International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 03 | Mar 2019 www.irjet.net p-ISSN: 2395-0072

HOUSE PRICE PREDICITION USING MACHINE LEARNING

Atharva Chouthai[1], Mohammed Athar Rangila [2], Sanved Amate[3], Prayag Adhikari[4], Vijay Kukre[3].

[1,2,3,4] Student, Dept. of Computer Engineering, A.I.S.S.M.S. Polytechnic Pune, Maharashtra, India
[5] H.O.D, Dept. of Computer Engineering, A.I.S.S.M.S. Polytechnic Pune, Maharashtra, India

---

Abstract - *We are predicting the deal price of the houses using various machine learning algorithms. Housing sales price are determined by numerous factors such as area of the property, location of the house, material used for construction, age of the property, number of bedrooms and garages and so on. This paper uses machine learning algorithms to build the prediction model for houses. Here, machine learning algorithms such as logistic regression and support vector regression, Lasso Regression technique and Decision Tree are employed to build a predictive model. We have considered housing data of 100 properties. Logistic Regression*

Key Words: Real Estate, Prediction Model, Linear Regression.

1. INTRODUCTION.

Real Estate Property is not only the basic need of a man but today it also represents the riches and prestige of a person. Investment in real estate generally seems to be profitable because their property values do not decline rapidly. Changes in the real estate price can affect various household investors, bankers, policy makers and many. Investment in real estate sector seems to be an attractive choice for the investments. Investment is a business activity that most people are interested in this globalization era. There are several objects that are often used for investment, for example, gold, stocks and property. In particular, property investment has increased significantly since 2011, both on demand and property selling.

2. EXISTING SYSTEM.

There are several approaches that can be used to determine the price of the house, one of them is the prediction analysis.
The first approach is a quantitative prediction. A quantitative approach is an approach that utilizes time-series data [5]. The time-series approach is to look for the relationship between current prices and prevailing prices. The second approach is to use linear regression based on hedonic pricing. Previous research conducted by Gharehchopogh using linear regression approach get 0,929 errors with the actual price. In linear regression, determining coefficients generally using the least square method, but it takes a long time to get the best formula. Particle swarm optimization (PSO) is proposed to find the coefficients aimed at obtaining optimal result . Some previous researches such as Marini and Walzack show that PSO gets better results than other hybrid methods. There are several advantages of PSO, in the small search space PSO can do better solution search. Although the PSO global search is less than optimal , but on the optimization problem the value of the variable on the regression equation can find a maximum solution using PSO

3. PROPOSED SYSTEM.

The land prices are predicted with a new set of parameters with a different technique. Also we predicted the compensation for the settlement of the property. Mathematical relationships help us to understand many aspects of everyday life. When such relationships are expressed with exact numbers, we gain additional clarity
Regression is concerned with specifying the relationship between a single numeric dependent variable and one or more numeric independent variables.
House prices increase every year, so there is a need for a system to predict house prices in the future. House price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house.

4. PROBLEM STATEMENT.

Last week, James Wilcox, a real estate appraiser in West Palm Beach, Florida, received a call from Madera & Associates, a law firm in nearby Boca Raton, Florida. Madeline Madera explained that she represents a client who is seeking compensation after discovering urea formaldehyde (also known as urea-methanal) foam insulation (UFFI) in a recently purchased home. Ms. Madera needed an assessment of the impact of UFFI on property values to support their client in pending litigation. Urea Formaldehyde Foam Insulation (UFFI) was a compound used in the 1960's and 1970's to improve the insulation quality of older homes. It was found that UFFI emitted the potentially hazardous formaldehyde gas as it decayed through time, and as a result the insulation fell out of favor with builders in the 1970s and, in fact, was permanently banned in Canada and Europe in the 1980s. Despite the ban, there remained considerable controversy concerning the health and financial risks of UFFI insulated homes – different independent studies reached conflicting conclusions about the health risks of UFFI homes, while

various court cases provided no precedent for determining lost real estate value when such properties were sold without disclosure of UFFI presence.

When appraising real estate property value, James Wilcox generally uses a combination of three traditional methods plus an analytical method developed by him and his associates. The three traditional methods used are:
Replacement Value: Determine the value by estimating how much it would cost to build a similar building in a similar area.
Net Rental Value: Determine the net present value of cash flows generated by renting the property.
Market Value: Compare the property with similar properties in the area that were recently sold.
As an alternative to the standard appraisal techniques George had been experimenting with quantitative analytical methods. To support these methods, his firm built a database of real estate sales which includes available comparable features of each property. For a particular valuation request – depending on the property type, availability of data, and time constraints – George would provide appraisal reports that included a combination of traditional methods and estimates derived from his proprietary analytics. His clients put trust into his valuations as they are generally quite accurate, particularly compared to appraisals provided by his competitors.
Ms. Madera's client had recently purchased a residential property in the Norwood section of West Palm Beach, an area with many older homes. While her client was generally happy with the property, she had not been informed of the presence of UFFI by the previous owner, or by the owner's real estate agent. The client was not worried about potential health risks associated with UFFI, but believed she had overpaid for the property. Ms. Madera agreed with this claim for two reasons: 1) They were aware of the lingering negative stigma associated with UFFI homes, and 2) The decrease in sales value was the only reason UFFI presence would not be disclosed by the previous owner or sales agent. While they were willing to represent the client and were confident in the case, they needed demonstrable evidence of the value lost because of the non-disclosure.
James informed Madeline that he believed he could supply them with an analysis and estimated valuation, as he had a database of residential home sales in Palm Beach County – both with and without UFFI – collected over the previous decade. The client's UFFI litigation would begin in less than two weeks, so James must get to work immediately to generate his appraisal

5. LINEAR REGRESSION ALGORITHM

Regression is concerned with specifying the relationship between a single numeric dependent variable (the value to be predicted) and one or more numeric independent variables (the predictors). As the name implies the dependent variable depends upon the value of the independent variable or variables. The simplest forms of regression assume that the relationship between the independent and dependent variables follows a straight line.
The origin of the term "regression" to describe the process of fitting lines to data is rooted in a study of genetics by Sir Francis Galton in the late 19th century. He discovered that fathers who were extremely short or extremely tall tended to have sons whose heights were closer to the average height. He called this phenomenon "regression to the mean".

5.1 Understanding Regression.

You might recall from basic algebra that lines can be defined in a slope-intercept form similar to y = a + bx. In this form, the letter y indicates the dependent variable and x indicates the independent variable. The slope term b specifies how much the line rises for each increase in x. Positive values define lines that slope upward while negative values define lines that slope downward. The term a is known as the intercept because it specifies the point where the line crosses, or intercepts, the vertical y axis. It indicates the value of y when x = 0. Regression equations model data using a similar slope-intercept format. The machine's job is to identify values of a and b so that the specified line is best able to relate the supplied x values to the values of y. There may not always be a single function that perfectly relates the values, so the machine must also have some way to quantify the margin of error. We'll discuss this in depth shortly.

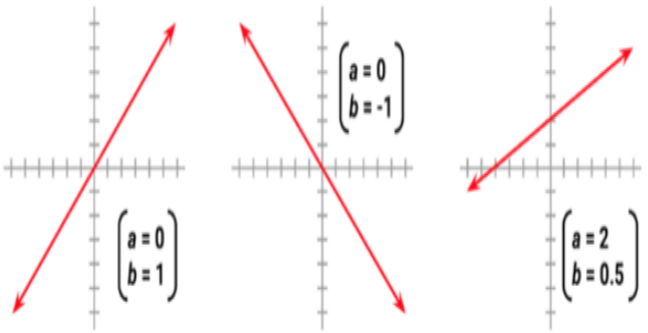

Fig -1: Slope Intercept Form.

6. SYSTEM ARCHITECTURE.

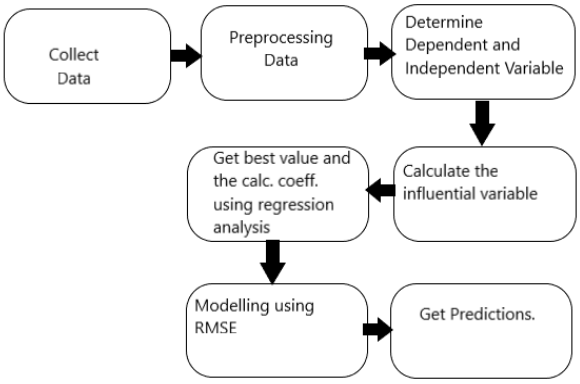

Fig -2: System Work Flow.

7. OUTPUTS.

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    99 obs. of  12 variables:
## $ Observation   : num  37 79 75 32 69 4 28 30 18 63 ...
## $ Year Sold     : num  2009 2009 2011 2011 2010 ...
## $ Sale Price    : num  76900 78000 79000 80000 82000 84000 84000 84000 85000 85000 ...
## $ UFFI IN       : num  1 1 0 0 1 1 0 0 0 1 ...
## $ Brick Ext     : num  0 0 0 0 0 0 0 0 0 1 ...
## $ 45 Yrs+       : num  1 1 1 1 1 1 1 1 1 1 ...
## $ Bsmnt Fin_SF  : num  0 154 400 0 157 ...
## $ Lot Area      : num  2772 4490 5840 5040 5441 ...
## $ Enc Pk Spaces : num  0 0 0 0 0 1 2 0 0 1 ...
## $ Living Area_SF: num  1018 536 721 513 672 ...
## $ Central Air   : num  0 1 1 0 0 0 0 0 1 0 ...
## $ Pool          : num  0 0 0 0 0 0 0 0 0 0 ...
```

Fig -3 : Data Set.

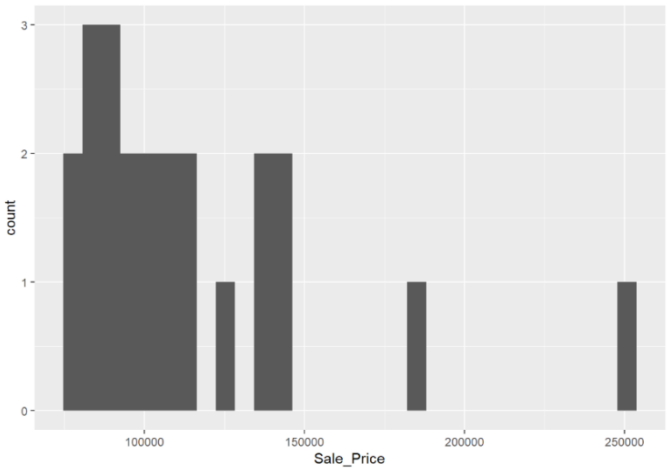

Fig -4: Histogram of UFFI and Price

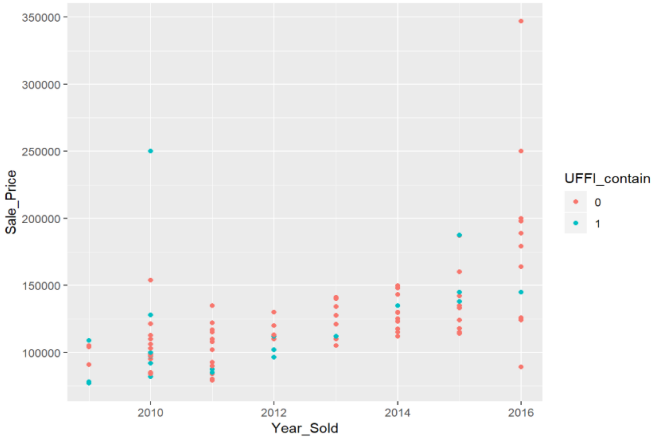

Fig -5: How Prices vary when UFFI is present.

Fig -6 : Variance of Price when UFFI is present using BarChart.

8. CONCLUSIONS.

We have used a data set of 100 houses with a number of parameters. We have used 50 percent of the data set to train the machine and 50 percent to test the machine. The results are truly accurate. And we have tested it with different parameters also. Not using PSO makes it easier to train machine with complex problems and hence regression is used.

ACKNOWLEDGEMENTS

The authors would like to thank Mr. S.K. Giram Principal, of A.I.S.S.M.S. Polytechnic Pune.

REFERENCES

House Price Prediction: Hedonic Price Model vs. Artificial Neural Network. Visit Limsombunchai Commerce Division, Lincoln University. P. Acharjya: A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools.

Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization. Adyan Nur Alfiyatin. (IJASCA) Volume 8 Nov 10 , 2017.

Real Estate Price Prediction with Regression and Classification. CS 229 autumn 2016 Project Final Report. Hijua Yu, Jiufa Wu.