

HOUSE PRICE PREDICTION USING MACHINE LEARNING

Prof. Archana lomte¹, Nikita Bhagwat², Kaveri Adhav³, Trupti bhavar⁴

¹Prof. Archana Lomte, Dept. of Computer Engineering, BSIOTR, Pune, Maharashtra, India

²Nikita Bhagwat, Dept. of Computer Engineering, BSIOTR, Pune, Maharashtra, India

³Kaveri Adhav, Dept. of Computer Engineering, BSIOTR, Pune, Maharashtra, India

⁴Trupti Bhavar, Dept. of Computer Engineering, BSIOTR, Pune, Maharashtra, India

Abstract - The development of the falling or rising of the house costs has attracted interest from the scientist as well as several alternative interested parties. There are several previous researches works that used numerous regression techniques to deal with the question of the changes house value. This work considers drawback of fixing house value as a classification problem and applies machine learning techniques to predict whether or not house costs can rise or fall. This work applies numerous feature choice techniques like variance influence issue, info price, principal element analysis and information transformation techniques like outlier and missing price treatment further as box-cox transformation techniques. The performance of the machine learning techniques is measured by the four parameters of accuracy, precision, specificity and sensitivity. The work considers 2 distinct values zero and one as various categories. If the worth of the category is zero then we tend to think about that the value of the house has faded and if the worth of the category is one then we tend to think about that the value of the house has raised.

Key Words: Machine Learning, House price prediction.

1. INTRODUCTION

Development of civilization is that the foundation of increase of demand of homes day by day. correct prediction of house costs has been continuously a fascination for the patrons, sellers and for the bankers additionally several researchers have already worked to unravel the mysteries of the prediction of the house worth. There area unit several theories that are born as a consequence of the analysis work contributed by the assorted researchers everywhere the planet. a number of these theories believe that the geographical location and culture of a selected space verify however the house costs can increase or decrease whereas there are a unit different college of thought United Nations agency emphasize the socio-economic conditions that for the most part play behind these house worth rises. we tend to all recognize that house worth may be a variety from some outlined assortment, thus clearly prediction of costs of homes may be a regression task. To forecast house worth one person typically tries to find similar properties at his or her neighbourhood and supported collected information that person can attempt to predict the house worth.

Machine learning (ML) is an area of artificial intelligence which works manually using trained model input Paid during the training period. Machine learning is used to perform a lot of computing tasks. It is also used to make predictions with the use of computers. Machine learning is sometimes also used to devise complex models. The principle point of machine learning is to permit the PCs to learn things naturally without the assistance of people. Machine learning is very useful and is widely used around the whole world. The process of machine learning involves providing data and then training the computers by building machine learning models with the help of various algorithms.

1.2 Python

Python is Associate in Nursing taken high-level all-purpose artificial language. Python's style philosophy emphasizes code readability with its notable use of great indentation. Its language constructs still as its object-oriented approach aim to assist programmers write clear, logical code for tiny and large-scale comes.

Python is dynamically-typed and garbage-collected. It supports multiple programming paradigms, as well as structured (particularly, procedural), object-oriented and purposeful programming. Python is usually represented as a "batteries included" language thanks to its comprehensive normal library.

2. LITERATURE REVIEW

There are a couple of components that impact house costs. In this exploration, partition these components into three essential get-togethers, there are state of being, thought and territory. States of being our properties constrained by a house that can be seen by human recognizes, including the range of the house, the number of rooms, the availability of kitchen and parking space, the openness of the yard nursery, the zone of land and structures, and the age of the house, while the thought is an idea offered by architects who can pull in potential buyers, for instance, the possibility of a moderate home, strong and green condition, and world class condition. Zone is a critical factor in shaping the expense of a house. This is in light of the fact that the zone chooses the normal land cost. Besides, the territory furthermore chooses

the basic passage to open workplaces, for instance, schools, grounds, crisis facilities and prosperity centers, similarly as family preoccupation workplaces, for instance, strip malls, culinary visits, or much offer awesome landscape.

3. METHODOLOGY

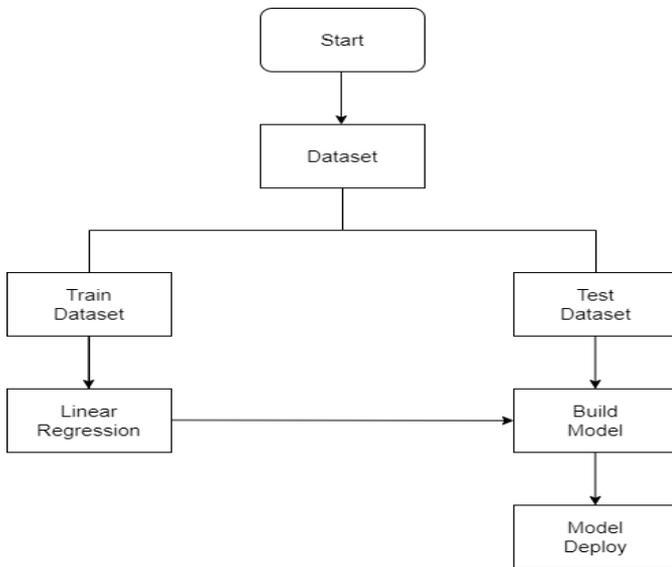


Fig.1. Block Diagram

3.1 Dataset:

A data set may be an assortment of information. within the case of tabular information, an information set corresponds to 1 or additional info tables, wherever each column of a table represents a selected variable, and every row corresponds to a given record of the info set in question.

3.2 coaching data:

A coaching dataset may be a dataset of examples used throughout the training method and is employed to suit the parameters (e.g., weights) of, for instance, a classifier. For classification tasks, a supervised learning formula appearance at the coaching dataset to work out, or learn, the best mixtures of variables which will generate an honest prognosticative model. The goal is to provide a trained (fitted) model that generalizes well to new, unknown information. The fitted model is evaluated victimization "new" examples from the held-out information sets (validation and take a look at datasets) to estimate the model's accuracy in classifying new data. to scale back the chance of problems like overfitting, the examples within the validation and take a look at datasets shouldn't be accustomed train the model. Most approaches that search through coaching information for empirical relationships tend to overfit the info, which means that they will determine and exploit apparent relationships within the coaching information that don't hold normally.

4. WORKING OF SVM:

In machine learning, support-vector machines (SVMs, additionally support-vector networks) are supervised learning models with connected learning algorithms that analyses knowledge used for classification and multivariate analysis. Developed at AT&T Bell Laboratories by Vapnik with, it presents one in all the foremost strong prediction ways, supported the applied mathematics learning framework or VC theory projected by Vapnik and Chervonenkis and Vapnik. Given a collection of coaching examples, every marked as happiness to at least one or the opposite of 2 classes, AN SVM coaching formula builds a model that assigns new examples to at least one class or the opposite, creating it a non- probabilistic binary linear classifier. AN SVM model could be an illustration of the examples as points in area, mapped so the instances of the separate categories are split by a transparent gap that's as wide as probable. New examples r then scheduled into that very same area and foreseen to belong to a class supported the aspect of the gap on that they fall.

4.1 APPLICATION OF SVM -

- SVMs area unit useful in text and machine-readable text categorization, as their application will considerably scale back the necessity for labeled coaching instances in each the quality inductive and transudative settings. Some strategies for shallow linguistics parsing area unit supported support vector machines
- Classification of pictures may also be performed victimisation SVMs. Experimental results show that SVMs attain considerably higher search accuracy than ancient question refinement schemes when simply 3 to four rounds of relevancy feedback. this can be additionally true for image segmentation systems, together with those employing a changed version SVM that uses the privileged approach as advised by Vapnik.
- The SVM formula has been wide applied within the biological and alternative sciences. they need been accustomed classify proteins with up to ninetieth of the compounds classified properly. Permutation tests supported SVM weights are advised as a mechanism for interpretation of SVM models. Support-vector machine weights have additionally been accustomed interpret SVM models within the past

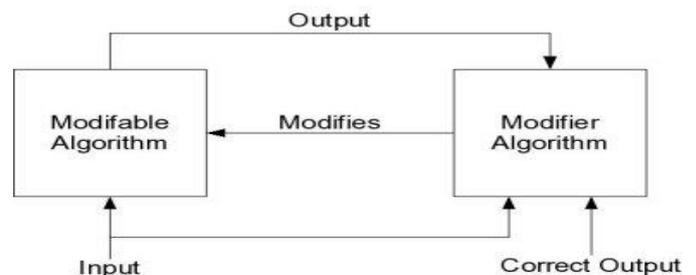


Fig 2.SVM

5. CHARACTERISTICS OF MACHINE LEARNING:

1.The ability to perform machine-driven information visualization:

A massive quantity {of information} of knowledge| of information} is being generated by businesses and customary folks on a daily basis by visualizing notable relationships in data, businesses can't solely build higher choices however build confidence also.

2.Automation at its best:

One of the largest characteristics of machine learning is its ability to change repetitive tasks and so, increasing productivity .A huge range of organizations are already victimization machine learning-powered work and email automation.

3.Customer engagement like ne'er before:

For any business, one in all the foremost crucial ways that to drive engagement, promote complete loyalty and establish durable client relationships is by triggering purposeful conversations with its target client base. Machine learning plays an important role in sanctioning businesses and types to spark a lot of valuable conversations in terms of client engagement.

4.Accurate information analysis:

As mil algorithms gain expertise, they keep up in accuracy and potency. This lets them build higher choices. Say you wish to create a weather outlook model. because the quantity of information, you have got keeps growing, your algorithms learn to create a lot of correct predictions quicker.

Machine learning has assisted to finish out task. Firstly, the information assortment is performed. Then information improvement is administered to get rid of all the errors from the information and build it clean. Then the information pre-processing is finished. Then with facilitate of knowledge visualized particle, completely different plots are created. This has portrayed the distribution of knowledge in several forms. Further, the preparation and testing of the model are performed. It has been found that a number of the classification algorithms were applied on our dataset whereas some weren't. So, those algorithms that weren't being applied on our house pricing dataset are born and tried to enhance the accuracy and exactness of these algorithms that were being applied on our house valuation dataset.

REFERENCES

- 1 Lisi G. Hedonic pricing models and residual house price volatility. *Letters in Spatial and Resource Sciences*, 2019: 1-10.
- 2 Caterina Contini, Fabio Boncinelli, Caterina Romano, Gabriele Scozzafava, Leonardo Casini. Price vector issue in a choice experiment: A methodological proposal. *Food Quality and Preference*, 2019,75.
- 3 Ang L I , Yi-Xiang L I , Xue-Hui L I . TensorFlow and Keras-based Convolutional Neural Network in CAT Image Recognition. 2017.
- 4 Wisanlaya Pornprakun, et al. Determining optimal policies for sugarcane harvesting in Thailand using bi-objective and quasi-Newton optimization methods. *Advances in Difference Equations*, 2019, Vol.2019 (1), pp.1-15
- 5 Dayhoff J E, Deleo J M. Artificial neural networks. *Cancer*, 2001, 91(8):1615-1634.
- 6 Lu Lili, Hu Bin, Li Hui, Duanmu Yiting. Simulation Analysis of China's Housing Price Composition and Forecast. *Computer Simulation*, 2014, 31(03):230-238.

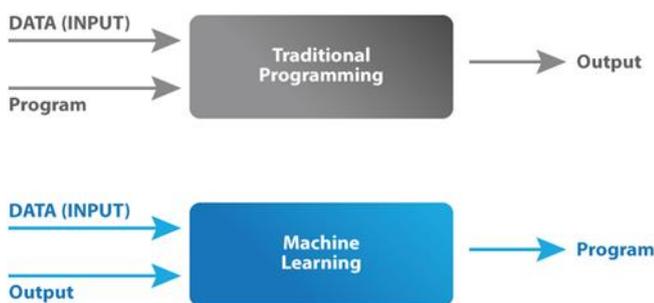


Fig -3: Machine Learning

6. CONCLUSIONS

The sales worth for the homes are calculated victimization completely different algorithms. The sales costs are calculated with better accuracy and exactness. this may be of nice facilitate for the folks. to attain these results, numerous information mining techniques are used in python language. The various factors that have an effect on the house valuation ought to be considered and work upon them.