International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Performance of Ridge Regression in Comprehensive Analysis of Residential Property Pricing

Prof. Pankaj Kumar
ISE department
Acharya Institute of Technology Bangalore
pankaj2472@acharya.ac.in

Deeksha H
ISE department
Acharya Institute of Technology Bangalore
deekshah.19.beis@acharya.ac.in

Manasa H
ISE department
Acharya Institute of Technology Bangalore
manasah.19.beis@acharya.ac.in

Bhuvika A M
ISE department
Acharya Institute of Technology Bangalore
bhuvikaa.19.beis@acharya.ac.in

Shilpa T H
ISE department
Acharya Institute of Technology Bangalore
shilpat.19.beis@acharya.ac.in

***Abstract*— The value of the world's economy is rising daily. The country's economic strength immediately raises the cost of real estate there. The adoption of a residential property pricing model is required to forecast residential property values, which enables customers to purchase homes at lower costs and in accordance with their needs. The bedrooms, location, type, and availability of the area, as well as the overall square feet, are all factors used in this model. This model calculates the price of a property using a variety of these attributes and their values, enabling the consumer to know the most accurate pricing. The most precise method for predicting the price of a residential property is to use ridge regression and linear regression with a lasso (least absolute shrinkage and Selection operator). The most reliable way to anticipate the value of a residential property is to know its pricing. The model is specifically made to calculate a dwelling price using pre-loaded data.**

Keywords- machine learning techniques include ridge regression, least absolute shrinkage and selection operator, and linear regression.

I. Introduction

In today's bustling cities, finding a home that meets our needs and wants is a top priority. For buyers and sellers alike, making an accurate prediction of a house in an unfamiliar location with specific specifications is always fascinating. Since the value of real estate is changing quickly, a machine learning application is being developed to offer a solution. The main component of machine learning is data. Data is used by machine learning algorithms to train their models and produce correct results. Potential homebuyers, appraisers, builders, tax assessors, and other real estate industry clients, like pledge and insurance, value an accurate projection of a home' price. Housing location, quality, site, and environment are the primary factors that influence house pricing. Each standard value has typically improved the predictive power of the models. Machine Learning is a field of Artificial Intelligence which enables PC frameworks to learn and improve in execution with the assistance of information. It is used to study the construction of algorithms that make predictions on data.

The evolve of Internet and better hardware and software has opened many doors such as we can now use Machine Learning to do prediction here houses so that it can address all the concern of buyers and the sellers. House is a basic necessity for a person and their prices vary from location to location based on the facilities available like parking space, locality, etc. Property investment has also increased significantly. Buying of a house is one of the greatest and significant choice of a family as it expands the entirety of their investment funds and now and again covers them under loans. It is the difficult task to predict the accurate values of house pricing. Our proposed model would make it possible to predict the exact prices of houses. Housing price trends are not only the concern of buyers and the sellers, it also indicates the current economic situation. There are many factors which has impact on house prices, such as location, BHK, floor etc. Also, a location with a great accessibility to highways, expressways, schools, shopping malls and local employment opportunities contributes to the rise in house price. Manual house prediction becomes difficult, hence there are many systems developed for house price prediction. The aim of this system is to create a website through which the user can give his house requirements as input which is then passed on to the most accurate model among Naïve Bayes, Linear Regression with SVM. Random Forest Regression, for predicting the house price. The website allows user to predict the house prices to a particular place, price-range and other amenities as specified by the user.

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly

programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

II .METHODOLOGY

3.a Existing system

Taxes are not taken into account by the current algorithm; instead, it only takes into account geography and other factors, which has a significant impact on prediction accuracy. Other regression methods yield rather poor accuracy. The existing models have used classification algorithms that predicts based on location and do not include many external features. Existing system does not consider taxes but only considers location and other features which makes a huge difference in accurate prediction. Using other regression techniques gives comparatively low accuracy. The existing models have used classification algorithms that predicts based on location and do not include many external features.

3.b Proposed system

The world is moving away from manual processes and toward automated ones. The goal of our project is to help the customer solve their challenges. In the current case, the customer goes to a real estate agent to ask for recommendations for good showplaces for his assets. However, the foregoing strategy is dangerous because the agent may provide the customer incorrect prices, resulting in a loss of the customer's investment. This manual procedure, which is now in use on the market, is obsolete and dangerous. To overcome the disadvantage, an updated and automated system is required. Data scraping is the first phase in our proposed system. It's a technique for extracting structured data from the web or any application and saving it to a database, spreadsheet, or CSV file. After the data has been extracted, we clean it. It speaks about the adjustments performed to the data prior to its input into the formula. The process of transforming unclean data into clean data includes handling missing data and category data as necessary. The outlier values have been trimmed after our entire dataset has been cleaned up. When we're done cleaning, we'll employ a range of algorithms. Both the lasso and the ridge regression algorithms were used in our system. The least absolute shrinkage and selection operator are used by the lasso. Ridge regression then utilizes associated features. When a user enters the necessary information about a house in a web application, these two regression models predict the price, compare it to one another, and then display the highly accurate result on the output screen.

3.c Algorithm

A. Linear Regression

It is an algorithm of supervised machine learning in which the predicted output is continuous which is having a constant slope. It is used to predict the values in a continuous range instead of classifying the values in the categories. Linear regression is used for performing different tasks such as house price prediction.

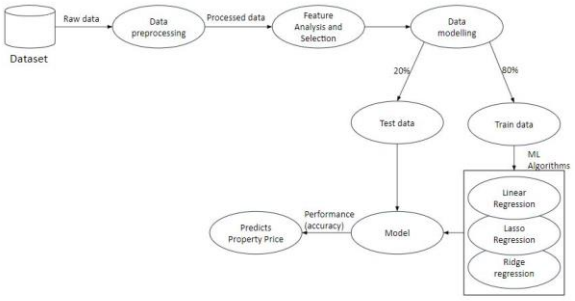

Fig1- SYSTEM ARCHITECTURE OF PROPERTY PRICE PREDICTION

B. Lasso regression

Least Absolute Shrinkage and Selection Operator is referred to as LASSO. One form of linear regression that employs shrinkage is the lasso regression. It is a regression analysis technique that includes both variable selection and regularization, as the name would imply. Lasso
Regression only chooses a portion of the offered covariates to be included in the final model. You normally begin by gathering a dataset that has numerous traits or variables that can potentially affect home prices when using lasso regression to forecast house price. The number of bedrooms, square footage, location, bathrooms, and other characteristics may be among these. The house price you want to predict would be the target variable. LASSO stands for Least Absolute Shrinkage and Selection Operator. The lasso regression is one type of linear regression that makes use of shrinkage. As the name would suggest, it is a regression analysis technique that incorporates both variable selection and regularization. Lasso Only a part of the available covariates are selected via regression to be incorporated into the final model. When using lasso regression to predict house prices, you often start by collecting a dataset that has a variety of characteristics or variables that could potentially effect home prices. These could include things like the number of bedrooms, square space, location, bathrooms, and other things. The target variable would be the house price you want to forecast.

C. Ridge regression

It is a tool for analysis of Multiple Regression on the data that have multicollinearity(mcl). Multicollinearity(mcl) is existence of near–linear relationships among the variables which are independent. When Multicollinearity occurs, least squares estimates are unbiased. Ridge regression reduces the standard errors by adding a degree of bias to the regression estimates. The formula for Ridge regression is ,adding a degree of bias to the regression estimates. Ridge regression is an extension of ordinary least squares (OLS) regression that introduces a regularization term to the loss function. This term penalizes the magnitudes of the regression coefficients, encouraging them to be smaller. By shrinking the coefficients, ridge regression helps mitigate the effects of multicollinearity and reduces the risk of overfitting. Ridge regression introduces a tuning parameter, commonly denoted as λ (lambda), which controls the amount of regularization applied. A larger λ value results in stronger regularization and greater coefficient shrinkage.

information criterion (AIC) or Bayesian information criterion (BIC). The ridge regression equation is an extension of the ordinary linear regression equation, where the least squares loss function is modified to include the regularization term. The ridge regression equation can be represented as:

$$y = \beta0 + \beta1x1 + \beta2x2 + ... + \beta nxn + \varepsilon + \lambda\Sigma(\beta i^2)$$

In this equation, y represents the dependent variable (house price), β0 to βn are the regression coefficients, x1 to xn are the independent variables (features), ε is the error term, and λΣ(βi^2) represents the regularization term. Ridge regression is effective in dealing with highly correlated independent variables, which is common in house price prediction datasets. By shrinking the coefficients, it reduces the impact of multicollinearity on the regression model's stability and interpretability. Ridge regression introduces regularization, which prevents the model from becoming too complex and overfitting the training data. This leads to better generalization performance when predicting house prices on unseen data.
Unlike feature selection techniques that discard some variables, ridge regression retains all the variables in the model while shrinking their coefficients. This can be useful when you want to retain all the available information without eliminating any features. Ridge regression can be implemented using various statistical software packages or programming languages that offer regression functionality. Common implementations include Python with libraries such as scikit-learn, R with packages like glmnet, or MATLAB with the Statistics and Machine Learning Toolbox. When using ridge regression for house price prediction, it is important to evaluate the model's performance. Common evaluation metrics include R-squared (coefficient of determination), adjusted R-squared, root mean squared error (RMSE), or mean absolute error (MAE). These metrics assess the goodness-of-fit and predictive accuracy of the ridge regression model. The choice of the regularization

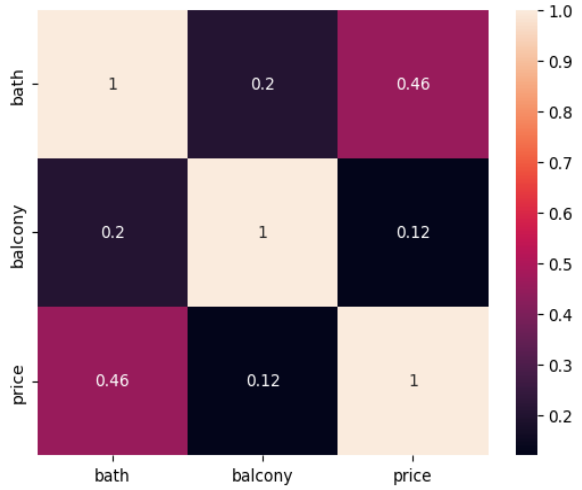

Fig 2 Performance of Ridge Regression

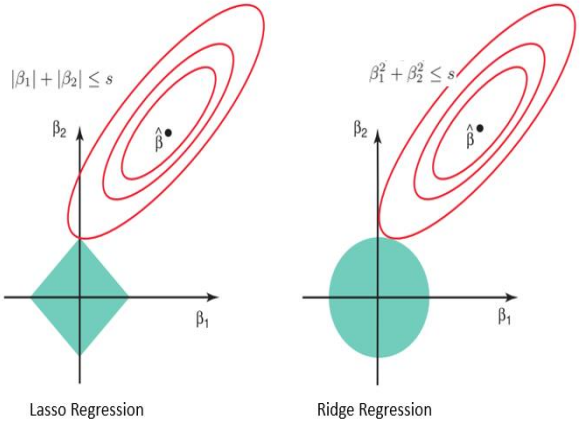

Fig 3 heat map

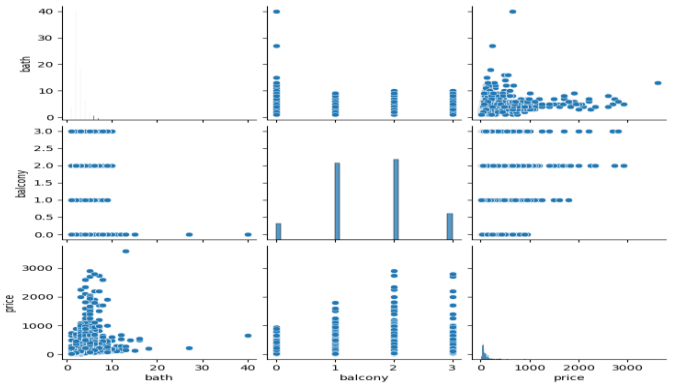

Fig 4 Lasso and Ridge Regression

ameter λ is crucial in ridge regression. It is important to find the optimal value that balances between fitting the training data well and avoiding excessive shrinkage. Cross-validation techniques can be employed to search for the best λ value or to perform a grid search over a range of λ values.

1V CONCLUSION

In this paper, we estimated the cost of homes in Bangalore. Based on test data, a system that aims to deliver a trustworthy prediction of house prices has been created. The precise prediction model would enable investors, homebuyers, and home builders to decide on an affordable house price as well as the realistic price of a home. We employed both lasso and ridge regression in this system. The system uses the parameter or attributes from the web application to compare them to training data. The users will be satisfied by this system's accurate prediction. These models were created using a variety of input attributes, and they have proven to be quite effective at predicting property prices. The user can receive pricing predictions based on tax in addition to the model's weight. Future updates to this system might involve the inclusion of big cities, unfavorable facilities, interior designs, more datasets, and commercial locations.

V . ACKNOWLEDGEMENT

We express our regard to Mr. Pankaj Kumar Assistant Professor, our project guide, and Mr. M K Dhananjaya Assistant Professor, for continuously monitoring the project's progress and establishing exact timelines. Their insightful counsel served as the driving force behind finishing the task progress and establishing exact timelines of the task.

V1. REFERENCES

[1] [2020] Machine Learning based Predicting House prices using Regression Techniques , Manasa j, Radha Gupta , Narahari N S.

[2] D2019] Using Machine Learning Algorithms for Housing Price Prediction : The Case of Fairfax county, Virginia Housing Data, Byeonghwa, and Jae Kwon Bae .

[3] [2019] Prediction of House Pricing using Machine Learning with Python, Namit Jain, Parikshay Purushottam Sharma, Vikas Deep.

[4] [2020] House Price Prediction Modelling using Machine Learning, Dr. M. Thamarai,Dr. S P. Malar vizhi.

[5] Khan Sohil Liyaqatullah1, Usman Malik2, Qureshi Mohammed Basheeruddin3, "International Journal of Advanced Research in Computer and Communication Engineering“, Vol. 10, Issue 4, April 2021,DOI 10.17148/IJARCCE.2021.10468.

[6] Vignesh M ,Vijay V, Krishna S, Satya moorthy, “International Research Journal of Modernization in Engineering Technology and Science,Volume:04/Issue:04/April-2022. eISSN: 2582-5208.

[7] Aldrin Fernandes, Abdullah, Qureshi , Amit Narote, Amir shaik ,“International Journal Of Advance Research ,Ideas And Innovation In Technology ”,Vol 7, Issue 6-V716-1155.

[8] Amena Begum, Nishad Jahan Kheya, Md. Zahidur Rahman, “International Journal of Innovative Technology and Exploring Engineering (IJITEE)”, ISSN: 2278-3075 (Online), Volume-11 Issue-3, January 2022

[9] Bindu Sivasankar1 , Arun P. Ashok2* , Gouri Madhu3 , Fousiya S. 4 JCSE International Journal of Computer Sciences and Engineering. Vol.8, Issue.7, July 2020 E-ISSN: 2347-2693

[10] International journal of creative research, Aniket Singh1 , Adarsh Kumar Singh2 , Aditya Raj3 , Harshit Jain4 ,Mrs. Asha M S 5 , House Price Prediction Analysis using Machine Learning,ISSN 2320-2882.

[11] Anand G Rawool, Dattatary V,Rogye, Sainath G,Rane , DR Vinayak A, Bharathi, “House price prediction using machine learning”. Iconic research and engineering” Vol 4, No11,pp 29-33,2021.

[12] Park, Byeonghwa, and joe Kwon Bae, “Using Machine Learning Algorithms for Housing price prediction: The case of FairFax country ,Virginia Housing Data”,pp.2928-2934, 2019.