IJMTARC – VOLUME – IX – ISSUE – 37, APR - JUN, 2022

ISSN: 2320-1363

A Novel Approach on Predicting House Prices using Regression TechniquesMachine Learning Methods

[1]G Damodar, [2]Dr. B. Madhav Rao

[1]M.Tech Student, Dept. of CSE, Sir C R Reddy College of Engineering College, Eluru.

[2]Associate Professor, Dept. of CSE, Sir C R Reddy College of Engineering College, Eluru.

Abstract: Predictive Fashions for determining the sale rate of houses in towns like bengaluru continues to be last as more challenging and complex mission. The sale price of houses in cities like bengaluru relies upon on some of interdependent elements. Key factors that could affect the rate consist of area of the belongings, vicinity of the property and its facilities. In this studies challenge, an analytical look at has been completed through considering the data set that stays open to the public by using illustrating the available housing houses in machine hackathon platform. The statistics set has 9 capabilities. In this examine, an attempt has been made to assemble a predictive model for comparing the fee based at the elements that affect the fee. Modeling explorations observe some regression strategies together with more than one linear regression (least squares), lasso and ridge regression models, guide vector regression, and boosting algorithms along with intense gradient boost regression (xgimprove). Such fashions are used to construct a predictive version, and to pick the high-quality acting model by performing a comparative analysis on the predictive errors acquired among these models. Right here, the try is to construct a predictive version for comparing the rate primarily based on elements that influences the rate.

Key Words:house price, lasso regression, ridge regression, regression methods.

Introduction: The Most regularly used model for predictive evaluation is regression. As we recognize, the proposed version for accurately predicting destiny consequences has packages in economics, commercial enterprise, banking area,

IJMTARC – VOLUME – IX – ISSUE – 37, APR - JUN, 2022 ISSN: 2320-1363

healthcare industry, e-commerce, amusement, sports etc. One such technique used to forecast residence fees are primarily based on multiple factors [7]. In metropolitancities like bengaluru, the possible home purchaser considers several elements inclusive of location, length of the land, proximity to parks, colleges, hospitals, electricity generation facilities, and most importantly the residence charge. A couple of linear regression is one of the statistical strategies for assessing the relationship between the (dependent) goal variable and several impartialvariables. Regression techniques are broadly used to construct a version based on several elements to are expecting fee. In this examine, we've got made a try to build residence charge prediction regression version for information set that remains on hand to the general public in system hackathon platform. We've taken into consideration five prediction fashions, they are everyday least squares model, lasso and ridge regression fashions, svr model, and xgboost regression model. A comparative examine became executed with evaluation metrics as properly. Once we get a great match, we are able to use the version to forecast financial value of that precise housing property in bengaluru.

Comparative Study: Pow, nissan, emiljanulewicz, and l. Liu [11] used four regression techniques particularly linear regression, assist vector machine, k-nearest acquaintances (knn) and random woodland regression and an ensemble approach through combining knn and random forest approach for predicting the assets's rate cost. The ensemble technique predicted the charges with least blunders of 0. 0985 and making use of pca didn't enhance the prediction errors. Several studies have additionally centered on the collection of features and extraction methods. Wu, jiao yang [12] has as compared numerous function selection and function extraction algorithms combined with help vector regression. Some researchers have evolved neural community models to predict residence charges. Limsombunchai, in comparison hedonic pricing structure with artificial neural community version to are expecting the house fees [13]. The r-squared value received for neural network version was more when in comparison to hedonic version and the rmse value of neural

IJMTARC – VOLUME – IX – ISSUE – 37, APR - JUN, 2022 ISSN: 2320-1363

network version turned into surprisingly lower. Subsequently they concluded that artificial neural network plays better when compared with hedonic model. Cebula applies the hedonic rate version to expect housing expenses inside the metropolis of savannah, georgia. The log rate of houses has been shown to be undoubtedly and appreciably related to the range of bathrooms, bedrooms, fireplaces, garage spaces, testimonies and the residence's general square toes [14]. Jirong, mingcang and liuguangyan follow guide vector machine (svm) regression to expect china's housing charges from 1993 to 2002. They have carried out the genetic set of rules to song the hyper-parameters within the svm regression model. The error rankings acquired for the svm regression version turned into less than four% [15]. Tay and ho in comparison the pricing prediction among regression evaluation and synthetic neural community in predicting apartment's costs. It was concluded that that the neural community version plays higher than regression evaluation version with a median absolute error of 3. 9%.

Existing System: In current gadget 4 regression techniques specifically linear regression, guide vector device, k-nearest buddies (knn) and random woodland regression and an ensemble approach through combining knn and random forest method for predicting the asset's rate price. The ensemble approach predicted the costs with least mistakes of 0. 0985 and applying pca didn't improve the prediction error.

Disadvantages

- It doesn't focus on Bangalore house price dataset.
- We are not getting much accurate model for predicting house price**.**

Proposed system: We Are going to assemble a predictive version for evaluating the price based totally on the elements that affect the charge. Modeling explorations apply some regression techniques together with multiple linear regression (least squares), lasso and ridge regression models, assist vector regression, and boosting algorithms such as excessive gradient enhance regression (xg increase). Such fashions are used to construct a predictive version, and to pick the pleasant performing model via acting a comparative analysis at

IJMTARC – VOLUME – IX – ISSUE – 37, APR - JUN, 2022 ISSN: 2320-1363

the predictive errors acquired among those models. Here, the try is to assemble a predictive model for comparing the price based on factors that influences the fee.

Advantages:

- Proposed system is totally focused on predicting house in Bangalore city.
- Multiple machine learning algorithms is used for predicting the house prices in different locations.

Modules:

Pandas: pandas are an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

Numpy: NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python.

MatPlotLib: matplotlib.Pyplot is a plotting library used for 2D graphics in python programming language. It can be used in python scripts, shell, web application servers and other graphical user interface toolkits

Scikit-learn: Scikit-learn is a free machine learning library for Python. It features various algorithms like support vector machine, random forests, and k-neighbors, and it also supports Python numerical and scientific libraries like NumPy and SciPy.

III. Data Understanding and Pre-Processing:

A. Data Description: The Two records sets-train set and test statistics taken into consideration in the undertaking is taken from system hackathon platform. It consists of functions that describe house-belongings in bengaluru. There are 9 functions in both the data units. The features may be defined as follows:

1. Place kind-describes the place
2. Availability-whilst it is possessing or while it is ready.
3. Charge- price of the belongings in lakhs.
4. Size- in bhk or bed room (1-10 or greater)
5. Society- to which it belongs.
6. Total_sqft - length of the property in sqft.
7. Bath-no of lavatories
8. Balcony- no of balcony
9. Vicinity – wherein it's miles positioned in Bengaluru with nine functions to be had, we attempt to build regression models to predict house price. We predicted the rate of test

IJMTARC – VOLUME – IX – ISSUE – 37, APR - JUN, 2022 ISSN: 2320-1363

facts set with the regression models built on teach records set [8].

B. Data understanding and basic EDA: The Motive is to create a model which can estimate housing fees. We divide the set of statistics into features and goal variable. In this segment, we can try and understand evaluation of unique data set, with its unique functions after which we are able to make an exploratory evaluation of the records set and attempt to get useful observations. The teach data set consists of 11200 facts with 9 explanatory variables. In take a look at statistics set, there have been around 1480 data with 9 variables. Whilst building regression models we are regularly required to convert the explicit i. E. Text functions to its numeric illustration. The two most common ways to do that is to apply label encoder or one warm encoder. Label encoding in python may be accomplished with the aid of using sklearn library. Label encoder encodes labels with a cost among zero and n-1.If a label repeats, it attributes the same value as previouslyassigned [6]. One hot encoding refers to splitting the columnthat contains numerical categorical data to many columnsdepending on the number of categories present in that column.Each column contains “0” or “1” corresponding to whichcolumn it has been placed [6].This dataset includes quite a few categorical variables (bothtrain and test data set) for which we will need to create dummyvariables or use label encoding to convert into numerical form.These would be fake/dummy variables because they areplaceholders for actual variable and are created by ourselves.Also, there are a lot of null values present as well, so we willneed to treat them accordingly. The features bath, price,balcony are numerical variables. Features like area_type, total_sqft, location, society, availability, and size appears ascategorical variables.

C. Data pre-processing

The general steps in data pre-processing are:

- Converting categorical features into numericalvariables in order to fit linear regression model.
- Imputing null records with appropriate values.
- Scaling of data
- Split into train –test sets.

Conclusion: An Surest version does no longer always represent a strong model. A

IJMTARC – VOLUME – IX – ISSUE – 37, APR - JUN, 2022 ISSN: 2320-1363

model that frequently use a studying set of rules that is not suitable for the given records shape. Now and again the statistics itself is probably too noisy or it may include too few samples to allow a model to appropriately capture the goal variable which means that the model stays match. When we observe the evaluation metrics acquired for advanced regression models, we will say both behave in asimilar way. We are able to pick out either one for residence rate prediction as compared to simple model. With the help of box plots, we are able to test for outliers. If present, we will take away outliers and test the model's performance for development. We can build fashions via advanced strategies specifically random forests, neural networks, and particle swarm optimization to enhance the accuracy of predictions.

References:

[1] R. Victor, Machine learning project: Predicting Boston house prices with regression in towards data science.

[2] S. Neelam, G.Kiran, Valuation of house prices using predictivetechniques, Internal Journal of Advances in Electronics and ComputerSciences:2018,vol 5,issue-6

[3] S. Abhishek. :Ridge regression vs Lasso, How these two popular MLRegression techniques work. Analytics India magazine,2018.

[4] S.Raheel.Choosing the right encoding method-Label vs One hotencoder. Towards datascience,2018.

[5] Raj, J. S., &Ananthi, J. V. (2019). Recurrent Neural Networks andNonlinear Prediction in Support Vector Machines. Journal of SoftComputing Paradigm (JSCP), 1(01), 33-40.

[6] Predicting house prices in Bengaluru(Machine Hackathon) https:// www.machinehack.com/course/predicting-house-prices-inbengaluru/

[7] Raj, J. S., &Ananthi, J. V. (2019). Recurrent neural networks andnonlinear prediction in support vector machines. Journal of SofComputing Paradigm (JSCP), 1(01), 33-40.

[8] Pow, Nissan, Emil Janulewicz, and L. Liu (2014). Applied MachineLearning Project 4 Prediction of real estate property prices in Montréal.

[9] Wu, Jiao Yang(2017). Housing Price prediction Using Support VectorRegression.

IJMTARC – VOLUME – IX – ISSUE – 37, APR - JUN, 2022

ISSN: 2320-1363

[10] Limsombunchai, Visit. 2004.House price prediction: hedonic pricemodel vs. artificial neural network.New Zealand Agricultural andResource Economics Society Conference.

[11] Rochard J. Cebula (2009).The Hedonic Pricing Model Applied to theHousing Market of the City of Savannah and Its Savannah HistoriLandmark District; The Review of Regional Studies 39.1 (2009), pp. 9–22

[12] GuJirong, Zhu Mingcang, and Jiang Liuguangyan. (2011).Housingprice based on genetic algorithm and support vector machine". In:Expert Systems with Applications 38 pp. 3383–3386.

[13] H.L. Harter, Method of Least Squares and some alternatives-PartII.International Static Review.1972,43(2) ,pp. 125-190.

[14] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol.2. Oxford: Clarendon, 1892, pp.68-73.

[15] Lu. Sifeietal,A hybrid regression technique for house prices prediction.In proceedings of IEEE conference on Industrial Engineering andEngineering Management: 2017.

About Authors:

G Damodar is currently pursuing his M.Tech (CST) in Computer Science and Engineering Department, Sir C R Reddy College of Engineering College, West Godavari, A.P.

Dr. B. Madhav Rao is currently working as an Associate Professor in Computer Science and Engineering Department, Sir C R Reddy College of Engineering.