

A Machine learning based Advanced House Price Prediction using Logistic Regression

Ravula Bala Siva Krishna
Pursuing Final Year,
Computer Science and
Engineering
KLEF Deemed to the University
Vaddeswaram, Andhra Pradesh

Kunamneni Surya Kumar
Pursuing Final Year,
Computer Science and
Engineering
KLEF Deemed to the University
Vaddeswaram, Andhra Pradesh

Tadiparthi Chandravas
Pursuing Final Year,
Computer Science and
Engineering
KLEF Deemed to the University,
Vaddeswaram, Andhra Pradesh

Manikandan J.
Assistant Professor
CSE Department
KLEF deemed to the University
Vaddeswaram, Andhra Pradesh

ABSTRACT

Machine learning has played a bigger role than in previous years' image detection, spam changes, normal conversation orders, product recommendations, medical diagnoses. Current machine learning algorithms help to improve security alarms, public safety and medical improvement and other advantages using machine learning penetration in to several global implementations and their improvement. The machine learning system provides better opportunities customer service and safe car system. Currently A paper on which we talk about future housing price forecasts machine learning algorithms. As for Choose from our comparative and various study hypotheses forecasting methods. We use refined-lasso regression as our model due to the method of adaptation and verification on the model selection. Research shows that we need a problem-solving approach able to succeed and develop assumptions It will be possible to compare the cost of the house with other models. On the other hand, the housing price index and progress forecast of housing costs that tend to make real progress real estate policy schemes. This study uses machine learning The algorithm is used as a research method to develop housing prices predictive models. We create a model to calculate the cost of housing an example of a machine learning algorithm, e.g. gradient-boosting framework, refined-lasso regression and the machine learning based system Execute orders accurately. We recommend building an apartment at that time a cost estimate model to support a home seller or real estate an agent for better information based on house valuations. These tests show that there is a lasso regression algorithm The appearance of accuracy and reliability of other models Implement preliminary housing cost estimates

Keywords

gradient-boosting framework, refined-lasso regression, machine learning, advanced predicting techniques.

1. INTRODUCTION

A similar machine learning mechanism The previous experience plays the same important role as the use of animals gain experience and knowledge in food safety detection. In the voice If the knowledge about food is negative, you will

make a mistake If there is a label, there will also be assumptions about the animal adverse effects and may occur in the future. Hamt Inspiration from previous examples of successful learning we show a typical machine learning algorithm. We want to program a learning machine as well How to filter spam emails[5]. The machine will remember the email that was there before called spam emails to people as needed user. The machine will search when another email arrives This has been established in the past about spam emails. Although it fits Among them, it will be trash. Otherwise it will do moved to the user's incoming email organization. At the same time the first "learning by heart" The methodology can be useful, it does not offer significant The so-called field of the training system - the capacity to record invisible emails. There must be an effective learner the ability to progress beyond uniqueness make the sample more comprehensive. It could be Similarly, inductive thinking or inductive based induction. The emotional sensations previously shown after rats to take certain types of food samples; they apply A new and invisible example is dealing with them Require comparable nutrition and taste. The spam judge needs to consolidate the task The student can check previously viewed emails extricate a. is about expressing existence Email is a spam feature. Then, on another When the email arrives, those machines can be weighed continuously among suspicious expressions gives an idea of it properly anticipates its sign. There may be such an arrangement be able to preview the name about unseen emails. Responsibilities other than human competence: a an additional team that declares to make a profit first The machine acquisition system is defined as follows Very thorough and thorough investigation data set: galactic data, modified reconstructed lever rehabilitation knowledge, weather forecasts, and Unpacking genome data, Web series engines, as well e-commerce. It is possible to increase with increasing volume will be numerically recorded data Most of the critical information was collected displays the most optimal information along with the best statue too little and also bothers people good bode about[8]. Take a serious example A significant set of information can also be a significant provide guarantees for their mixed projects to get an unlimited amount of unlimited memory Expanding the conversion speed of PCs

opens up new possibilities layers. Unsupervised monitoring since entry learner, environment, Suggestions that can be taken separately should be discarded tasks defined by the nature of the connection. Professional start to Contrast the among regulated and uncontrolled acquisitions Give an example, and think about preparing for it spam emails must be identified ignore recognition. We are for spam detection Consider making the settings The learner receives prepare an email mark spam / not spam, like wise house price prediction helps in fast growing cities in order to help end users in predicting the prices based on several parameters.

2. LITERATURE SURVEY:

The latest global financial crisis has recovered sharply Inspired by both research and strategy The portion of the cost of the asset and the amount of the cost of the special location overlap parallel to the movement of money. Lamer (2007) noted, The exhibition of these hotels predicted eight out of ten post offices in the world Drama ii crisis and exemplary drama She is a real segment of the economy. Tell the truth he is diving "Housing is the benefit of the business cycle. Vargas and Silva (2008) argue that housing is costly changes play an important role in determining business cycle phase. When that economy work in the growth, development and housing department The demand should increase significantly. quickly pushes the cost of the house. By the way The phase of attracting money and the decline of personal money decreases aggravates the interest. Amazingly The number of family members gradually decreases after leaving home do not want to lower the cost. Most of them The deal will slow down as it intensifies The price is cheaper than the discount development section and building skills. Moreover, it is true during liquidation and settlement The house price will fall quickly Similarly, general inflation The style makes the cost of a real house sticky recognized costs. Recently, few writers have decided to experiment found that the cost of the house made it possible to mould the tools determine the yield. (Forney, 2003; stocks and Watson, 2003; Gupta Further Das, 2010; das et al., 2009; 2010; 2011; Gupta to Hartley, 2013). Those people The development department is extensive and speaks only Comprehensive monetary measures linked to GDP. Therefore, when it comes to decoration, it is a a house with a wide range of general economic wealth Cost fluctuations can serve as an indication of development On GDP (Case et al., 2005). As for the picture, it is materials with different properties and these certificates The development of the cost of the house can be an addition an indication of the future direction of expansion (Gupta And Kabundi, 2010). In general, the exact definition The development method that results from invoicing the cost of a house is a on house business members and budget strategic authority. There is a huge amount of literature about houses in the United States price. Roach Further (2007) uses cars regressive diffuse (ARDL) model range, The hypothesis possesses 25 identifying capacities with real premises cost development in selected unique states Resource Zone 8. They found ARDL models The tend should hit the reference AR model. reach and Stress (2009) added to these The test on the largest 20 u. Acquainted with the city due to the development of industrialization and innovation government representative. Countries dependent on the ARDL model View state, territory, and state level variables. Again, the creators will cover the comparison The preview of the models who are joining weak alternative structure. Gorgas and Prodigies (2011) use hazard premiums Similarly, they are relatively different among different people long-term motivation level and fulfils the wishes of the delegates as an information

variable that can predict future flight levels What do you do for a living? They draw conclusions about it gurus Investigators also used most of these items appropriately provided in accordance with those investment risk premium information therefore, the probability of claiming must be assessed today Get the S&P CS-10 list for three months ahead. Gupta and Das (2010) also predict a recent decline At the price of the largest real price increase in the United States in twenty years provinces. Authors use spatial Bayesian VARs (BVARs), based only on the months of real house price increases Forecast for the decline in the time period from January 1, 2007 to 2008: 01. They believe that BVAR models are well equipped although predicting the future direction of the actual house The decline is highly valued. They determine this is a presumption that BVAR models are lacking information on the rationale in the settlement process. reach and Stress (2009) expand individuals equally The exam on the 20 most amazing. -I met you urban decay associated with industrialization and progress developed and administrative. Reliable countries ARDL models are gander in the state as well as in the region national level variables. Once again, he will be a designer A similarity is the aggregation of forecasts about independently structured models. Gog's and prairies (2011) use the maximum level of danger The uniqueness of the white colour was similar of all levels of energy and agents data anticipate what will move into the cost of the house.

3. PROPOSED APPROACH:

3.1 Linear regression:

The simple linear regression statistical method allows the study of the relationship between two continuous numerical variables. A variable called x is considered a predictive, explanatory, and independent variable. Other variables denoted by y are considered to be response, result, or dependent variables.

3.2 Multiple Regression Analysis:

Multiple regression analysis is used to check for a statistically significant relationship between multiple variables. Used to detect these patterns personal information package. Numerous relapses Investigations will be almost identical. The main difference between straight-line relapses is that there are also multiple relapses in the number of predictors (variable " x ") that use those relapses. Straight iteration analysis Each sub-variable " y " has an absolute x variable. Point of use: (x_1, Y_1). Numerous " x " variables are used for each free variable: (x_1) 1, (x_2) 1, (x_3) 1, Y_1). For a single-variable linear regression, you can inform yourself about a specific case-dependent variable (for example, "Sales"). You may be interested to see how the diversity of claims suggests a recurrence. You can make your X_1 clear from the sale.

3.3 The cost Function:

So hopefully you have enlarged the size of a particular store, and you said that these deals are likely to be high. Although the size was expanded, the cheap size in the store did not expand. Link these costs together In the past, expanding the size of the store has had negative consequences for you. Therefore, we need to keep these costs to a minimum. Therefore, we present a cost function that the template is primarily used to slide and measure these pages.

$$\begin{aligned} PR &= \frac{TP}{TP+FP} \\ RE &= \frac{TP}{TP+FN} \\ CA &= \frac{TP+TN}{TP+TN+FP+FN} \\ F_1 &= \frac{2TP}{2TP+FP+FN} \end{aligned}$$

$\hat{\pi}_{i-1}$	Precision of the prediction onto the level below
π_i	Posterior precision at the current level
δ_{i-1}	Prediction error for the volatility of the level below

3.4 Gradient-boosting framework, refined-lasso Regression:

Information on Lasso regression will be studied. Furthermore, it is possible to sample the regression model, and the formula is added as a reference. LASSO acts as the operator of the minimum absolute contraction and selection. Lasso regression is one of the stabilization regimes that leads to the modelling of the surrounding area with a number of unique features, and neither of the following applies. • Enough to enhance fashion extremes. A minimum of ten variables can serve as the basis for the fastening. • Enough is enough to do a computational test. This situation may be due to numerous or billions of claims for features. The tester recurrence performs the L1 adjustment, which includes a penalty equal to the maximum value of the coefficients. Here's a few basic facts about a stomp pad and how it is used. Reduction goal = LS OBJ + λ (sum of the values for which the coefficient is estimated). LS OBJ remains in place It is the goal of direct recurrence without stabilization, and the goal of the smallest square. The size will increase with the expanding ones, and the difference will decrease (λ) and the growth scale will decrease. The Lasso regression calculation is defined as follows

$$\Delta\mu_i^{(k)} = \frac{1}{2} \kappa_{i-1} v_{i-1}^{(k)} \frac{\hat{\pi}_{i-1}^{(k)}}{\pi_i^{(k)}} \delta_{i-1}^{(k)}$$

Here, the rotating component controls the quality of the fine, ie. When $\Lambda = 0$, we get the same direct coefficient. When $\Lambda = \infty$: the coefficient is always zero. The point at which the coefficient is taken between $0 < \lambda < \infty$: 0 is the most important thing to have a basic over-frequency, and we set it under the two plans when the extreme is between the two poles. Connect Get a straight model X at a time. Contract these coefficients.

3.5 Linear Regression based Gradient Boosting algorithm:

Linear Regression and Gradient growth is a strategy for machine abuse. Also, organizational issues create a predictive model rather than a grouping of powerless predictive models. There are two ways to help make a predictive model more accurate. It is possible to take turns building a building with special features. To implement the calculation directly. There are many calculations to make. These include: Gradient Boosting XG Boost Ada Boost Gentle Boost. Each algorithm needs its own math to improve.

$$\Rightarrow \Delta\mu_i \propto \frac{\hat{\pi}_{i-1}}{\pi_i} \delta_{i-1}$$

Implemented formula for valid prediction And a little more can be seen when using them. Increasing the computation is important among these. Most of them are able to get the thoughts they have been acquainted with for the last twenty years. Even if it is intended to order the problem, everything that can be developed for it must also be a relapse. Inspiration to increase the gradient can be a technique. It can produce a large number of "loose" classifiers. These develop "competent" committees. "A non-electrical classifier (e.g., G. selection tree) is one in which the sliding speed has an advantage over the irregular assumption.

4. IMPLEMENTATION:

4.1.1 Execution And Outputs:

When the code starts to run, we get the results, and then the assumptions. These conspiracies help us understand the relationship between the target variable (price) and the variables of different predictors. This area shows a graphical table showing the number of bedrooms and houses. The number of 3-bedroom houses was considered high and the number of 7-bedroom houses was considered to be the lowest

4.1.1 Input Datasets:

	id	date	price	bedrooms	bathrooms
0	7129300520	20141013T000000	221900.0	3	1.00
1	6414100192	20141209T000000	538000.0	3	2.25
2	5631500400	20150225T000000	180000.0	2	1.00
3	2487200875	20141209T000000	604000.0	4	3.00
4	1954400510	20150218T000000	510000.0	3	2.00

4.1.2 Splitting data:

```
#Splitting the dataset into training a testing set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state = 0)
```

```
regressor_LR = LinearRegression()
```

```
regressor_LR.fit(X_train,y_train)
```

4.1.3 Implementing Linear Regression:

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

```
y_pred_lin = regressor_LR.predict(X_test)
print (y_pred_lin)
```

```
[ 384885.98850099 1522091.68631129 541596.44539849 ... 320303.71125759
 224536.90644671 140903.44896095]
```

5. RESULTS

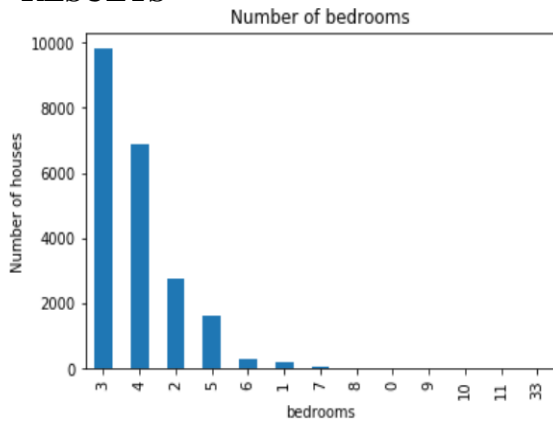


Fig – 5.1 : No. of houses vs. no. of bedrooms

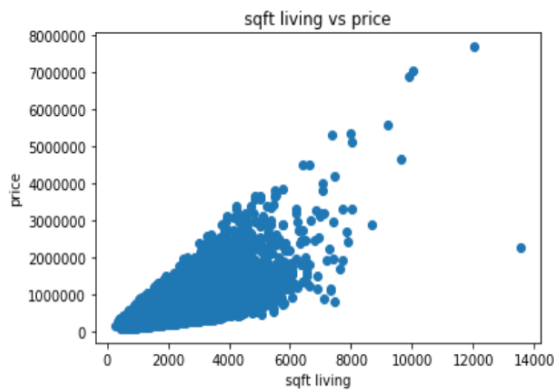


Fig-5.2: Price vs. Square feet living

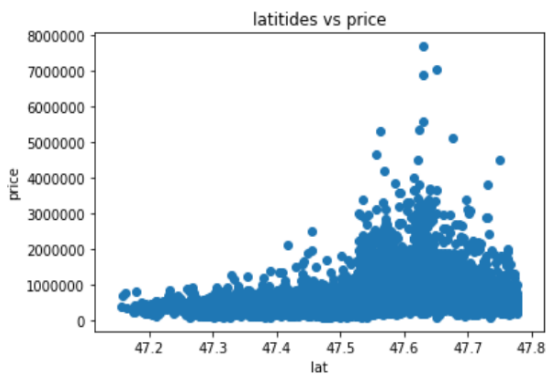


Fig5.3: Price vs. Latitude values

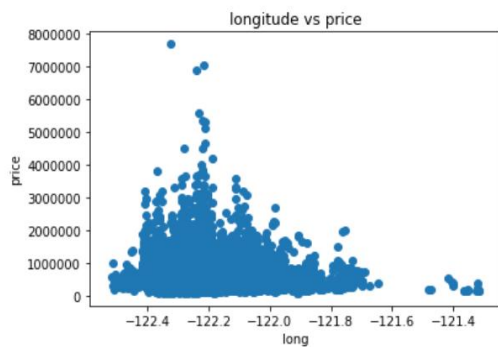


Fig – 5.4: Price vs. Longitude values

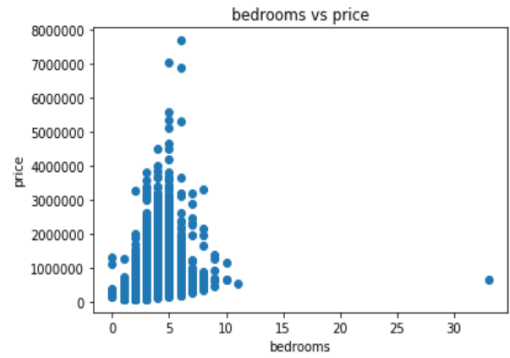


Fig –5.5 Price vs. Bedrooms

5.1.1 Confusion Matrix Results:

```
plt.plot(y_test)
plt.plot(y_pred_lin)
plt.xlabel('count')
plt.ylabel('TRUE AND PREDICTED VALUES ')
plt.show()
```

5.1.2 Predicted Results:



Fig 5.6 Sqft living vs Price

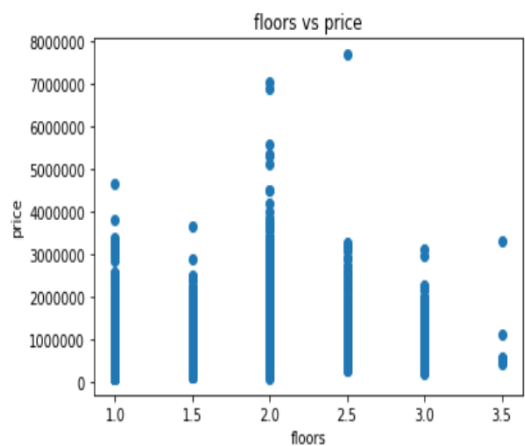


Fig 5.7 Floors vs Price

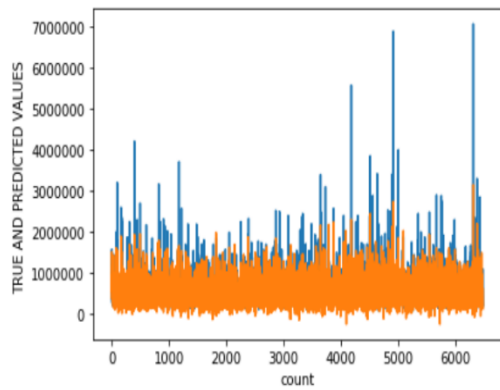


Fig 5.8 True and Predicted Values

6. CONCLUSION

Looking at the future value assumptions, we decided how to prepare a model that would give our customers the best approach to acquiring gander. Overcoming strategies have been studied. This was compared when a predictive strategy was developed with ML support. Previous straight works are used in our model, and future value assumptions tend to be worth all the senses. We implemented the idea of requiring a gradient based regression techniques increase using a method similar to the use of information as time allowed for the prediction system. Hosting incentives have made all the effort to meet our presentation requirements, so different updates can be made later. These are included We have not updated due to time constraints. For the predictive range, the real concern may be the stockpile phase. In addition, our data set must be prepared in no more than one day. In addition to sequential calculations, we can use different processors and similar calculations, which may reduce the preparation time. Our experimental results shows that the prediction rate improve upto 2.8% based on the traditional predicting

methodologies, If more features are included according to the model, customers with a location-friendly location should be listed and a high-temperature map should be provided and can be used by majority of rental and house purchasing promotion application in future.

7. 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”
- [6] Gupta and Das (2010) Forecasting the US Real House Price Index: Structural and Non-Structural Models with and without Fundamentals
- [7] Rangan Gupta “Forecasting US real house price returns over 1831– 2013: evidence from copula models”
- [8] Number of homes sold in the United States between 1995 and 2016. Wwww.statista.com. Entered: 11/10/2017.
- [9] Quick Fact: Resident Demographics. National Council on Multi-Housing. Accessed:11/11/2017.
- [10] Random memories of Leo Breiman and Adele Cutler.
- [11] Sherwin Rosen. “Hedonic Prices and the Underground Market: Product Differences Pure competition. ’ Submitted by: Journal of Political Economy 82.1 (1974).