International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/IJSRCSEIT

Using Enhanced Machine Learning Techniques, House Price Prediction

[1] **T. Swaroopa**, [2] **K. Roopa**

[1] MCA student, [2] Assistant Professor

Department of Computer Applications, Madanapalle Institute of Technology and Science, Angallu, Andhra Pradesh, India

ARTICLEINFO

Article History:

Accepted: 01 Aug 2023

Published: 09 Aug 2023

Publication Issue

Volume 9, Issue 4

July-August-2023

Page Number

285-291

ABSTRACT

The goal of the paper is to help prospective homeowners make well-informed selections that take into account their finances and market trends. It attempts to forecast house prices for those who do not own a home by examining their financial plans and goals. Various regression techniques, such as Linear Regression and Decision Tree Regression, are employed to estimate speculated prices. The analysis involves assessing the accuracy and R2 score values of the predictions using relevant data. The study's objective is to determine the most effective regression technique for house price prediction. The ultimate objective is to assist sellers in accurately calculating the price at which to sell a home and to assist buyers in determining when it is best to make a real estate acquisition.

Keywords : Machine learning, House price, Prediction.

I. INTRODUCTION

The real estate market experiences significant fluctuations, influenced by numerous complex financial indicators. However, the advancements in technology offer an opportunity to achieve stable returns on house prices and assist experts in identifying the most informative indicators for better predictions. The primary goal of the HOUSE PRICE PREDICTION project was to make highly accurate house price predictions using suitable algorithms and determine the best approach with the lowest error rate. This problem is intriguing as it directly impacts the majority of people who will eventually buy or sell a home. As housing market analysts, solving this problem enables us to gain deeper insights into the real estate market and make more informed decisions. Precise house price predictions are crucial for maximizing profits in real estate transactions while minimizing risks.

II. RELATED WORKS

IoT-based Water Quality Monitoring System: When buying a new home, people take prudence, taking their budgets and market strategies into account. The purpose of this article is to provide prospective homebuyers who do not already own a home with an accurate prediction of house prices. Based on their financial plans and goals, this prediction. The study involves analyzing various factors, including previous market trends, fare ranges, and development forecasts, to estimate speculated prices. Multiple regression techniques such as Elastic Net Regression, Multiple Linear Regression, Gradient Boosting Regression, LASSO Regression, Ridge Regression, and Ada Boost Regression are employed to predict house prices using a dataset. The main aim is to identify the best-performing regression technique among them. The paper intends to assist sellers in determining the optimal selling price for a house and to aid potential buyers in predicting the ideal time to make a purchase. Additionally, the research considers various related factors that influence house prices, such as physical conditions, conceptual aspects, and location.

Analysis of Regression and Optimization Using Particle Swarm for Modeling House Price Prediction Case Study: As house prices consistently rise each year, there is a growing demand for a reliable system to predict future house prices. These forecasts can assist builders in selecting the ideal home sale price and assist buyers in choosing the best time to buy. The physical state, design, and location of a home all have an impact on its cost. Using NJOP houses as a starting point, this study employs regression analysis and particle swarm optimization (PSO) to estimate home values in Malang city. PSO is employed to select significant variables that affect house prices, while regression analysis is utilized to find the optimal coefficients for accurate predictions. The study's results demonstrate that the combination of regression and PSO is well-suited, yielding minimal prediction errors.

Using SDLC software metrics, design and implementation of a hybrid phase-based ensemble technique for fault finding: Accurately estimating programming effort remains one of the most significant challenges for software engineers. Initial assessments made during the proposal stage often suffer from a high degree of inaccuracy, as requirements are not yet defined in the finest details. However, as the project progresses and requirements become clearer, the accuracy and confidence in the estimates improve. Selecting the appropriate software effort estimation techniques is crucial for accurate prediction. In this article, Artificial Neural Network (ANN) and Support Vector Machine (SVM) are applied to a reliable dataset to forecast programming effort. These techniques are employed to enhance the accuracy of the estimates and aid in making informed decisions during software development.

Comparative Analysis of Statistical and Neural Networks for Estimating Home Prices: The housing market holds tremendous importance for the overall economy. Activities related to housing construction and renovation contribute to economic growth by increasing aggregate expenditures, employment opportunities, and the volume of house sales. Furthermore, these activities stimulate demand for related industries, including household durables. Fluctuations in house prices directly impact the asset portfolios of many households, as a house often represents their largest single asset. Accurate house price predictions are essential for various stakeholders, such as aspiring homeowners, builders, investors, tax assessors, appraisers, mortgage lenders, and insurance companies.

Without an agreed benchmark or certification process, traditional methods of estimating property prices frequently rely on comparing costs and sale prices. Therefore, the availability of a trustworthy model for predicting house prices can close this knowledge gap and improve the real estate market's general effectiveness. Recent findings suggest that houses with more bedrooms and bathrooms tend to be priced higher, newer houses are generally more expensive than older ones, and houses with gardens typically command higher prices than those without. Studies highlight the crucial role of the housing sector as a leading indicator of the broader economy, with house prices providing valuable insights into inflation and output forecasts. Numerous previous studies have established empirical evidence of significant interrelations between house prices and various economic factors, including income, interest rates, construction costs, and labor market variables.

Analysis of House Price Variations Using Multiple Regressions: The application of rigorous statistical analysis to assist in investment decision-making is gaining momentum in both the United States of America and the United Kingdom. However, in Malaysia, the response from local academics has been relatively slow, and even slower among practitioners. This paper showcases the utilization of Multiple Regression Analysis (MRA) and its extension, Hedonic Regression Analysis, in explaining price variations for selected houses in Malaysia. Each attribute identified as a theoretical price determinant is quantified, and the perceived contribution of each attribute is explicitly displayed. The paper demonstrates how statistical analysis can effectively analyze property investment by considering multiple determinants. This rigorous consideration of various characteristics enables better-informed investment decision-making processes.

III. Methodology

Proposed system:

The proposed system utilizes efficient supervised learning methods for house price prediction. Given the consistent annual increase in house prices, the need for a reliable prediction system becomes evident. Additionally, land prices are predicted using a different technique with a new set of parameters, and compensation for property settlements is also forecasted. Mathematical relationships, when expressed with precise numbers, offer enhanced clarity in understanding various aspects of everyday life. Technological advancements present an opportunity to achieve stable returns in real estate and assist experts in identifying the most informative indicators for improved predictions.

Regression analysis plays a crucial role in establishing the relationship between a numeric dependent variable and one or more numeric independent variables. Developers may tremendously profit from house price prediction by using it to determine the best selling price, and customers can use it to decide when is the best time to buy. The proposed system successfully achieves expected and predicted values through the use of efficient Regressor methods. Users can conveniently predict house prices using our application.

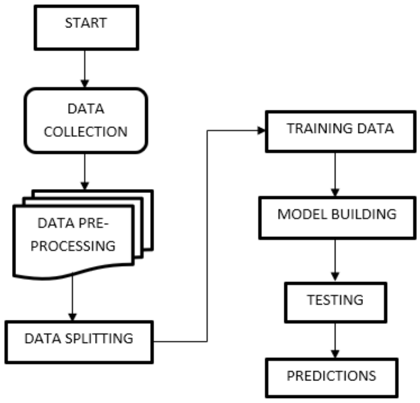


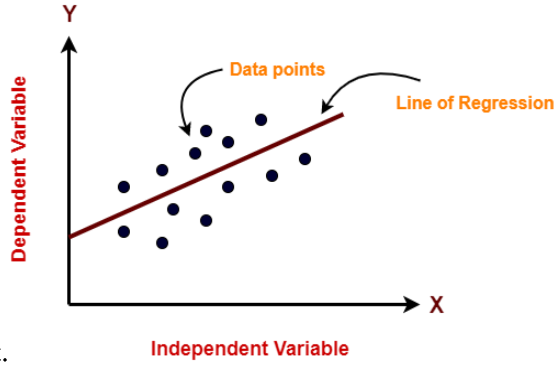
Figure 1: Block diagram

IV. Implementation

The project has implemented by using below listed algorithm.

LINEAR REGRESSION

A supervised machine learning approach called linear regression is made to forecast continuous outputs with a fixed slope. It is especially helpful for forecasting within a continuous range, such as sales or pricing, as opposed to categorizing data into separate groups, like "cat" or "dog." In order to make accurate predictions based on this linear relationship, the algorithm creates a linear relationship between the input features and the



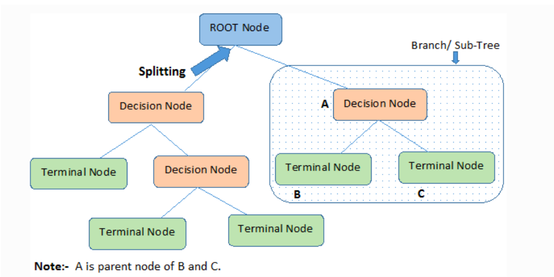
output.

DECISION TREE:

A non-parametric supervised learning technique called decision trees is used for both classification and regression applications. The main goal is to build a model that can forecast the value of a target variable by deriving simple decision-making guidelines from the data attributes.

A decision tree is depicted as being drawn upside-down, with the root at the top. A condition or internal node, shown as bold text in black within the left-hand image, specifies how the tree is branched into several edges or branches. Each branch eventually leads to a decision or leaf, where no further splitting occurs. In the case described, the decision could be whether a passenger survived or died, represented respectively as red and green text. Decision trees are powerful tools for making informed predictions based on learned rules from the data.

<u>**Block Diagram for Decision Tree Algorithm:**</u>



Root Node: It acts as the tree's starting point and represents the complete population or sample. This node is then split into two or more subcategories according to specific criteria.

Splitting: Splitting is the action of dividing a node into two or more homogenous sub-nodes. Each split is determined by a distinct characteristic or quality.

Decision Node: A sub-node is referred to as a decision node when it further divides into further sub-nodes. These nodes assist in the tree's branching.

Leaf / Terminal Node: Leaf or terminal nodes are nodes that stop further division. They stand in for the decision tree's final result or outcome.

Pruning: Pruning is the process of deleting sub-nodes from a decision node. By doing so, the tree is made simpler and overfitting is avoided.

Branch / Sub-Tree: A branch or sub-tree is a division of the complete decision tree. Each branch illustrates a series of choices that resulted in a certain result.

AdaBoost:

The common ensemble method in machine learning for classification tasks is called AdaBoost, which stands for Adaptive Boosting. It is called "Adaptive Boosting" because it gives different weights to each instance, with heavier weights going to cases that were mistakenly categorised. The goal of boosting is to reduce bias and variance in supervised learning by sequentially growing learners. Each subsequent learner is built based on the errors from the previously grown learners, converting weak learners into strong ones.

During the training phase, several decision trees are built, which is how boosting functions. The first decision tree or model is built, and the following model gives precedence to the records that were erroneously classified in the initial model. This process keeps going until a certain threshold of basic learners is reached. It's crucial to remember that all boosting approaches permit record repetition.

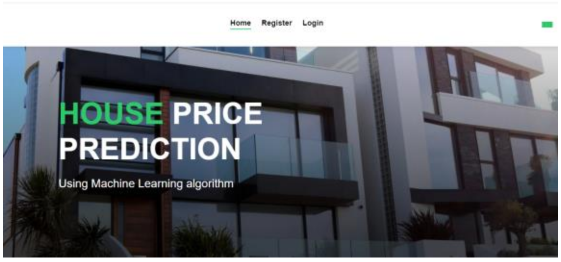
The trees produced by the AdaBoost algorithm are referred to as stumps, and they are made up of a single node and two leaves. The arrangement of these stumps is important in AdaBoost since they are regarded as weak learners. The first stump's mistake has an impact on how subsequent stumps are built.

AdaBoost is particularly effective when used to boost the performance of decision trees for binary classification problems. It can be applied to any machine learning algorithm, but it is most beneficial when used with weak learners. Originally called AdaBoost.M1, it is now commonly referred to as discrete AdaBoost due to its classification focus, as opposed to regression.

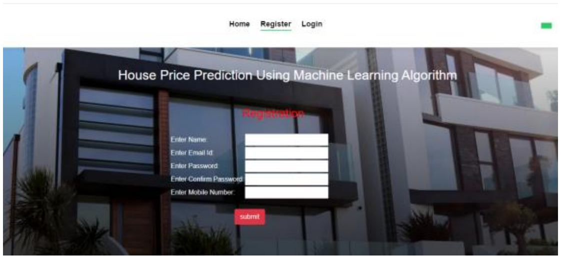
V. Results and Discussion

The following screenshots are depicted the flow and working process of project.

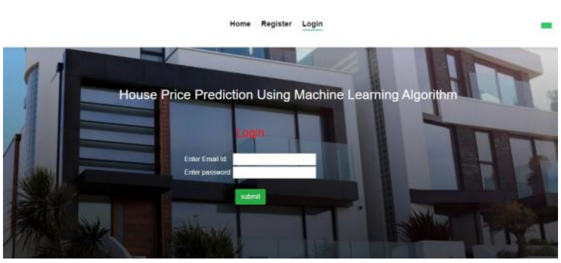
Home page: This is the home page of house price prediction.



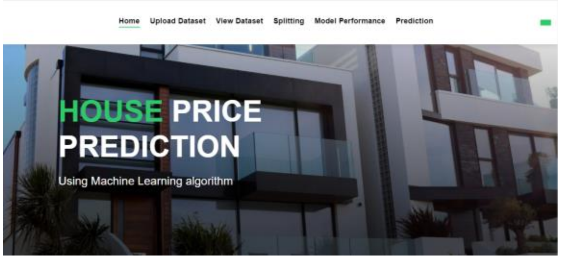
Registration page: Here the user can gets registered into the application.



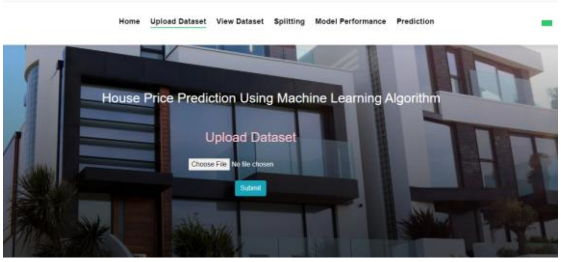
Login page: User can login with valid credentials.



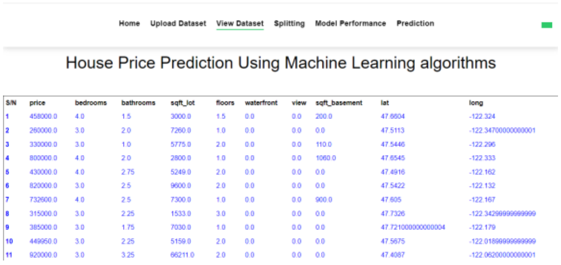
User Home page: After login user can view the home page.



Upload dataset: Here the user can upload dataset.

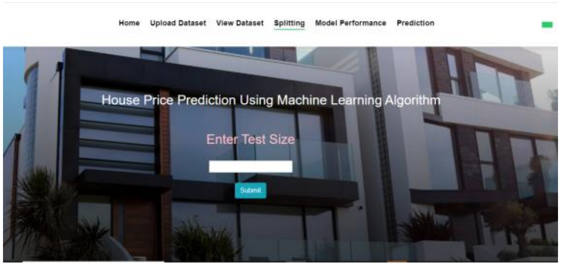


View Dataset: Here we can view the all uploaded data.

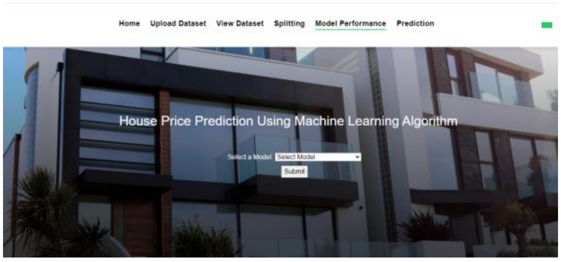


S/N	price	bedrooms	bathrooms	sqft_lot	floors	waterfront	view	sqft_basement	lat	long
1	458000.0	4.0	1.5	3000.0	1.5	0.0	0.0	200.0	47.6604	-122.324
2	260000.0	3.0	2.0	7260.0	1.0	0.0	0.0	0.0	47.5113	-122.34700000000001
3	330000.0	3.0	1.0	5775.0	2.0	0.0	0.0	110.0	47.5446	-122.296
4	800000.0	4.0	2.0	2800.0	1.0	0.0	0.0	1060.0	47.6545	-122.333
5	430000.0	4.0	2.75	5249.0	2.0	0.0	0.0	0.0	47.4916	-122.162
6	820000.0	3.0	2.5	9600.0	2.0	0.0	0.0	0.0	47.5422	-122.132
7	732600.0	4.0	2.5	7300.0	1.0	0.0	0.0	900.0	47.605	-122.167
8	315000.0	3.0	2.25	1533.0	3.0	0.0	0.0	0.0	47.7326	-122.34299999999999
9	385000.0	3.0	1.75	7030.0	1.0	0.0	0.0	0.0	47.721000000000004	-122.179
10	449950.0	3.0	2.25	5159.0	2.0	0.0	0.0	0.0	47.5675	-122.01899999999999
11	920000.0	3.0	3.25	66211.0	2.0	0.0	0.0	0.0	47.4087	-122.06200000000001

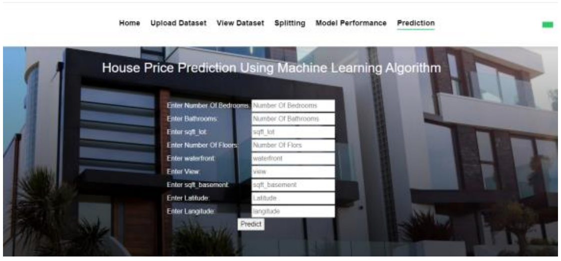
Splitting data: Here we can split the data.



Model performance: Selected model will work on the given data.



Prediction: Here we can see the predicted results from selected models.



VI. Conclusion

The primary focus of this article is to compare two machine learning algorithms, Linear Regression, and Decision Tree Regressor, for House Price Prediction Analysis. Through the conducted experiments, the results indicate that the Decision Tree Regressor algorithm achieves higher accuracy compared to all other algorithms when predicting house prices. Accuracy is assessed using the Mean Square Error (MSE) and R2 score, which both measure how well the method performed on the King County Dataset, a publicly available dataset. To further enhance the paper, future research can extend the application of these algorithms to predict the resale value of houses.

VII. REFERENCES

[1]. Aminah Md Yusof and Syuhaida Ismail's paper, "Multiple Regressions in Analyzing House Price Variations." IBIMA Publishing, Communications of the IBIMA Vol. 2012 (2012), Article ID 383101, 9 pages.

[2]. What you see may not be what you get: 2. Babyak, M. A. a succinct, layman's guide on over fitting regression-type models. Medical Psychology, 66(3), 411-421.

[3]. Jinal Jain, Akshata Gaud, Priyanka Khaire, Atharva Chogle.Forecasting housing prices using data mining methods International Journal of Advanced Research in Computer and Communication Engineering, volume 6, issue 12, December 2017, with ISO 3297:2007 certification.

[4]. Predicting Zillow Estimation Error Using Linear Regression and Gradient Boosting, IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Page(s):530 - 534.

[5]. Model, Comparative Study On Estimating House Price Using Statistical And Neural Networks, International Journal of Scientific and Technology, Research Volume 3, Issue 12, December 2014, Pages 126–131, Nur Khalidah Khalilah Binti Kamarudin, Azme Bin Khamis.

[6]. Malang, East Java, and Indonesia for Modeling House Price Prediction Using Regression Analysis and Particle Swarm Optimization by Adyan Nur Alfiyatin, Hilman Taufiq, Ruth Ema Febrita, and Wayan Firdaus Mahmudy. 2017's Vol. 8, No. 10 of the International Journal of Advanced Computer Science and Applications (IJACSA); pages 323–326.

[7]. "Using SDLC software metrics, design and implementation of a hybrid phase-based ensemble technique for fault finding ", An International Conference by IEEE, pp. 268-274, 2016.

[8]. Data mining-based house price forecasting by Ankit Mohokar, Nihar Bhagat, and Shreyash.The International Journal of Computer Applications 152(2):23–26 in October 2016.

[9]. VU Amsterdam, March 30, 2017, Valeria Fonti, Paper titled " Research Paper on Business Analytics: Feature Selection Using LASSO

[10]. Peter B. Luh, " Improved Extended Kalman Filter Method for Neural Network-Based Market Clearing Price Prediction and Confidence Interval Estimation," IEEE Transactions on Power Systems 20(1):59–66, March 2005.

[11]. New Zealand's University of Lincoln, Canterbury 8150" Hedonic Price Model vs. Artificial Neural Network for House Price Prediction, American Journal of Applied Sciences 1 (3): 193-201, 2004, Christopher Gan and Minsoo Lee.

[12]. "IoT-based Water Quality Monitoring System ", An International Conference by IEEE, pp. 109-113, 2018.

[13]. Ahmed Khalafallah, " Model Using Neural Networks to Predict Housing Market Performance," Tsinghua Science & Technology 13(S1):325–328, Oct.

[14]. Martin Edward Ralph Hoesli, Eva Cantoni, and Steven C. Bourassa, " House price forecasting using spatial dependence, and housing submarkets" 2007; 143–160; The Journal of Real Estate Finance and Economics.

[15]. Chris Brunsdon, A. Stewart Fotheringham, and Martin E. Charlton, " A Technique for Investigating Spatial Nonstationarity: Regression with Weights Based on Location," in Geographic analysis, volume 28, issue 4, pages 281-375, 1996.

Cite this article as :

T. Swaroopa, K. Roopa, "Using Enhanced Machine Learning Techniques, House Price Prediction ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 4, pp. 285-291, July-August-2023.