Journal of Engineering Sciences
Vol 11, Issue 6,June/2020
ISSN NO:0377-9254

HOUSE PRICE PREDICTION USING ADVANCED REGRESSION TECHNIQUES

Mr. P. Ravindra, Assistant Professor, Department of Computer Science and Engineering, Matrusri Engineering College, Saidabad, Hyderabad, Telangana
Katta Meghana, Gattu Bhavitha, Donkada Neelima, B.E Scholars, Department of Computer Science and Engineering, Matrusri Engineering College, Saidabad, Hyderabad, Telangana

ABSTRACT - House prices increase every year, so there is a need for a system to predict house prices. Usually, House price index represents the summarized price changes of residential housing. While for a single-family house price prediction, it needs more accurate method based on location, house type, size, year of build, local amenities, and some other factors which could affect house demand and supply. With limited dataset and data features, a practical and composite data pre-processing, creative feature engineering method is examined.

People are attentive when they are trying to buy a new house with their budgets and market strategies. The motive is to help the seller to estimate the selling cost of a house perfectly and to help buyers to predict the exact time slap to accumulate a house. Some of the related factors that impact the cost were also taken into consideration such as physical conditions, concept and location etc.

House price prediction can also help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house. It needs more accurate method based on location, house type, size, build year, local amenities, and some other factors which could affect house demand and supply. This model uses regression techniques to predict the price of the house.

In this project we train a model based on the given data to predict the house price as accurate as possible. So that it predicts the house price for customers with respect to their budgets and priorities by analyzing previous market trends and price ranges.

1. INTRODUCTION

Investment is a business activity on which 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. Housing price trends are not only the concern of buyers and sellers, but it also indicates the current economic situation. There are many factors which has impact on house prices, such as number of rooms, location, size etc. Even the nearby location, a location with a great accessibility to highways, schools, shopping malls and local employment opportunities contributes to the rise in house price. Manual house price prediction becomes difficult, hence there are many systems developed for house price prediction.

Nowadays, e-education and e-learning is highly influenced. Everything is shifting from manual to automated systems. The objective of this project is to predict the house prices in order to minimize the problems faced by the customer. The insights gathered to make predictions could be used by individuals to complement their decision making process when purchasing a house. This helps to maximise the value users can gain while keeping to a budget.

There are several approaches that can be used to determine the price of the house, one of them is the prediction analysis. In determining the price of home, the developer must calculate carefully because property prices always increase continuously and almost never fall in the long term or short. It efficiently analyses previous market trends and price ranges, to predict future prices. It predicts the house price efficiently for customers with respect to their budgets and priorities.

Our approach to prediction is to use regression techniques. There are various regression techniques available like Linear Regression, Polynomial Regression, Stepwise Regression, Ridge Regression, Logistic Regression, Lasso Regression and Elastic net Regression.

2. LITERATURE SURVEY

Surveyon "Valuation of House Prices Using Predictive Techniques"(NeelamShinde, Kiran Gawande)

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 3000 properties. These show the R-squared value of 0.98, 0.96, 0.81 and 0.99 respectively. Further, we have compared these algorithms based on parameters such as MAE, MSE, RMSE and accuracy.

we can select the parameters which are not correlated to each other and are independent in nature. We found that Decision Tree overfits our dataset and gives the highest accuracy of 84.64%. Lasso gives the least accuracy of 60.32%. Logistic Regression and Support Vector Regression giving an accuracy of 72.81% and 67.81% respectively.

3. OVERVIEW OF THESYSTEM

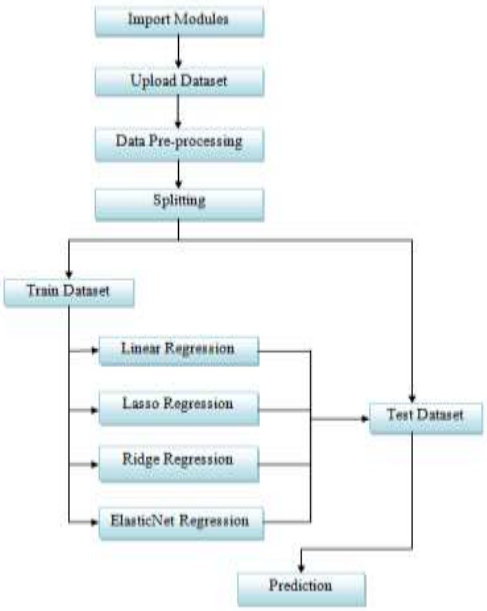

Fig 3.1: System Architecture

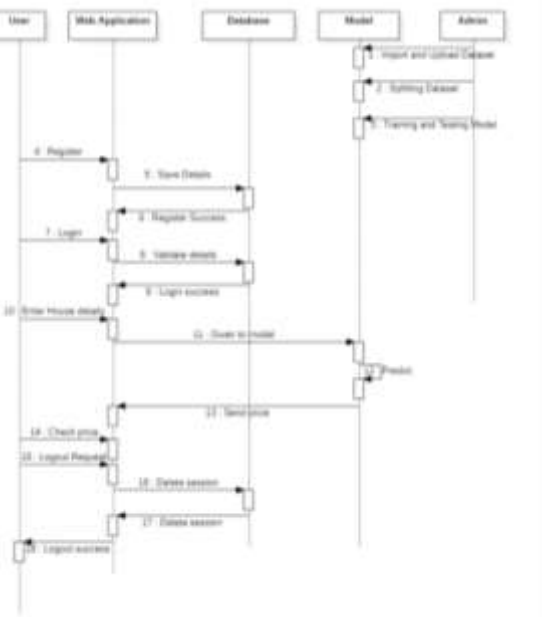

Fig 3.2: Sequence Diagram

3.1 Problem Statement

People looking to buy a new house tend to be more conservative with their budgets and market strategies. House price prediction can help the customer to arrange the right time to purchase a house.

Nowadays, everything is shifting from manual to automated systems. So, we decided to develop a model that predicts the price of the house when user provides location and size as input.

3.2 Existing System

In the existing system property buying or selling is hectic and expensive. As the customer has to roam places and has to pay commission to the Real estate agent. Also, the customer/buyer does not know whether the property is profitable in future or not.

In this system houses are being sold by third party agents who themselves may take some stake of the money resulting in frauds and sometimes may make wrong predictions. So, this method leads to loss of the customer's investments and has high risk. So as to overcome this fault, there is a need for an updated and automated system.

3.2.1 Problems with Existing System

1. HUMAN resource: - The current system has too much manual work from filling a form to filing a document, delivering manifesto. This increases burden on workers but does not yield the effective results it should.
2. THORNY Job: - In current system if any modification is to be made it increases manual work and is error prone.
3. ERROR: - As the system is managed and maintained by workers, errors are some of the possibilities.

3.3 Proposed System

Nowadays, e-education and e-learning is highly influenced. Everything is shifting from manual to automated systems. The objective of this project is to predict the house prices so as to minimize the problems faced by the customer. The present method is that the customer approaches a real estate agent to manage his/her investments and suggest suitable estates for his investments. But this method is risky as the agent might predict wrong estates and thus leading to loss of the customers' investments. The manual method which is currently used in the market is out dated and has high risk. So as to overcome this fault, there is a need for an updated and automated system.

To overcome the problems, we create a model that predicts price of the house efficiently by using regression techniques. Here Regression algorithms such as Linear, Lasso, Ridge, and Elastic Regressions are employed to build the predictive model. Also, the new system will be cost and time efficient. This will have simple operations.

3.4 Process logic

Our approach to prediction is to use regression techniques. In this project we are using Linear Regression, Lasso Regression, Ridge Regression, ElasticNet Regression Techniques for price prediction analysis.

Algorithms used

Linear Regression: It establishes a relationship between dependent variable (Y) and one or more independent variables (X) using a best fit straight line (also known as regression line). It shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable. Mathematically, we can represent a linear regression as:

$$y= a_0+a_1x+ \varepsilon$$

Here,

Y= Dependent Variable (Target Variable)
X= Independent Variable (predictor Variable)
a0= intercept of the line (Gives an additional degree of freedom)
a1 = Linear regression coefficient (scale factor to each input value).
ε = random error

Lasso Regression: LASSO stands for *Least Absolute Shrinkage and Selection Operator*. In statistics and machine learning, lasso is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces. It Shrinkage is where data values are shrunk towards a central point, like the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters).

Ridge Regression: Ridge Regression is a technique for analyzing multiple regression data that suffer from multicollinearity. When multicollinearity occurs, least squares estimates are unbiased, but their variances are large so they may be far from the true value. Ridge regression belongs to a class of regression tools that use L2 regularization.

ElasticNet Regression: Elasticnet is a regularized regression method that linearly combines the L_1 and L_2 penalties of the lasso and ridge methods. The technique combines both the lasso and ridge regression methods by learning from their shortcomings to improve on the regularization of statistical Model.

4.2 Evaluation Metrics

R Squared Value: It determines how much of the total variation in Y (dependent variable) is explained by the variation in X (independent variable). Mathematically, it can be written as:

$$\mathrm{R-Square} = 1 - \frac{\sum(Y_{actual} - Y_{predicted})\hat{}2}{\sum(Y_{actual} - Y_{mean})\hat{}2}$$

Adjusted R Squared Value: The Adjusted R-Square is the modified form of R-Square that has been adjusted for the number of predictors in the model. It incorporates model's degree of freedom. The adjusted R-Square only increases if the new term improves the model accuracy.

$$R^2\mathrm{adjusted} = 1 - \frac{(1-R^2)(N-1)}{\mathrm{N-p-1}}$$

Where

R^2 = Sample R square

P= No of predictors

N= Total sample size

3.5.FUNCTIONAL REQUIREMENTS

- The application must have a user-friendly interface that stores user details.
- The application must work efficiently to predict the result.
- The user must register and login which enables the user to get price of the house.
- The user should enter some details of house in order to know the price.
- User should get accurate search results.

5. RESULTS

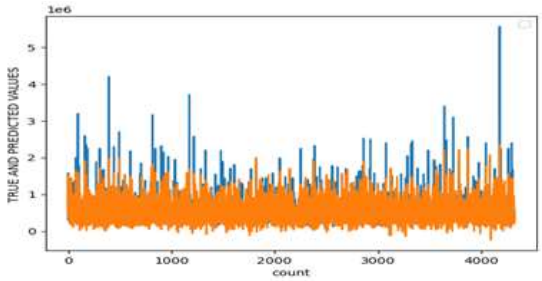

Fig 5.1: Linear Regression Graph

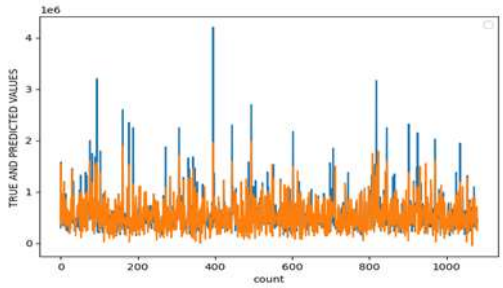

Fig 5.2: Lasso Regression Graph

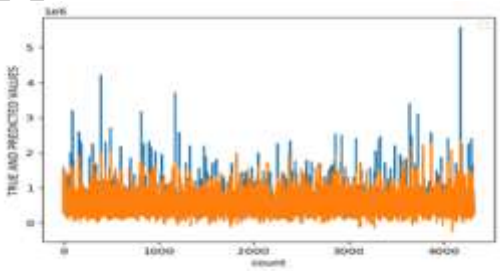

Fig 5.3: Ridge Regression Graph

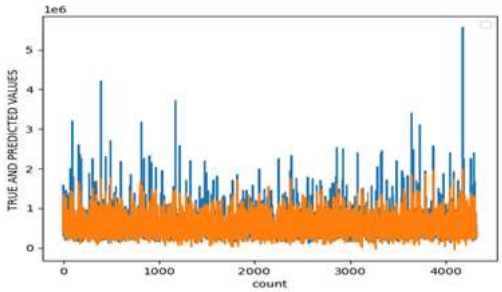

Fig 5.4: ElasticNet Regression Graph

Fig 5.5: User Interface Page displaying predicted House Price.

Table 1: Analysis of Regression Techniques

Algorithm	R- Squared Value	Adjusted R-squared value
Linear regression	0.6952490098854764	0.6939744936628786
Lasso regression	0.6949536925136919	0.6936779412277361
Ridge regression	0.6949539226552225	0.6936781723317546
Elastic net regression	0.6674101549595732	0.6660192122990882

6. CONCLUSION

In this project we used regression algorithms to predict house prices of real dataset of Seattle, USA. This dataset is used to train and test four different regression models which include Linear Regression, Lasso Regression, Ridge Regression and ElasticNet Regression model.

In order to evaluate performance of each model we have used evaluation metrics like R-squared value and adjusted R-squared value. The value of R2 and Adjusted R2 are compared to select accurate model. By making a clear analysis we have found that linear regression has high R squared and Adjusted R squared values and ElasticNet has lower values. Hence, we considered Linear Regression model to predict the house prices. So, we can conclude that system makes optimal use of linear regression to predict house prices. The application developed provides a user interface which allows the user to Register, Login, and Logout and enter features of the house as input based on his/her requirements. Based on the inputs provided by the user accurate house price is predicted by the model and displayed on the Interface. Thus, the system developed helps to fulfil user providing accurate price and reducing risk of huge investment.

7. FUTURE ENHANCEMENTS

In this paper, we proposed a house price prediction model using regression techniques which minimizes the problems with existing system and helps the user to get the right price of his/her house.

Future Enhancements may include:

- Working on large dataset that would yield a better and real picture about the model.
- We have undertaken only few Machine Learning algorithms that are actually regression algorithms but we need to train many other classifiers and understand their predicting behaviour.
- Developing a mobile application with good features.

8. REFERENCES

[1] Vishal Raman, May 2014. Identifying Customer Interest in Real Estate Using Data Mining.

[2] Park, Byeonghwa, and Jae Kwon Bae. "Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data." Expert Systems with Applications 42.6 (2015): 2928-2934.

[3] Hromada, Eduard. "Mapping of real estate prices using data mining techniques."Procedia Engineering 123 (2015): 233-240.

[4] Data pre-processing https://www.javatpoint.com/multiple-linear-regression-in- machine-learning.

[5] Django settings https://www.digitalocean.com/community/tutorials/how-to-create-a-django- app-and-connect-it-to-a-database

[6] https://www.analyticsvidhya.com/blog/2017/06/a-comprehensive-guide-for-linear-ridge- and-lasso-regression/
[7] Regression Algorithms https://scikit-learn.org/stable/
[8] Django https://docs.djangoproject.com/en/3.0/
[9] HTML and CSS https://www.w3schools.com/.