RESEARCH ARTICLE OPEN ACCESS

Estimating house Rent using machine learning over Cloud Computing

Shubham Badyal, Raminder Kaur
Research Scholar Center for cloud infrastructure and security Suresh GyanVihar University, Jaipur
Assistant Professor Dept. of Computer Science Suresh GyanVihar University, Jaipur
Corresponding Author; Shubham Badyal

ABSTRACT—Property is the most transparent industry in our ecosystem. Home prices vary from day to day and sometimes depend on the price. Predicting accommodation with the main reasons is the main crux of our research project. Our values are targeted here based on each of the key parameters considered during pricing. The decline in house prices has attracted interest from researchers as well as a number of other interested parties. There are many previous studies that used different regression techniques to address the question of house price changes. It considers home price changes as a classification problem, and applies machine learning strategies to determine if home prices will rise or fall. It employs specialty selection techniques such as differential impact factor, data quality, policy component analysis and data transfer techniques such as external and missing value treatment as well as box-to-cox exchange techniques. The efficiency of machine learning techniques is measured by four parameters of precision, accuracy, precision and sensitivity. We use different regression techniques in this way, and our results are not only a strategy, but a weighted way of different strategies to deliver accurate results.
Keywords—House Price Predction, linear regression, machine learning, prediction, parameters, boosted regression, forest regression, neural network.

DATE OF SUBMISSION: 06-11-2019 DATE OF ACCEPTANCE: 26-11-2019

I. INTRODUCTION

The progress of civilization is the basis of the increasing demand for housing by the day. Accurate home mortgage forecasts have always been a boon for buyers, sellers and bankers alike. Many researchers have already worked to uncover the mystery of predicting home prices. Many similar ideas were born as a result of research work that has been participated by various researchers around the world. Some of these theories believe that the geographical location and culture of a particular place determines how home prices rise or fall, and there are other theories that emphasize the socioeconomic status that largely plays into this house's price tag. Is behind. Home prices are the number of selected stockholders, so apparently home prices are worth a pint Rbabhasa. To predict home prices, a person will usually try to identify similar national characteristics around them, and based on the information collected, that person will try to estimate home prices. All of this indicates that home mortgage forecasts are a growing research area that requires machine learning knowledge. It inspired me to work in this domain.

The Ames Housing Price Dataset at Cagle [2] was recently published [2] "is an updated and expanded version of the recurring Boston Housing Dataset". It covers all residential sale prices recorded from January 25 through July 25 in US IS.

Each home description has features such as home area, garage capacity, overall home and kitchen estimates, city specific, building data, sales model, building year ('built year') and more.

The dataset contains 1460 observations in the train set and 1459 observations in the test set. There are 46 categorical variables in the category, including 23 23 nominal and general general manager, and n 33 numeric variables in the dataset. The training set also has a sales price in response when the test set does not.

We used parameters such as 'square foot area', 'no'. We consider the approved dataset with modifications to give precise results for all bedroom conditions ", 'no toilet count', 'floor type', 'lift presence', 'parking availability', 'proper condition' '. We have used the following different algorithms in different combinations and the weight accuracy for each algorithm is given on a percentage basis. After evaluating the different test curves, we come to the conclusion that a series algorithm instead of a separate algorithm gives the best results.

II. RELATED WORK

There are two major challenges that researchers face. The biggest challenge is identifying the maximum number of properties that can help you accurately estimate the direction of home prices. General Chat Chat Lounge The model that the priest worked with showed how home prices can be relatively trendy where home assets grow faster than incomes in the long run, then fall, and experience longevity.

Reduction [2] in his doctoral dissertation noted that he was the most influential factor in determining the cost of housing with his research work on indoor space. He also cites average census income, which remains a very effective factor in determining home prices.

Paradox [1] uses features such as the use of linear numeric technology to predict the floor size, size category, number of bathrooms, number of bedrooms, standard age and garage size and home prices.

The second major challenge that researchers face is finding machine learning techniques that will be highly effective in predicting home prices accurately. Ang and Dizonart []] developed a mobile-based application using the Gauss process for stress. According to Hu et al [5] use the Maximum Data Coefficient (MIC) to construct a precise mathematical model for predicting home prices. Lamsambucha also built a model using properties such as home size, age of home, number of bedrooms, number of bedrooms, number of bathrooms, number of garages, amenities around the house and geographical location. His work on home pricing in New Zealand compares precision performance between hedonic and artificial neural network models and observes that neural networks perform better than hedonic models in accurately predicting home prices. Bjورrk and Mلller [3] use time series models to predict home prices.

Current work is unique to all of these tasks, because instead of seeing the problem as a regressive view that tries to predict home prices, the work creates a problem as a classification problem, namely that house prices will rise. Or fall.

III. WORK PROGRESS

Our main focus here is to develop a model that predicts the cost of ownership for a buyer based on his or her benefit. Our model analyzes the customer-selected parameter set to find an ideal value that suits their needs and interests. It uses a classical technique called line risk, forest regression and enhanced regression to estimate and analyze the gains. In addition, neural networks are further used to improve the accuracy of algorithms which subsequently develop with induced stress. This helps to strengthen the relationship between the dependent variable and the label attribute and other dependent independent variables, respectively, as a regular attribute.

Our dataset contains various essential parameters and data mining that are at the core of our system. We initially cleaned up all of our datasets and adjusted the style values as well. Cooking [9]

However, as we followed a specific path to increase accuracy, our survey concluded that real estate value is also closely related to local amenities such as train stations, schools, hospitals, temples, parks and more. Now we offer our own unique way. It can address this need. We use the Google Maps API, and based on local searches, we limit the radius to km. Now if we find such a public place within the circle, we are raising the value of the property appropriately. We did this with a handy example and it gave us great results in our prediction accuracy [2].

The whole work process can be divided into four categories: These are feature definitions, feature selection, implementation of machine learning techniques and performance measurement methods.

A. Feature Definition

The present work uses datasets from web source Kaggle.com and the competition hosted by its web application.

B. Feature Selection

It uses feature selection techniques such as impact factor, data values, policy component analysis and data exchange techniques such as external and missing value treatments as well as selection and subsequent selection of Cox-Cox exchange techniques. For the exchange process. These techniques are used in the following ways:

i) Information Value Computation

The value of predictor variables is a non-linear measure that calculates the level of information available on predictor variables among predictor variables. In our work, the target variable has 2 values, namely 0 and 1, while 0 indicates a price drop and 1 a price increase. The data values are calculated for all data and then the properties of the big data values are chosen as the most important feature for further development. A canopy mounting tool is used in this process.

ii) Data Transformation

Information exchange techniques apply to all features selected from the data quality accounting process. The data conversion process includes a follow-up shot of the Box-Cox exchange process and value-removal techniques. In the process of exchanging box cooks, the default value is converted to square, opposite, and significance values.

***iii)* Analysis of key materials**

The original form is assigned to the component process analysis or PCA form. Key components of the policy are analyzed via the Python "PCA" package. This is done to ensure that there are no mates in the feature set.

***iv)* Variation Influence Factory**

The variable input inflation factor (VIF) is the measure of the correlation of this variable with other variables. If the correlation between variability is high and therefore the inflation factor (VIF) is high as a rule of thumb, we try to maintain a set of variations that allow the inflation factor (VIF) of all variations to be 1.5- Is less than 2.0. To implement the change effect element, we use the "Status Models" package in Python. Now we look at the table below, which is a very important feature with corresponding data values.

Table I. Most Important Features With Iv Values

Features	Information Value
Overall Quality of the building	100.212714
Total Basement Square Foot	79.045944
Year Built	75.879952
Garage Area (Square Foot)	67.871142
Garage Cars	63.907900
1st Floor (Square Foot)	60.820567
Year Remodeled	58.179838
Garage Year Built	52.611296
Lot Area (Square Foot)	43.699276
Total Rooms Above Grade	38.715080
Open Porch (Square Foot)	37.455398
Lot Frontage	36.74
Fireplaces	33.579414
Basement (Square Foot)	32.67648
Mas Vnr Area	30.083962

While constructing the above table we have considered that the minimum cut off for the Information Value is 30.

***C.* Machine Learning Technique**

Once we have selected the features we use three technology support vector machines (SVM), random forest and artificial neural network (ANN).

***i)* Support Vector Machines**

Support vector machines are better off with maximum linear performance (classification). Margin is defined as the width at which the boundary can be extended before multiplying the data point

***ii)* Random Forest**

Random forests are hierarchical classes created from the set of decision trees, the product of the class is the product of the decision tree. Random Forests combines the Beiman concept with the "Beijing" concept of the concept of random selection of features. The algorithm was created by Leo Bremen and Adele Cutler to encourage Random Forest.

***iii)* Artificial Neural Network**

Artificial neural networks use neurons or predators as primary units. This audience uses a vector of true Hollywood stuff. These artifacts are always a literal combination of themselves. If the result is greater than the threshold value and the output is 0 then the function is output.

***D.* Performance Measurement**

This work uses the following metrics to measure performance. We consider true values as category values where class 1 predicts when the target value is 1, true negatives are values where classifier 0 is the target value 0, false negative values where class 1 is false. The target value is 0 and 0 is the false negative value where the class predicts 0, but the target values are 1.

Algorithms used

i) Linear Regression

Linear pressure is the simplest method of presentation. It uses two things as variables, which are predictive variables and variables that are the most important first, whether hunting cues and su. These stress estimates are used to describe the relationship between the dependent variable and one or more independent variables. The regression equation is determined by the equation formula with one dependent and one distinct variable [8]. **b = y + x * a** where, **b = dependent variable score**,

y = constant, x = regression coefficient, **and a = score on the independent variable**.

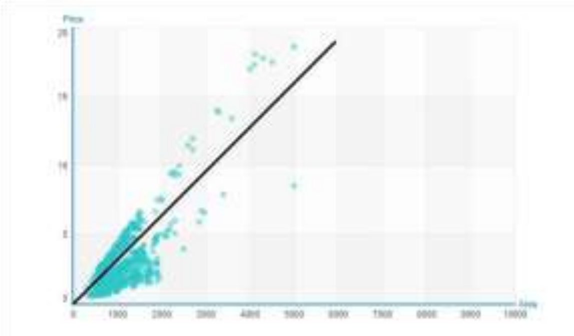

Figure 1: Linear regression scatter plot

ii) Forest Regression

Forest stress uses a technique known as tree cutting. The key idea here is to decorate a few trees. Then we reduce the average variance of the trees. Using this method, a large number of decision trees are made [3].

Rarely does a forest training algorithm use the technique of training or training a tree trainer.

Given a training set X = x1, ...,xn with responses Y = y1, ...,yn, bagging repeatedly (B times) selects arandom sample with replacement of the training set and fits trees to these samples:

For b = 1, ...,B:

1. Sample, with replacement, n training examples from X, Y; call these Xb, Yb.
2. Train a classification or regression tree fb on Xb, Yb.

After training, predictions for unseen samples a' can be made by averaging the predictions from all the

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(x')$$

individual regression trees on a':

In addition, the prediction uncertainty can be estimated as the standard deviation of the predictor from all individual stressors.

trees on *a'*:

$$\sigma = \sqrt{\frac{\sum_{b=1}^{B}(f_b(x') - \hat{f})^2}{B-1}}.$$

IV. EXPERIMENTAL RESULTS

A system has been created to provide accurate predictions of housing prices. The system uses linear regression, forest regression, boosted regression best. The efficiency of the algorithm is further enhanced with the use of neural networks. The system will satisfy customers by providing the correct output and preventing the risk of investing in the wrong home. Additional features can also be added to the system for the benefit of the customer without disrupting its core functionality. A major update of the future could be adding big cities to the database, which will allow our users to explore more buildings, get more accuracy and thus make a right decision.

The accuracy of the system can be improved. As the size of the system and the computation power increase, a few more quotes can be included in the system.

Various classifications have been assigned for this purpose. All of these classifiers are supervised types. The appropriateness of the use of supervised classification is that prior knowledge is needed to predict the value of a home, and such classifier works with prior knowledge. Various classifications have been edited based on accuracy, precision, sensitivity and specificity. Comparative analyzes are tabulated in Table II.

Table II. Comparision Between Different Types Of Classifiers

Classifier	Accuracy (in %)	Precision (in %)	Sensitivity (in %)	Specificity (in %)
Artificial neural network	80	74	81	86
Support Vector Machine	82	75	84	81
Random Forest	86	79	82	83

V. CONCLUSION

From our experimental results we can conclude that Random Forest offers more precision but at the same time this particular classification is at high risk so the performance vector machine category function may be more reliable and consistent than other rest of the two classifiers.

REFERENCES

[1]. Pardoe, I.: Modeling home prices using realtor data. 16(2), 1-9 (2008).

[2]. A. Adair, J. Berry, W. McGreal, Hedonic modeling, housing submarkets and residential valuation, Journal of Property Research, 13 (1996) 67-83.

[3]. O. Bin, A prediction comparison of housing sales prices by parametric versus semi-

parametric regressions, Journal of Housing Economics, 13 (2004) 68-84.
[4]. Lowrance, E.R.: Predicting the market value of single-family residential real estate. 1st edn. PhD diss., New York University, (2015).
[5]. Bork, M., Moller, V.S.: House price forecast ability: a factor analysis. Real Estate Economics. Heidelberg (2016).
[6]. Ng, A., Deisenroth, M.: Machine learning for a London housing price prediction mobile application. Imperial College London, (2015).
[7]. The elements of statistical learning, Trevor Hastie - Random Forest Generation
[8]. C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006
[9]. S. Yin, S. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring," IEEE
[10]. Hu, G., Wang, J., & Feng, W.: Multivariate regression modeling for home value estimates with evaluation using maximum information coefficient. Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. 1(2), 69-81 (2013).
[11]. Limsombunchao, V.: House price prediction: hedonic price model vs. artificial neural network. Lincoln University, NZ, (2004).
[12]. Kahn, J.: What drives housing prices? Federal Reserve Bank of New York Staff Reports, New York, USA, (2008).

Author Profile

ShubhamBadyal
M.Tech Research Scholar,
Center for Cloud Infrastructure And Security
Suresh GyanVihar University, Jaipur
(doluv09@gmail.com)

Ms. Reminder Kaur,
Assistant Professor,
Department of Computer Science and Engineering,
Suresh GyanVihar University, Jaipur
(raminder.kaur@mygyanvihar.com)

Shubham Badyal "Estimating house Rent using machine learning Over Cloud Computing" International Journal of Engineering Research and Applications (IJERA), vol. 9, no. 11, 2019, pp 08-12