International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

HOUSE RESALE PRICE PREDICTION USING CLASSIFICATION ALGORITHM

NILESH KHOT

Department of Masters in Computer Application, SIES College of management studies, Mumbai, Maharashtra, India

---

Abstract: *House resales are mainly seen in metro cities these days. Predicting the resale price of a house on a temporary basis for long-term residents is important, It is specially for people who intend to stay long-term, however are now not permanent, and people who do no longer wish to take any chances during the residence construction process.*

For the purpose of this study, the resale price prediction of the house is done by using various classification algorithms, including Decision trees, Naive Bayes as well as AdaBoost for boosting up weak learners.

There are several factors affecting the resale price of a house, including its physical characteristics, its location and the economy at the time.

Right here accuracy is taken into consideration as the performance metric for different datasets, and those algorithms are applied and as compared to discover the most appropriate technique for determining the resale rate by sellers.

Key Words *— Prediction, Decision Tree, AdaBoost, Naïve Bayes*

1.INTRODUCTION:

It is important you take into account the cost of the land, the existence of the building, the construction age, and the center availability when buying a resale flat.

The process of determining the actual market value of your property has to be followed, and one must recognize that information about the property, including its jurisdiction, taluka, division, and village, type of assets (land, residential, office, and store), and the size of the property.

Lets take an example a 20 years old flat with 1500 sq ft built up 1000sq ft UDS. According to current market prices, the land would cost 10,000/sqft if the flat were to be built today.

Land value = (uds * current price) = 1,000 * 10,000 = 1,00,00,000.

Construction cost = (built up area * construction cost) = 1500 * 1200 = 18,00,000

Total = 1,18,00,000 including all other charges

60 years is the average age of a building.

Depreciation is calculated so that at the end of 60 years the building has no value.

Generally, the residual value of scrap or salvaged assets is 10% of the asset value and may be ignored.

Over the next 60 years, 90% of the constructing costs will be spread out.

Therefore, the annual charge of depreciation is 1.5%.

Suppose a building is 20 years old and has a depreciation of 30%. i.e., The construction is 70% of its initial value.

If the value of the creation twenty years ago was 250/square foot, the cost of the resale flat would be as follows.

Price of land = 1,00,00,000

Construction cost = (1-(30/100)) * (1500*250) = 2,60,000

Overall cost of resale flat = 1,02,60,000

Because the flat would have been bought at a lower entry cost, at a lower land rate, and because of the lack of protection, the owner may be willing to accept a 30-40% discount.

If a person who holds a old flat (25-30 years+) wants to obtain the advantage as per the true valuation above. An excellent idea would be to carry out a joint construction project in coordination with each and every building's residents by hiring a builder or contractor.

2. RELATED WORK:

A resale is when a property is being sold by a person who was initially assigned the property. In other words, when a property is sold by the original purchaser, the property is said to be on sale.

Therefore, predicting the price of a home presents unique challenges because the value of a home depends on a variety of factors.

Though there are a lot of units dedicated to the present task, their performance and applications are limited by the extremely long delay in data handling, the lack of real-world settings, and so the lack of housing options.

Through this paper, we will try to predict the effectiveness of house marketing mistreatment classification techniques.

Currently, most studies look at how to break down the distraction of predicting house costs.

Researchers around the world produce different theories based on their analysis work. Economic predictors can use this paper to get the most recent results on prediction analysis.

In short, it summarizes prediction markets and present markets together, enabling easier prediction.

3. LITERATURE SURVEY:

Calculation of house prices uses several methods to predict the value of land and house.

Real estate has grown to be a pretty aggressive business since there are a lot of unstructured assets and documents.

The information mining process in such an enterprise empowers the builders via processing the information, forecasting future trends and supporting them to make profitable choices based on that understanding.

In this paper, the main focus is on the records mining approach, which is used to broaden a model that predicts the most suitable place of real estate based totally in your hobbies and ranks the most popular properties in accordance with their ranking.

A ranking approach is used to predict a favorable region.

Based on the customer's location selections, it analyses the data.

It extensively works on fundamental stages.

The primary phase ranks a group of customers defined locations to find a perfect area and the second one section predicts the most suitable place in line with their requirements and interest.

It uses a classification technique called Naïve bayes and tries to predict the price of houses.

Drawbacks: It does not estimate future costs on properties mentioned by the consumer.

Because of this, the hazard in investment in an rental or a place will increase notably. To decrease this error, clients tend to hire an agent which once more will increase the price of the method. This leads to the amendment and development of the existing machine

The accuracy of prediction is defined as: Accuracy rate= (The total number of prediction / Total number of sample sets) * 100

4. PROBLEM STATEMENT:

An important limitation of Naive Bayes is the assumption of independent predictor features. It assumes that all attributes are independent. Here, in

our example any attribute for calculating price cannot be independent because for predicting the price we need the size of the area, number of rooms, etc. then only we can predict the price of the house.

If we are using decision tree for prediction of house price and if there is a small change in the data then it may lead to large change in our prediction. Most often they are inaccurate.

And also for weak classifiers it can lead to low margins.

5. ALGORITHM:

Decision Tree:

In general, Decision tree analysis is a predictive modelling tool that can be applied across many areas. Decision trees can be constructed by an algorithmic approach that can split the dataset in different ways based on different conditions. Decisions trees are the most powerful algorithms that falls under the category of supervised algorithms. They can be used for both classification and regression tasks. The two main entities of a tree are decision nodes, where the data is split and leaves, where we got outcome. The example of a binary tree for predicting whether a person is fit or unfit providing various information like age, eating habits and exercise habits, is given below –

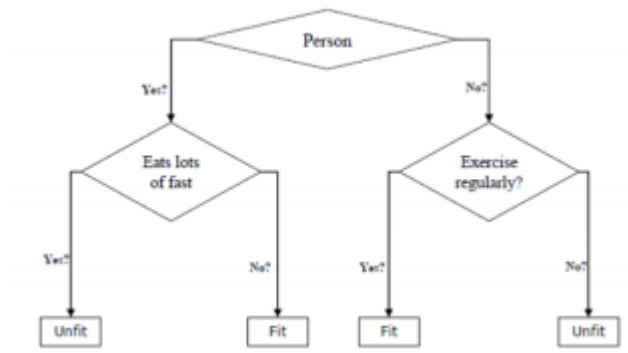

Fig 1. Decision Tree

Decision tree is used to make a particular decision on price prediction. Like, to predict for the future price the decision tree can be helpful.

Naïve Bayes:

Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes' theorem is stated mathematically as the following equation:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where A and B are events.

- Basically, we are trying to find probability of event A, given the event B is true. Event B is also termed as evidence.

- P(A) is the priori of A (the prior probability, i.e. Probability of event before evidence is seen). The evidence is an attribute value of an unknown instance(here, it is event B).

- P(A|B) is a posteriori probability of B, i.e. probability of event after evidence is seen

Now, with regards to our dataset, we can apply Bayes' theorem in following way:

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)}$$

Naïve Bayes is the algorithm that is used to generate sentiment score. And on the basis of this sentiment score the prediction of stock is done. Hadoop Map Reduce is used for managing the large amount of big data.

6. RESULT AND DISCUSSION:

Size of house	No. of bedrooms	No. of bathrooms	Price of house	Level of efficiency
950	2	1	1500000	Efficient
1500	2	2	3000000	Very Efficient

2222	2	2	4000000	Very Efficient
3275	3	1	5000000	Very Efficient
4150	3	2	10000000	Very Efficient
1435	1	1	2500000	Efficient
5220	4	2	15000000	Very Efficient
2005	2	1	3500000	Efficient

NAÏVE BAYES:

In the above table, there are some values for house price and we need to predict value for that.

P(Level of efficiency="Efficient") = 3/8 = 0.38

P(Level of efficiency="Very Efficient") = 5/8 = 0.63

So for predicting value for 2 bedroom and 2 bathroom that whether for that the price for the house is efficient or not.

For predicting the values for 2 bedrooms and 2 bathrooms,

P(Bedroom="2" | Level of efficiency="Efficient") = 2/3 = 0.67

P(Bedroom="2" | Level of efficiency="Very Efficient")=2/5=0.4

P(Bathroom="2" | Level of efficiency="Efficient") = 0/3=0

P(Bathroom="2" | Level of efficiency="Very Efficient")=4/5=0.8

P(X | Level of efficiency="Efficient") = 0.67 * 0 = 0

P(X | Level of Efficiency="Very Efficient") = 0.8 * 0.4 = 0.32

P(X|Ci) * P(Ci) → For efficient

0*0.38 = 0

→ For Very Efficient

0.32 * 0.63 = 0.20

P(Level of efficiency="Very Efficient") > P(Level of efficiency="Efficient")

So, from the above prediction it is proved that for **"Very Efficient"** value the probability value is more.

Therefore, Level of efficiency is **Very Efficient.**

DECISION TREE:

For Level of Efficiency:

Info = I(3,5)= -3/8 Log(3/8) – 5/8 Log(5/8)

= 0.38*0.43 + 0.63*0.20

= 0.16 + 0.13

= 0.29

For bedroom,

For 1 bedroom,

Info(bed)D = 1/8 I(1,0)

= 013 * 0

=0

For 2 bedroom,

Info(bedroom) D = 4/8 I(2,2)

= 0.5 * [0.5 + 0.5]

= 0.5

For 3 bedroom,

Info(bedroom) D = 2/8 Log(0,2)

= 0

For 4 bedroom,

Info(bedroom) D = 1/8 Log(0,1)

= 0

For bathroom,

For 1 bathroom,

Info(bathroom) D = 4/8 I(3,1)

= 0.5 * [3/4 Log(3/4) -1/4 Log(1/4)]

= 0.5 * [0.8+0.12]

= 0.5 * 0.92

= 0.5

For 2 bathroom,

Info(bathroom) D = 4/8 I(0,4)

= 0.5 [0/4 Log(0/4) – 4/4 Log(4/4)]

= 0.5 (1 * 0)

= 0

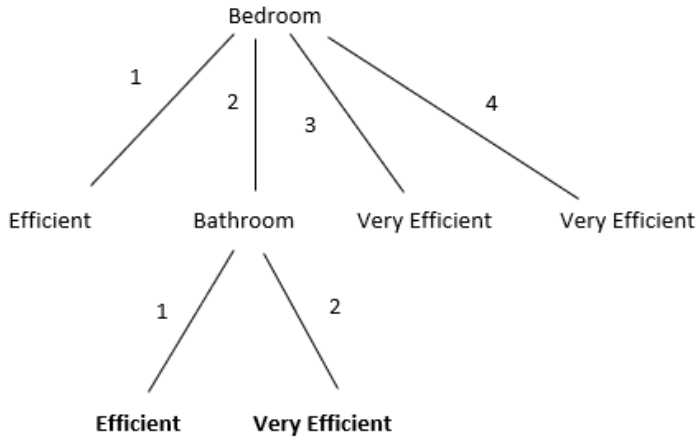

Fig 2: Decision Tree Example

We can draw a decision tree for this if we want to make a decision for how many bathroom and bedroom is efficient for us then this decision tree is helpful.

7.CONCLUSION:

In today's material world, it has become increasingly difficult to keep track of such big data and to bring it out in personal way. Also, the data that is extracted is useful.

The system makes good use of the Naïve Bayes Algorithm and decision tree. The system uses such data in the most efficient way. The naïve bayes algorithm helps to reach customers by increasing the accuracy of asset selection and reducing the risk of investment in the asset. And decision tree helps to make a proper decision for number of rooms. Many features can be added to make the program more widely accepted. One of the great future gaps is adding asset information to many cities that will allow the user to explore multiple locations and come to the right conclusion. Other factors such as the economic downturn in housing prices will be added. This will help the system to operate at a higher level.

8. REFERNCES

[1] https://www.geeksforgeeks.org/decision-tree/

[2] Data Mining concepts and techniques second edition by Jiawei Han and Micheline Kamber

[3] https://www.geeksforgeeks.org/naive-bayes-classifiers/

[4] https://www.slideshare.net/Shrutisoodhi/how-to-calculate-resale-value-of-house

BIOGRAPHIES

Nilesh B. Khot
Department of Masters in computer Application
SIES college of management studies & technology, Mumbai-400706