addition, it aims to predict the value of real estate, so it can be used to measure the recession in the real estate market in recent years.

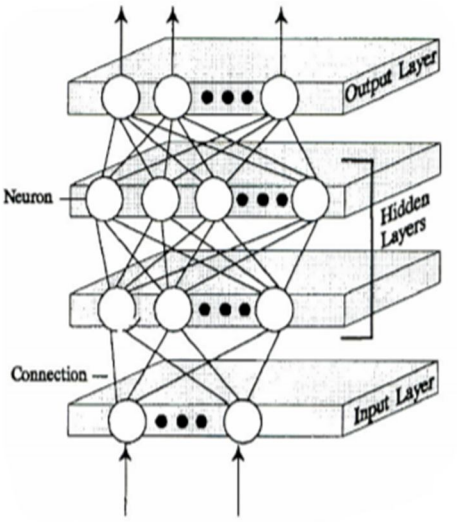


Fig.1 Feed-forward Neural Network Structure with Two Hidden Layers

Furthermore, the results of all the aforementioned algorithms are fed as input to the neural network. We use a neural network that applies regression to increase the precision of the result. The neural network calculates them by comparing all the predictions and to show the most accurate result did the job as well.

I. Implementation

This research used the Jupyter IDE. It is an open-source web application that helps us to share and create documents with code in real time, visualization, equations and narrative text. Here we accumulated residence income associated records to estimate the residence fees primarily based totally on actual global dataset USA. It is a public output dataset of that particular location in the USA. In Order to discover the efficient regression approach for prediction, we require sure parameters to perform comparison a few of the strategies. The parameters selected for the evaluation are Scores of the algorithm, [MSE] Mean Square Error and [RMSE] Root Mean Square Error, R2Square, cross validation. The underneath Table 1 represents the consequent precis of the parameters, whilst above strategies are carried out practically.

Table 1:Comparison of Algorithms

	Model	MAE	MSE	RMSE	R2 Square	Cross Validation
0	Linear Regression	82819.662944	1.049064e+10	102423.820750	0.917654	0.917379
1	Polynomail Regression	83155.418051	1.058190e+10	102868.357999	0.916938	0.000000
2	Robust Regression	83355.536337	1.061246e+10	103016.776390	0.916698	0.911415
3	Ridge Regression	83705.764658	1.065735e+10	103234.445007	0.916346	0.917379
4	Lasso Regression	82820.108151	1.049068e+10	102424.008667	0.917654	0.917379
5	Elastic Net Regression	83056.368651	1.052769e+10	102604.522004	0.917363	0.879545
6	Stochastic Gradient Descent	82819.664255	1.049064e+10	102423.820196	0.917654	0.000000
7	SVM Regressor	91817.938656	1.314823e+10	114665.715503	0.896794	0.000000
8	Random Forest Regressor	96547.150435	1.504251e+10	122647.916184	0.881924	0.000000
9	Artficial Neural Network	98139.106370	1.551028e+10	124540.290374	0.878253	0.000000

From the above table, we are able to without difficulty carry out comparison of unique algorithms virtually to discover the satisfactory among them. Figure 2 underneath is used to virtually visualize the overall performance of diverse strategies in a graphical layout based on their scores. In Figure 2,x axis represents the diverse regression strategies taken into consideration for take a look at and y-axis represents the rating values observed

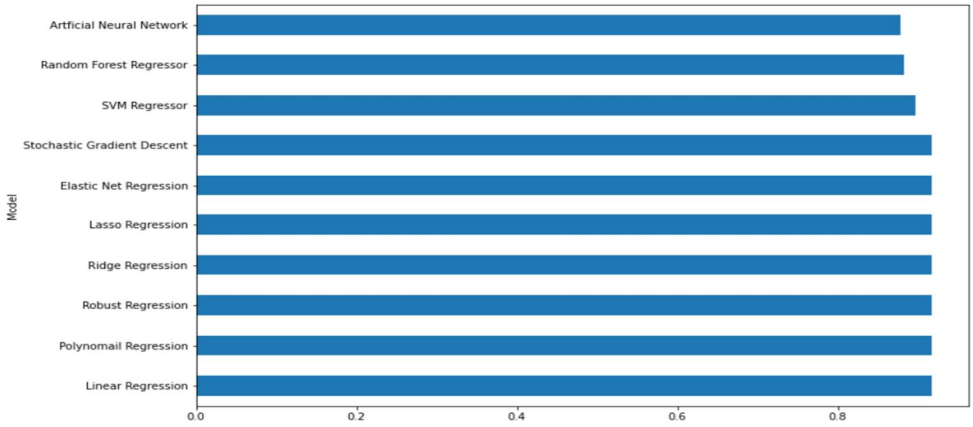


Fig.2 Comparison of Various Regressions

The graphical representation of all the different regression techniques

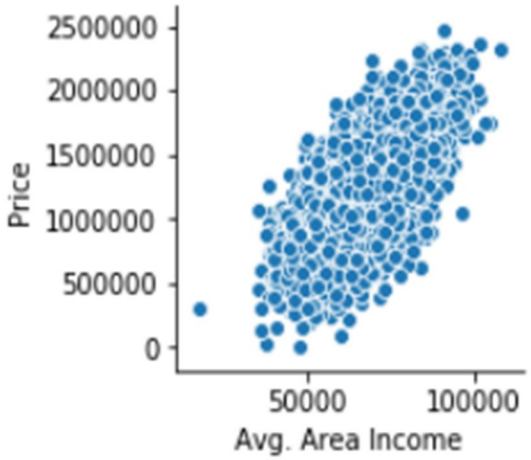


Fig.3 Price vs Area Income

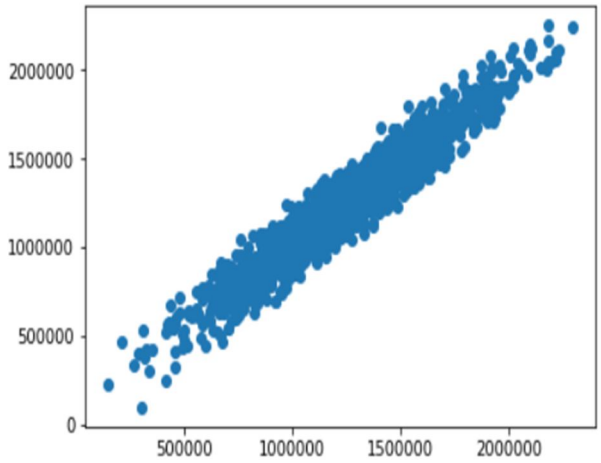


Fig 4 Ridge Regression

V. CONCLUSION

This article focuses on the comparison of different machine learning algorithms (linear regression, random forest regression, polynomial regression, robust regression, lasso regression, elastic net regression, stochastic gradient descent, svm regressor, artificial neural network). Of the experimental results above, the precision rate of the gradient augmentation algorithm is higher than that of all the other algorithms. The article can be extended by applying the algorithm described above to predict the resale value of houses. Further, most we add several cities to compute the house price values.

REFERENCES

[1]. Model,Azme Bin Khamis, Nur Khalidah Khalilah Binti Kamarudin ,Comparative Study On Estimate House Price Using Statistical And Neural Network, International journal of scientific and technology,research volume 3, ISSUE 12, December 2014,Page(s):126-131.

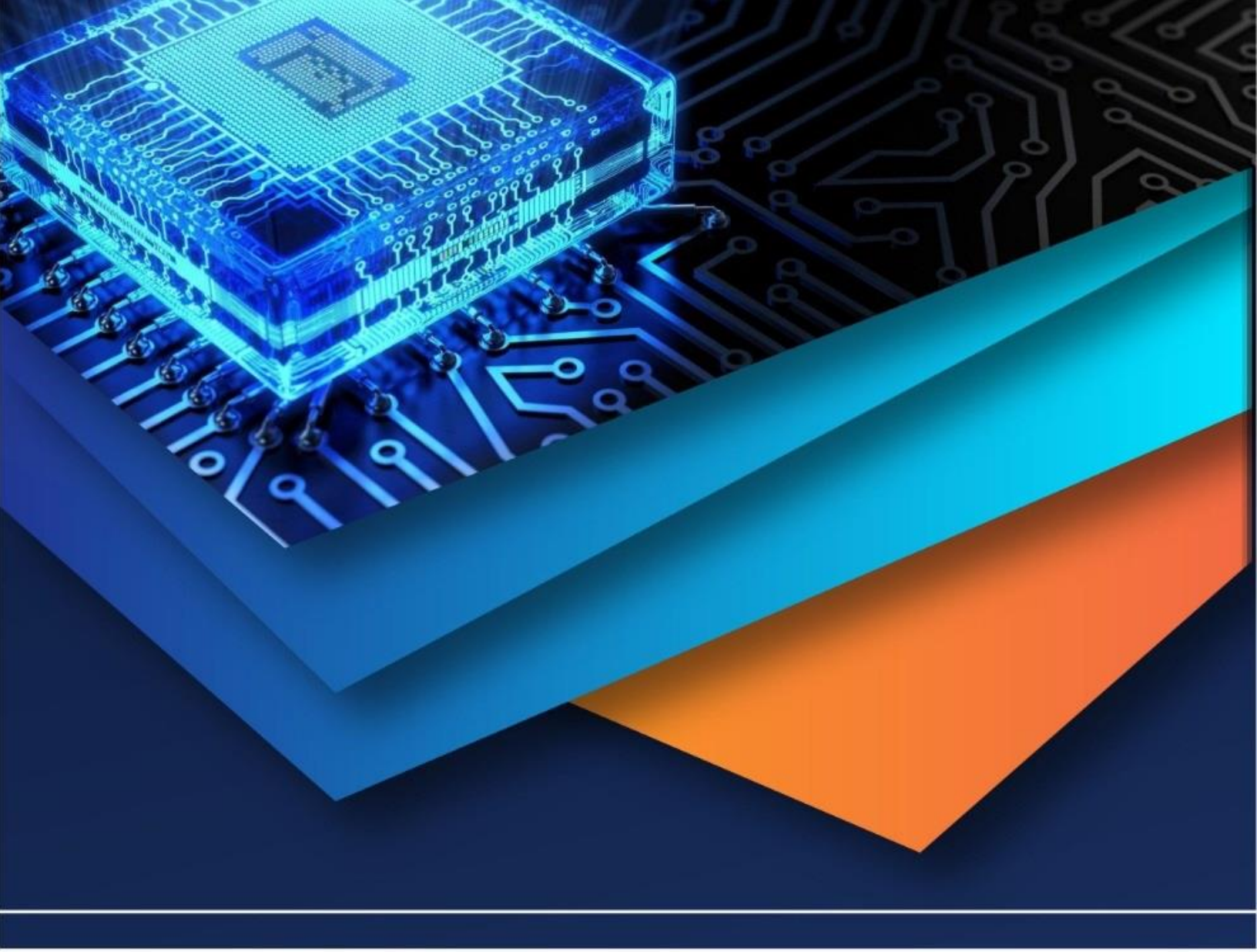
[2]. N. N. Ghosalkar and S. N. Dhage, "Real Estate Value Prediction Using Linear Regression," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-5.

[3]. 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, 2012, pp. 154–168, ISBN: 9783642315367. DOI: 10 . 1007 / 978 - 3 - 642 - 31537-4\ 13 [4] J.

[4]. Madhuri, C. R., Anuradha, G., & Pujitha, M. V. (2019). *House Price Prediction Using Regression Techniques: A Comparative Study. 2019 International Conference on Smart Structures and Systems (ICSSS).* doi:10.1109/icsss.2019.8882834

[5]. Varma, A., Sarma, A., Doshi, S., & Nair, R. (2018). House Price Prediction Using Machine Learning and Neural Networks. 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT). doi:10.1109/icicct.2018.8473231

[6]. Housing Price Prediction via Improved Machine Learning Techniques Author links open overlay panel Department of Computer Science, Texas Christian University, Fort Worth, TX 76129, United States.
[7]. Banerjee, Debanjan, and Suchibrota Dutta. "Predicting the housing price direction using machine learning techniques." 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI). IEEE, 2017.
[8]. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-9, July 2019 717 Published By: Blue Eyes Intelligence Engineering & Sciences Publication Retrieval Number: I7849078919/19©BEEP DOI:10.35940/ijitee.I7849.078919 House Price Prediction Using Machine Learning G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu.
[9]. A hybrid regression technique for house prices prediction December 2017DOI:10.1109/IEEM.2017.8289904Conference: 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) Project: AI theory and application.
[10]. Housing Sale Price Prediction Using Machine Learning Algorithms 2020 Author(s): Zhou, YichenAdvisor(s): Wu, Yingnian
[11]. Joep Steegmans, Wolter Hassink, "Financial position and house price determination: An empirical study of income and wealth effects", Journal of Housing Economics, vol. 36, pp. 8-24, June 2017
[12]. [11] Lakshmi, B. N., and G. H. Raghunandhan. "A conceptual overview of data mining." 2011 National Conference on Innovations in Emerging Technology. IEEE, 2011.
[13]. Manjula, R., et al. "Real estate value prediction using multivariate regression models." Materials Science and Engineering Conference Series. Vol. 263. No. 4. 2017.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY