iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37451

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 VIII Aug 2021- Available at www.ijraset.com

House Price Prediction using Machine Learning

Millee Samukcham[1], Riya Choudhary[2], M. Bhavana[3], Yeluru Sreeja Reddy[4]
[1, 2, 3, 4]*Department of Computer science and Engineering, MVJ College of Engineering, Bengaluru, India*

Abstract: ***With the increase in industrialisation, people are also a lot more careful today when they make an attempt to shop for a brand new house with their budgets and market strategies. Until date, existing websites gift solely the house costs given by the homeowners and details of the house largely infrastructure. Some websites even offer comparison between completely different homes with the same infrastructure. But, some individuals aren't awake to what quantity a house with an exact infrastructure is meant to value and are not ready to find how much is sweet enough to be ready to find frauds. individuals additionally want alternative factors however infrastructure to come to a decision whether or not or to not obtain a house Machine learning algorithmic program helps us in enhancing security alerts, guaranteeing public safety and improve medical enhancements.***
Keywords: Logistic Regression, Linear Regression, House Price Prediction

I. INTRODUCTION

Data is at the heart of technical innovations, with predictive models it is now possible to achieve any result. Machine learning is widely used in this approach. Machine learning means providing a valid data set, and furthermore, when predictions are based on it, the machine learns from its preloaded data and accordingly how important a certain event can be for the entire system. Several modern uses of this technique include predicting stock prices, predicting the possibility of an earthquake, predicting company sales, and the list is endless. For our research project, we considered Bangalore as our main location, and are predicting house prices for various locations in and around Bangalore.

What is machine learning? Machine Learning is an application of artificial intelligence (AI) that enables systems to automatically learn from experience and improve without being explicitly programmed .The learning process begins with observations or data, such as examples, direct experience, or instructions, to look for patterns in the data, and use the examples we provide to make better decisions in the future. The primary aim is to enable computers to learn automatically without human intervention or input and to adjust actions accordingly.

Supervised machine learning algorithms can apply what they have learned in the past to new data using flagged examples to predict future events The system is able to provide goals for each new input after sufficient training. The learning algorithm can also compare its output with the correct and expected output and find errors to modify the model accordingly. Supervised learning has input variables (x) and an output variable (y) and uses an algorithm to learn the mapping function from input to output. Supervised learning is the task of learning a function that maps an input to an output based on example input and output pairs. It derives a function from the designated training set consisting of a set of training examples . A supervised learning algorithm analyses the training data and generates a derived function that can be used to map new examples. The optimal scenario will allow the algorithm to correctly determine the class labels for invisible instances.

Unsupervised machine learning algorithms are used when the information used for training is unclassified or marked . Unsupervised learning examines how systems can derive a function to describe a hidden structure from unlabelled data ,examines data and can pull data set inferences to describe hidden structures from unlabelled data. Unsupervised learning is the training of a model using information that is not classified or tagged. This model can be used to group input data into classes based on its statistical properties. Unsupervised machine learning is the task of deriving a function to describe a hidden structure from "unlabelled" data (no classification or categorization is included).Since the examples given are not labelled, it is the accuracy of the structure that generates the relevant algorithm ,which is a way of distinguishing unsupervised learning from supervised learning and reinforced learning.

A drawback of unsupervised learning is the problem of density estimation in statistics, although learning without supervised learning involves many other problems that summarize and explain Networks are subsumed according to Donald Hebb's principle, i.e. neurons that fire together, connect with each other.

Unsupervised learning occurs when an algorithm learns from simple examples without any associated response, allowing the algorithm to determine the data patterns on its own. This type of algorithm tends to restructure the data into something else, such as new features that can represent a class or a new set of uncorrelated values.

They are very useful for providing people with information about the importance of data and useful new input to supervised machine learning algorithms. As a way of learning, it is similar to the methods people use to find out that certain objects or events are of the same class, for example by observing the degree of similarity between objects. Some recommendation systems that we find on the web in the form of marketing automation are based on this type of learning.

When the machine is used to make certain decisions, reinforcement learning is used. While doing this, the model is in a environment where it trains itself to make it more accurate through trial and error. The model learn from experience and capture the knowledge of this area to make accurate decisions. Example of Reinforcement Learning : Markov Decision Process -hot encoding is one of these reinforcement learning algorithms.

II. LITERATURE SURVEY

House Price Prediction Using Machine Learning by G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu used various models to predict the price. They have compared linear regression,lasso regression and gradient boosting for the same dataset. Gradient boosting regression has the highest accuracy out of the three algorithms for their dataset. Their dataset contained only numeric values. They plotted the relation between the independent features for the data mining and fitted the models to the cleaned data which took them more than one day to only just prepare the data. They have mentioned that further improvements could be made with parallelising the computations.

House Price Prediction Using Machine Learning and RPA Prof. Pradnya Patil and Darshil Shah, Harshad Rajput, Jay Chheda Undergraduate Student, Computer Engineering Department, K. J. Somaiya Institute of Engineering and Information Technology, Mumbai, India has proposed a system that uses RPA for data scraping. After data extraction, they have performed data cleaning. They have mentioned that the dataset they have used is quite small and hence they have opted for random forest. They state that random forest works faster and accurately for smaller dataset. It reaches a state easily where more dataset improve the accuracy. So,they have paired with it up with Catboost and then again with LightBGM. They received better accuracy with CatBoost for a dataset that had categorical features.

Housing Prices Prediction with a Deep Learning and Random Forest Ensemble Bruno Klaus de Aquino Afonso1 , Luckeciano Carvalho Melo2 , Willian Dihanster Gomes de Oliveira1 , Samuel Bruno da Silva Sousa1 , Lilian Berton1 have used two different architectures to predict the model. They have used enriched RF for the first and KISS model for the other architecture. The location features were enriched and then mixed with Regex for textual attributes. After all these steps, Random forest was passed to predict the house prices. On the other hand, KISS which stands for 'Keep it simple, stupid' is a deep architectures which uses a shared embedded layer of textual attributes, processed image features and enriched tabular data. According to them, the KISS model was the best architecture for housing prices prediction, since it shared the embedding layer among all the textual attributes, without losing information by training models separately for these features. Although enriched RF worked well with numeric features it couldn't handle raw image or text data. KISS, on the other hand, can represent all kinds of data through the embedding layers. A disadvantage of KISS is that it did not handle numeric attributes as well as the random forest did, and had to rely on more data types to edge it out.

III. PROPOSED SYSTEM

Our main focus here is on developing a model that predicts the cost of ownership for a customer according to their interests. Our model analyses a number of parameters chosen by the customer to find an ideal price according to their needs and interests. It uses a classic technique called logistic regression and collaborative filtering for prediction and tries to give an analysis of the results obtained. Our data set consists of several essential parameters and data mining was the root of our system.

First we clean up our entire data set and also truncate outliers. Additionally, we weigh each parameter according to its importance to determine the price of the system, and this has led us to increase the value that each parameter retains in the system. We have selected 3 different machine learning algorithms and tested our system with different combinations, which can guarantee the best possible reliability of our results.

Even after that, we took a unique approach to increase accuracy, our survey led to the result that the real value of real estate also depends on nearby local amenities like train station, supermarket, school, hospital, temple, parks, etc. And now we propose our unique approach that can address this need. If we find any places in the district we are increasing the property value accordingly. We did it with manual examples and this gave us tremendous results in terms of prediction accuracy.

This project contains two modules: Admin module and user module. The features for both module are explained below. In the Admin module: The admin can view his profile and edit his profile details. He/ She can also view the users and delete each of them. He/ She can load and display the data record and can also change his/her password and log out.

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

In the User module, a user can log in and view his data. He/ She can also edit the profile details.
A user can check the city, area and type of house they
are looking for and results are shown for the same. A user can also check the house price based on the city, area and budget they have. Finally, he/she can check the house price based on the entire city. A user can change his/her password and log out of his/her profile. Here in our proposed system, we will be using logistic regression and collaborative filtering which are further discussed in detail below.

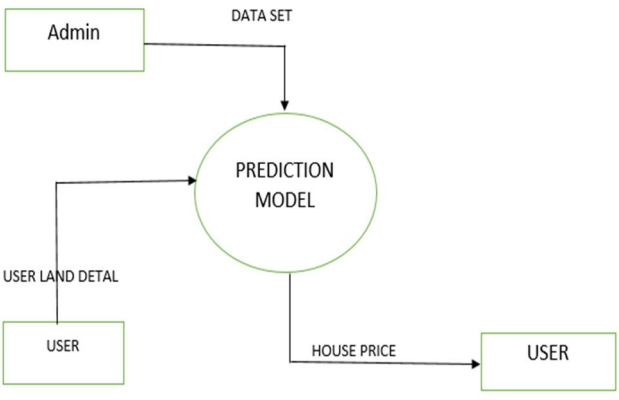
Logistic regression comes under supervised learning classification algorithm and is used to predict the probability of a target variable. It is one of the simplest ML algorithms. It is used for various classification problems such as spam detection, Diabetes prediction, cancer detection etc. Logistic regression is actually used to calculate the probability of a binary event occurring, and to deal with the issues of classification. For example, predicting if an incoming email is spam or not , or predicting if a credit card transaction is fraudulent or not. In a medical context, logistic regression may also be used to predict whether a tumor is benign or malignant. In the field of marketing, it may be used to predict if a given user (or group of users) will buy a certain product or not.

An online education company can make use of logistic regression to predict whether a student will complete their course on time or not. In short, logistic regression is used to predict the likelihood of all kinds of “yes” or “no” outcomes. By predicting such kinds of outcomes, logistic regression helps data analysts (and the companies they work for) to make informed decisions. In conclusion, this helps to both minimize the risk of loss and to optimize spending in order to maximize profits.

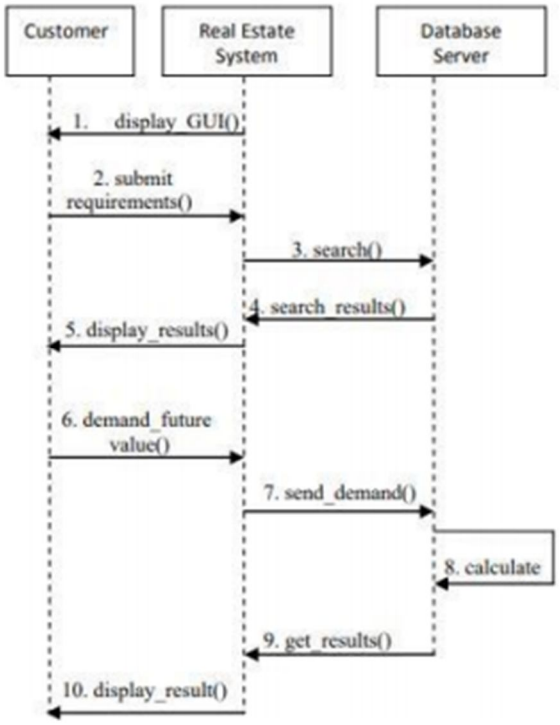
Linear regression models the relationship between two variables known as dependent and independent variables. This is done by fitting a linear equation to the observed data. One variable is taken an explanatory variable and the other is taken a dependent variable. For example, a modeler would like to include the weights of individuals. relate their height to a linear regression model. Before attempting to fit a linear model to the observed data, a modeler must determine whether or not there is a relationship between the variables of interest, which does not necessarily mean that one variable causes the other (e.g. causes higher college grades), but that there is a significant relationship between the two variables.

A scatter plot is a useful tool for determining the strength of the relationship between two variables. If the explanatory variables and the proposed dependencies do not have any relationship between them (i.e. the scatter plot does not show rising or falling trends) then fitting with a linear regression model is unlikely to provide a useful model for the data. A valuable numerical measure of the association between two variables is the correlation coefficient, which is a value between -1 and 1 that indicates the strength of the association between the observed data for the two variables.

IV. SYSTEM DESIGN AND IMPLEMENTATION



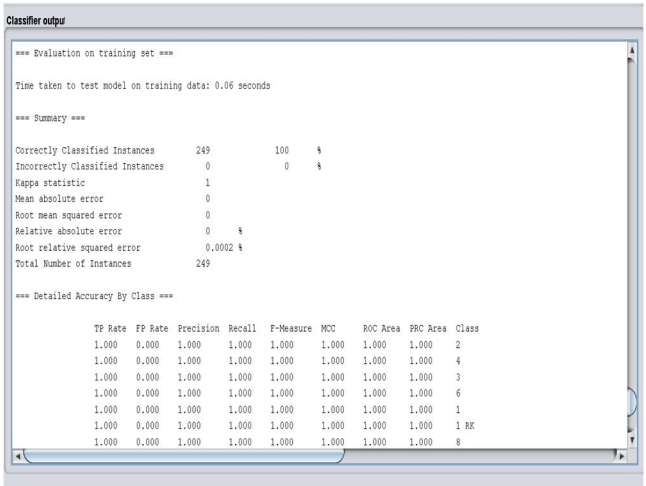
The admin module provides the data set for the predictive model. The user module then provides the required properties / parameters according to its interest for the predictive model. The predictive model takes as input the price of the house and generates it, which is delivered to the user module

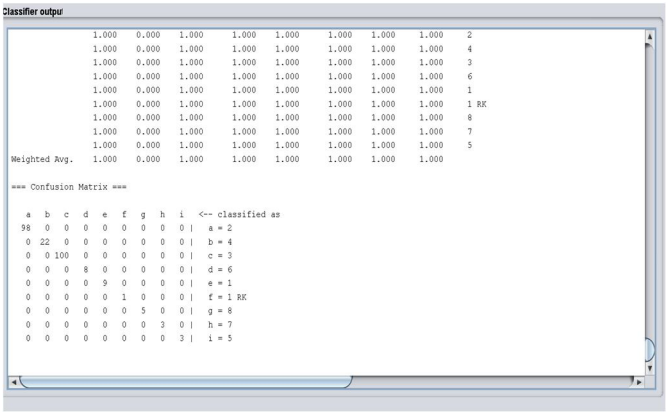


The real estate system shows the user the GUI in order to make the necessary entries for authorization. The client enters the data required for authentication and transmits it to the system. The real estate system searches for the entered data if it is available in the database. The search result is transferred from the database to the real estate system. If the user exists, the user is redirected to the next corresponding page; otherwise the GUI will display an invalid message to the user. Now the user, once authenticated, requests future values from the system. Then the system sends this request to the database. The database performs the calculation required for the query. The database then sends the results to the system. Finally, the system displays the result to the user.

V. RESULTS

We have used linear regression and logistic regression as a model to predict the house prices. For the data that we have used linear regression turns out to fit much better than other models as the features are mostly numerical features and there is less presence of categorical independent features.



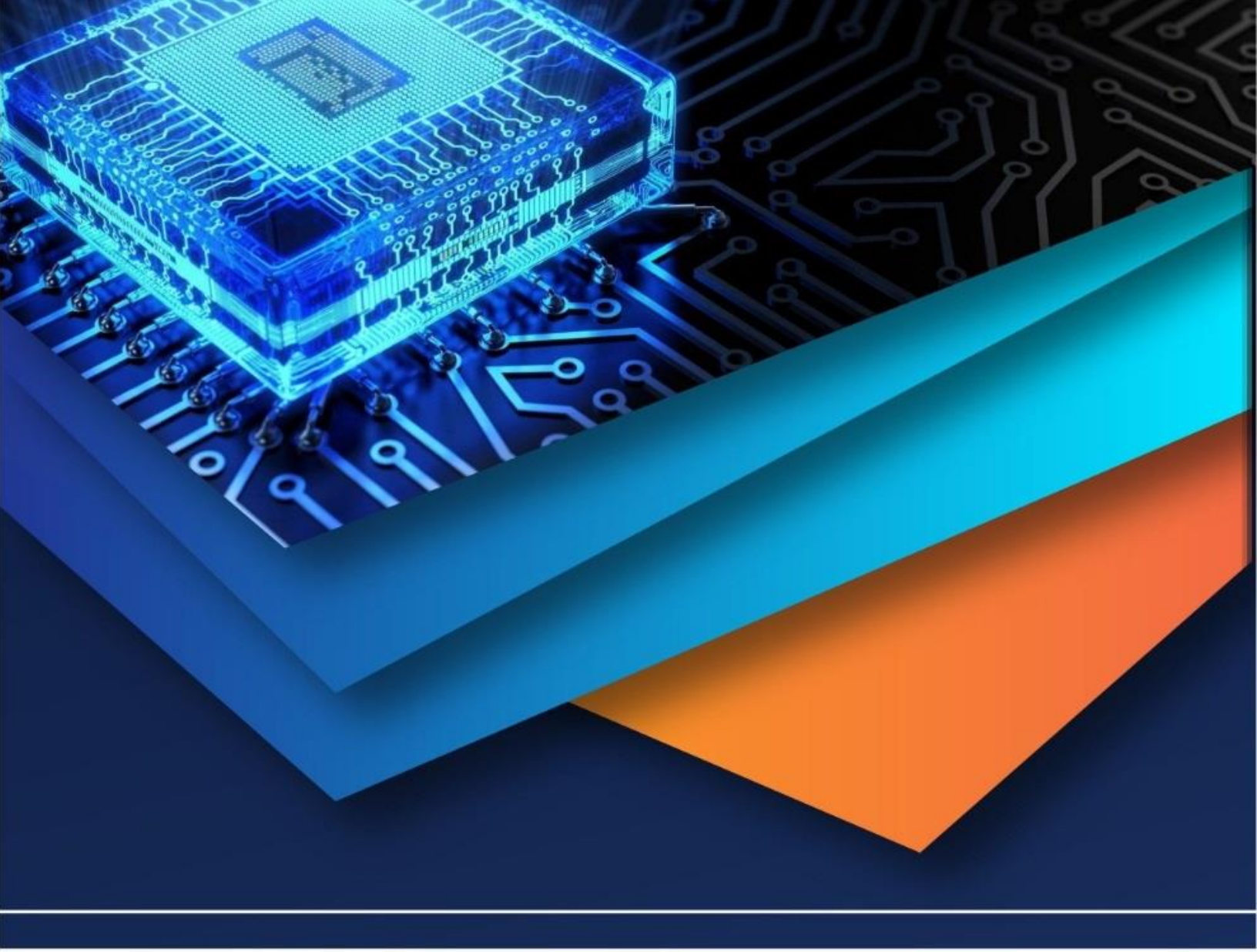


VI. ACKNOWLEDGEMENT

We would like to thank our project guide Dr. S. Umamaheswaran, Dept. of CSE,MVJCE for discusssions about the topic and providing valuable and helpful comments on drafts of the paper.

REFERENCES

[1] https://medium.com/@ageitgey/machine-learning-is-fun-80ea3ec3c471
[2] https://www.sas.com/en_us/insights/analytics/machinelearning.html#machine-learning-importance
[3] http://www.wired.co.uk/article/machine-learning-ai-explained
[4] https://deeplearning4j.org/ai-machinelearning-deeplearning 5. David E. Rapach , Jack K. Strauss " Forecasting real housing price growth in the Eighth District states"
[5] Sifei Lu, Zengxiang Li, Zheng Qin, Xulei Yang, Rick Siow Mong Goh, "A Hybrid Regression Technique for House Price Prediction", December 2017.
[6] Ayush Varma, Abhijit Sharma, Sagar Doshi, Rohini Nair, "House Price Prediction Using Machine Learning And Neural Networks", INSPEC number 18116205, April 2018.
[7] Adyan Nur Alfiyatin, Hilman Taufiq, Ruth Ema Febrita, Wayan Firdaus Mahmudy, "Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 10, 2017.
[8] Neelam Shinde, Kiran Gawande, "Valuation of House Price Using Predictive Techniques", International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835(IJAECS), Volume-5, Issue-6, June-2018.
[9] Li, Li, and Kai-Hsuan Chu. "Prediction of real estate price variation based on economic parameters." Applied System Innovation (ICASI), 2017 International Conference on.IEEE, 2017.
[10] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." Journal of machine learning research 12.Oct (2011): 2825-2830.
[11] Byeonghwa Park , Jae Kwon Bae (2015). Using machine learning algorithms for housing price prediction , Volume 42, Pages 2928-2934.
[12] Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, 2015. Introduction to Linear Regression Analysis



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY