

HOUSE PRICE PREDICTION AND RECOMMENDATION

Mary Anitha T

Assistant Professor

Department of Computer Application

The Oxford College of Engineering

mary.anitha.charlton@gmail.com

Vivek V

PG Student

Department of Computer Application

The Oxford College of Engineering

vivekreddy056@gmail.com

ABSTRACT

Our Project is a Machine Learning model/program that predicts the correct esteem of your future domestic based on certain highlights of your future domestic. For individuals who are arranging to purchase a house, the house cost estimate is an assessed future cost run so that they can arrange their financing. In addition, house cost figure is valuable for genuine bequest speculators to know the slant of house costs in certain areas. In genuine domain, most genuine bequest companies contract a few specialists to appraise the esteem of the house and recommend the company to contribute in the house, but some of the time it is not 100% precise. Some of the time they are not able to provide an precise gauge of the house cost. So we need to make an machine learning show that can be utilized to anticipate the esteem of this house. This gives 90% adjust expectation, but gives an exact expectation almost the esteem of the house, since we prepare this demonstrate with information from the past year, which predicts the surmised but exact esteem of the house that employments this machine learning demonstrate. In this extend, we utilize machine learning models to foresee house costs in cities utilizing a dataset given by Kaggle.

Keywords: House Price, Machine Learning, Linear Regression

1. INTRODUCTION

House cost expectation and proposal utilizing machine learning includes predicting calculations and models to foresee the cost of a house based on different highlights of the houses and characteristics. This can be valuable for genuine bequest operators, domestic buyers, and dealers to make educated choices around the cost of a property.

The prepare regularly includes collecting information on different highlights of a house, such as its area, estimate square feet , number of roomsand washrooms, age, and condition. This information is at that point used to prepare a machine learning show, which can foresee the cost of a house based on these highlights. The proposed inquire about paper alludes to the expectations on the recent patterns and for the plans of economy. The fundamental drive behind the article is forecast of the genuine bequest costs to build best of the house cost forecast frameworks utilizing the machine learning calculations with greatest exactness. Beneath the space of ML and Information Science the planning of the genuine estate cost forecast along with the full-fledged site is done. Concurring to the census of 2011 as it were 80 percent of people claim their houses. And as it were individuals based in provincial areas claim most extreme houses but individuals in urban segment as it were about 69 % possess a house. This is due to the raising costs of the properties and unclear house costs. The fundamental point to design and create this demonstrate is to deliver cost forecast system along with a user-friendly front conclusion that will facilitate the clients to select the craved goal and get

an thought almost the cost rates. The Investigation that has been made in the paper is primarily utilizing the dataset from the trusted site that gives adequate of test focuses for way better analysis. One must be mindful of the correct cost of house before concluding the bargain. As the cost of house depends on many variables like Zone, area, populace, measure and number of rooms & washrooms given, stopping space, lift, style of development, overhang space, condition of building, price per square foot etc. The proposed demonstrate points to make an precise result by taking into thought all diverse factors. For House cost forecast one can utilize various prediction models (Machine Learning Models) like bolster vector relapse, Bolster vector machine (SVM), Calculated regression, k-means, manufactured neural arrange etc. House pricing show is advantageous for the buyers, property speculators, and house builders. This demonstrate will be enlightening and knowledgeable for the substances related to the genuine bequest and all the partners to assess the current showcase patterns and budget neighborly properties. Thinks about at first concentrated on investigation of the properties which impact costs of the houses based on which show of ML is utilized and still this article brings together both anticipating house cost and attributes together.

2. LITERATURE SURVEY

The Danh Phan, 2018. "Housing Cost Forecast Utilizing Machine Learning Algorithms.": The Case of Melbourne City, Australia.

The Case ponder around of Melbourne City, Australia Maker: The Danh Phan, 2018 Property taken a toll desire is an crucial point in the honest to goodness space field. Composing endeavors to choose up important bits of information from bona fide property data. Machine learning strategies are associated to consider chronicled property trades in Australia and get models that can be profitable for household buyers and merchants. A basic refinement in property costs between the most expensive and sensible locales of Melbourne city is revealed. In expansion, tests show up that a

combination of stepwise and reinforce vector machines supporting brutal squared botch measures can be a competitive approach.

Parasich Andrey Viktorovich; Parasich Viktor Alexandrovich; Kavtannikov Igor Leopoldovich. "Predicting The Offering Cost Of A House Utilizing Machine Learning Regression."

Article depicts the course of action to the machine learning "House Fetched: Advanced Backslide Methodologies" held on the Kaggle arrange. The objective is to predict the bargain taken a toll of a house based on its characteristics, such as the locale of the house, the year of improvement, etc. Our course of action livelihoods classical machine learning calculations and our have procedures, which are portrayed here. In the 4,444th competition, we put 18th out of 2,124 individuals from around the world.

N Nehal Ghosalkar; Sudhir Dhage, 2018. "Predicting Property Esteem utilizing Straight Regression."

Market of Genuine domain exhibit is one of the most most smoking markets in the shape of assessing and is continuously fluctuating. This is one of the key districts where machine learning concepts are associated to make strides and expect costs with tall precision. There are three factors that choose the taken a toll of a house: physical condition, concept, and region. Current systems evaluate the regard of a house without expecting exhibit taken a toll or taken a toll improvement. The objective of this paper is to expect the honest to goodness inheritance costs inby considering the cash related plans and needs of the buyers. The future costs are expected by analyzing the past grandstand plans and regard ranges as well as the up and coming improvements. This think about points to anticipate the genuine bequest costs in banglore city utilizing direct relapse, which will empower clients to contribute their reserves in blessing without going through a broker. The comes about of this ponder have demonstrated that direct relapse gives a least forecast blunder of 0.3713.

Güven Velthorst; Cicek, 2019. “Predicting genuine bequest advertise patterns utilizing data on Twitter data.”

The think about that we endeavor to predict designs in the Dutch honest to goodness inheritance promote utilizing substance mining and machine learning as an application of data science methodologies in support. Our essential objective is to expect short-term upward or plummeting designs of ordinary property costs in the Dutch publicize utilizing substance data collected from Twitter. Twitter is in addition broadly utilized and has illustrated to be a profitable data source. Be that as it may, Twitter, substance mining (tokenization, bag-of-words, n-grams, weighted term repeat) and machine learning (classification calculations) have not be that as it may been combined to anticipate short-term designs in the honest to goodness space promote. In this consider, tweets containing predefined see terms are collected with space data in judgment skills, so that the comparing compositions are collected by month as files. At that point, words and word strings are changed over into numerical values. These values serve as qualities to expect whether the veritable space grandstand will rise or drop. That is,9H. We considered this as a parallel classification issue to relate the substance data of one month to the incline (up or down) of the taking after month.

Our essential revelations show up that there is a relationship between the (weighted) repeat of words and short-term honest to goodness space designs. In other words, by combining diverse machine learning techniques and substance mining, we were able to accurately predict short-term designs.

3. EXISTING SYSTEM

An investigation led by Nils Landberg. Nils examined how prices have changed in the Swedish real estate market as well as how quality factors affect home values there. Landberg looks at the cost per square foot, employment, population, new residences, new companies, unemployment rate, overall crime rate, and number of jobs. According to Nils, rates of unemployment, crime, interest, and contemporary housing all have a detrimental impact

on housing expenses. Landberg emphasizes that analyzing the hotel showcase is more difficult than analyzing the product display since several optional expenses affect the increase in home costs. Empirical studies indicate that rising population numbers and quality-related factors positively impact,home,values.

Limitations Of The existing system

There are numerous shortcomings with the current The current house price prediction methods have a number of shortcomings, some of which are:

- **Lack of data:** A significant drawback of the current house price prediction algorithms is the dearth of high-quality data. These algorithms base their predictions on previous data, which could produce false results if information is erroneous, old, or incomplete.
- **Lack of transparency:** To create projections, a lot of the house price prediction systems in use today rely on intricate algorithms and machine learning models. Nevertheless, it can be hard to determine how the system arrived at a specific forecast because these models can be opaque and hard to understand.
- **Restricted scope:** Certain home price prediction algorithms might only include a small number of variables, such the location and size of the property, neglecting other crucial factors like the state of the local economy, the level of crime, and the caliber of the schools.
- **Unforeseen occurrences:** Because house price prediction models rely on past data and patterns, they might not take into consideration unanticipated events like natural catastrophes or significant economic.
- **Human bias:** Predictive algorithms for house prices may be biased due to the prejudices of its developers or users, which could result in unjust or erroneous forecasts. For instance, the system may reinforce biases in its predictions if the data used to train it is skewed toward particular neighborhoods or populations.

4. PROPOSED SYSTEM

Our main goal in the proposed project is to develop a proposal and dwelling cost prediction framework. Expected house costs and A common task in the field of information science was developing a recommendation framework, which is why machine learning computations are used in part. We will apply slope boost, arbitrary woods, choice tree, and straight relapse computations for house cost forecasting in this suggested framework.

This suggested system employs a variety of machine learning computations to predict house costs and provide accurate housing prescriptions. We can increase the accuracy of housing cost projections by evaluating the performance of various models and sending the best one. In order to predict housing costs and recommend the houses, the suggested framework combines four machine learning computations: choice tree, slope boost, irregular timberland, and straight relapse.

This strategy can advance the accurate precision of the predictions and recommendations by merging the features of several algorithms. Pre-processing of the data and construction procedures are also included in this suggested framework, which can advance the estimations' correctness. Additionally, the framework includes methods for evaluating and determining which shows to submit, which can help choose which shows perform the best.

One machine learning demonstration known as "direct relapse" aims to show a clear correlation between an independent variable (X) and a dependent variable (Y). The true value of the target (Y) is compared to the expected value of the target (Y) for every observation that the display evaluates. The main difference between these values is referred to as the leftover. A straight relapse indicates that the total squared residuals should be minimized. The phrase science of a straight relapse is given by

$$Y = a_0 + a_1X + \epsilon.$$

The algorithm begins determining the optimum fit line for using a_0 and a_1 when the client conducts direct relapse. It is therefore more precise than the real data points. We may use the model to assess the output because we are aware of the values of a_0 and a_1 .

5. SYSTEM DESIGN

Software design is the process of translating requirements into a software representation. After software requirements have been specified, software design involves three technical activities like design, code generation, and testing. Systems are designed in a modular fashion. H. Software is logically divided into components that perform specific functions and subfunctions. During the design phase, modules with independent functional characteristics are created. It also leads to interfaces that reduce the complexity of connections between modules and with the external environment. The design phase is the most important because the decisions made during this phase ultimately affect the success of the implementation and maintenance of the software.

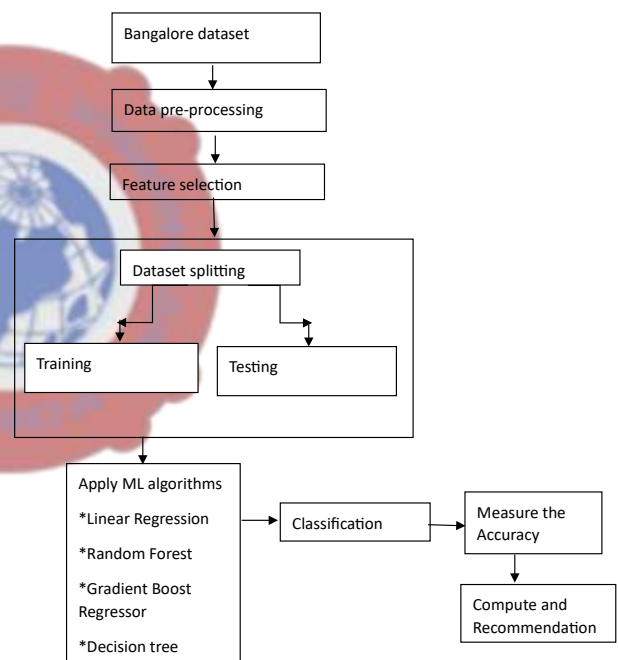


Fig. System Architecture

Grouping charts demonstrate the stream of rationale inside your framework in a visual way, empowering you both to report and approve your rationale, and are commonly utilized for both examination and plan purposes. Arrangement graphs are the most prevalent UML artefact for energetic modelling, which centres on distinguishing the behaviour inside your framework. Other energetic modelling methods incorporate movement diagramming,

communication diagramming, timing diagramming, and interaction outline diagramming. Arrangement graphs, along with course charts and physical information models are in my supposition the most imperative design-level models for cutting edge trade application advancement. These have appeared as labelled bolt sets close a connect. These messages are communications between objects that pass on information roughly the activity and can consolidate the course of action number. The most crucial objects are set in the middle of the chart, with all other sharing objects branching off. After all, objects are put, joins and messages should to be included in between.

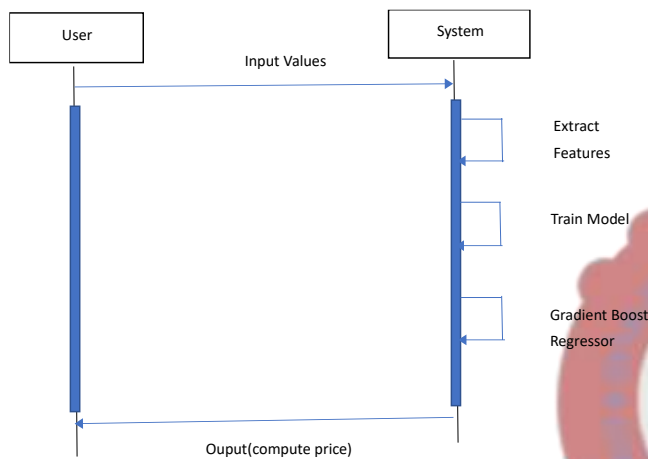


Fig. Sequence Diagram

6. IMPLEMENTATION

```
In [41]: data
```

```
Out[41]:
```

	location	size	total_sqft	bath	price	bhk	price_per_sqft
0	1st Block,Jayanagar	4 BHK	2650.0	4.0	428.0	4	15017.543880
1	1st Block,Jayanagar	3 BHK	1630.0	3.0	194.0	3	11901.840401
2	1st Block,Jayanagar	3 BHK	1875.0	2.0	236.0	3	12533.333333
3	1st Block,Jayanagar	3 BHK	1200.0	2.0	130.0	3	10633.333333
4	1st Block,Jayanagar	2 BHK	1235.0	2.0	148.0	2	11963.805668
...
10282	other	2 BHK	1200.0	2.0	70.0	2	5633.333333
10283	other	1 BHK	1800.0	1.0	200.0	1	11111.111111
10286	other	2 BHK	4363.0	2.0	110.0	2	8130.081301
10297	other	1 Bedroom	812.0	1.0	26.0	1	3201.970443
10300	other	4 BHK	2600.0	5.0	400.0	4	11111.111111

7361 rows x 7 columns

Fig. Data from CSV file

Substance relationship charts give a visual beginning point for a database plan that can moreover be utilized to offer assistance decide data framework prerequisites all through an organization. After a social database is rolled out, an ERD can still serve as a referral point, should to any investigating or trade prepare reengineering be required afterwards.

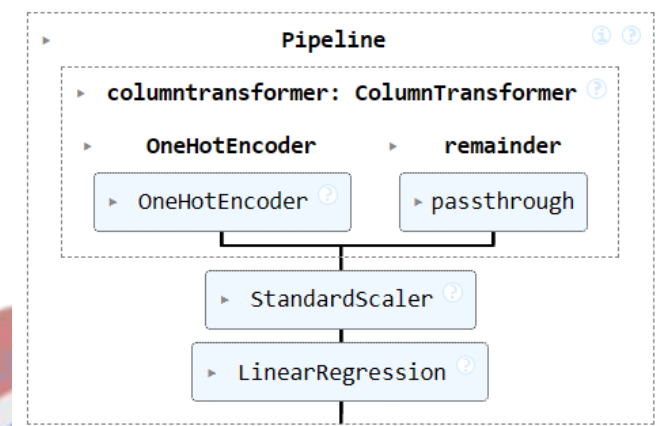


Fig. Linear Regression



Fig. Estimated Price of the Object

7.SOFTWARE TESTING

The Program Testing Life Cycle (STLC) is an efficient approach to testing a program application to guarantee that it meets the prerequisites and is free of abandons. It is a handle that takes after an arrangement of steps or stages, and each stage has particular targets and deliverables. The STLC is utilised to guarantee that the program is of high quality, solid, and meets the needs of the end-users. The fundamental objective of the STLC is to

distinguish and report any abandons or issues in the program application as early as conceivable in the improvement handle. This permits issues to be tended to and settled sometime after the program is discharged to the public. The stages of the STLC incorporate Test Arranging, Test Investigation, Test Plan, Test Environment Setup, Test Execution, Test Closure, and Deformity Retesting. Each of these stages incorporates particular exercises and deliverables that offer assistance to guarantee that the program is completely tried and meets the necessities of the conclusion clients. The Program Testing Life Cycle (STLC) is an efficient approach to testing a program application to guarantee that it meets the prerequisites and is free of abandons. It is a handle that takes after an arrangement of steps or stages, and each stage has particular targets and deliverables. The STLC is utilised to guarantee that the program is of high quality, solid, and meets the needs of the end-users. The fundamental objective of the STLC is to distinguish and report any abandons or issues in the program application as early as conceivable in the improvement handle.

8.CONCLUSION

In this project, we have considered four machine learning algorithms for house price computation and recommendation: linear regression, decision tree, random forest, and gradient boost. Comparing the predictive accuracy and overall performance, it can be established that gradient boost stands out from the rest. It has high accuracy, works on different types of features, and is able to capture complex interactions quite easily.

Linear regression provides a baseline and a simple model for house price prediction, but its linearity assumption can be too restrictive to capture the complex patterns in the data. Decision trees may be quite interpretable but are prone to overfitting. Random Forest resolves the overfitting problem but is rather computationally expensive. Gradient Boosting combines advantages of decision trees with regularization techniques to gain superior performance.

9.REFERENCES

- Key, T. Sch, and J. A. Med, "Artificial Intelligence & Medical Diagnosis," Sch. J. Appl. Med. Sci. (SJAMS) Artif. Intell. Med. Diagnosis, no. December, pp. 4982-4985, 2019.
- J. B. O. Mitchell B.O., "Machine Learning Methods in Chemoinformatics," Wiley Interdiscip. Rev. Comput. Mol. Sci., vol. 4, no. 5, pp. 468-481, 2014.
- K. H. Sadia, A. Sharma, A. Paul, S. Padhi, and S. Sanyal, "Stock Market Prediction Using Machine Learning Algorithms," Int. J. Eng. Adv. Technol., vol. 8, no. 4, pp. 25-31, 2019.
- Afonso, Bruno, Luckeciano Melo, Willian Oliveira, Samuel Sousa, and Lilian Berton. "Housing prices prediction with a deep learning and random forest ensemble." In Anais do XVI encontro nacional de inteligência artificial e computacional, pp. 389-400. SBC, 2019.
- Hjort, Anders, Johan Pensar, Ida Scheel, and Dag Einar Sommervoll. "House price prediction with gradient boosted trees under different loss functions." Journal of Property Research 39, no. 4 (2022): 338-364.
- Thuraiya Mohd , Syafiqah Jamil , Suraya Masrom, "Machine learning building price prediction with green building determinant" IAES International Journal of Artificial Intelligence (IJ-AI) Vol. 9, No. 3, September 2020, pp. 379~386, September 2020
- Khan Sohail Liyaqatullah , Usman Malik , Qureshi Mohammed Basheeruddin, "HOUSE PRICE PREDICTION USING MACHINE LEARNING", International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), Vol. 10, Issue 4, April 2021.
- V. Vijay Bhasker Reddy, Burugu Chandana, Burugu Chandana, Tandava Charitha, Sravya Siripragada, "HOUSE PRICE PREDICTION", YMER ISSN : 0044-0477, VOLUME 21 , 2022.