HOUSE PRICE PREDICTION USING MACHINE LEARNING

Dharani.R*,Banumathi.M[1],Soundarya.S[2].

**Assistant professor,department of Information technology,Panimalar institute of technology,391, Bangalore TrunkRoad,Poonamallee, Chennai, E-Mail:dharani.rit@gmail.com*

1,Student, department of Information technology,Panimalar institute of technology,391, Bangalore Trunk Road,Poonamallee, Chennai, E-Mail:banusuba72@gmail.com

2,Student, department of Information technology, Panimalar Institute of Technology, 391, Bangalore Trunk Road,Poonamallee, Chennai, E-Mail:bhavsiva@gmail.com

ABSTRACT

House prices increases every year so this makes difficult for customers to evaluate house price based on existing values at the time of buying. This motivated us to propose a system that predicts the price of the houses using various Machine Learning algorithms such as AdaBoost, XGBoost, LightGBM&LogisticRegressionalgorithms.The prediction is done using the features of the previously sold houses over the years.Using the interior and exterior features of the house, the appropriate price can be learned by the algorithms.The algorithm is trained and the price of the house in the near future can be predicted.The prediction would be very much useful in real estate fields where the customers can check the appropriate price of the house before even visiting the site of the house.

INTRODUCTION

The price of House increases every year which makes complications for customers to predict the price of the house during the time of buying. This motivated us to propose a system that predicts the price of the houses using various Machine Learning algorithms. Here prediction is done using the features of the previously sold houses over the years. Thus, by using the interior and exterior features of the house, the appropriate price of a house can be learned and predicted by such machine learning algorithms. Hence, The algorithm is trained so that the price of the house in the near future can be predicted. This prediction system finds its major application in fields like real-estate.

PROPOSED SYSTEM

In our Proposed system, we make use of machine learning algorithms such as XGBoost, LightGBM in order to carry out the various phases of house price prediction such as Data preprocessing, Training the model and testing the model. Here, the dataset is initially preprocessed by removing the unwanted data, null values and labeling textual values. The preprocessed data is split into two for the purpose of testing and training, which is then sent through the applied machine learning algorithms for

training the model with the data and then testing with the test data. Finally, the real time data is provided to the algorithm for predicting the price of thehouse.

MODULES SPLITUP

- DataPreprocessing.
- Training the model.
- Testing the trained algorithm.

DATAPREPROCESSING

Data Preprocessing is actually a crucial procedure of making the available dataset suitable for machine learning algorithm. This involves data mining techniques to transform the raw data into efficient format; the dataset is being initially preprocessed by removing the unwanted data and extracting only the data relevant to the problem, formatting the data so that the data should be suitable for machine learning algorithm which is accomplished by the removal of null values and irrelevant data, labeling textual values so that data can be cleaned and transformed into the required format. Here the data set consists of the various features of the house which helps in the process of price prediction and the data are extracted based on the feature selector and finally the standardized data is sent to the successive modules.

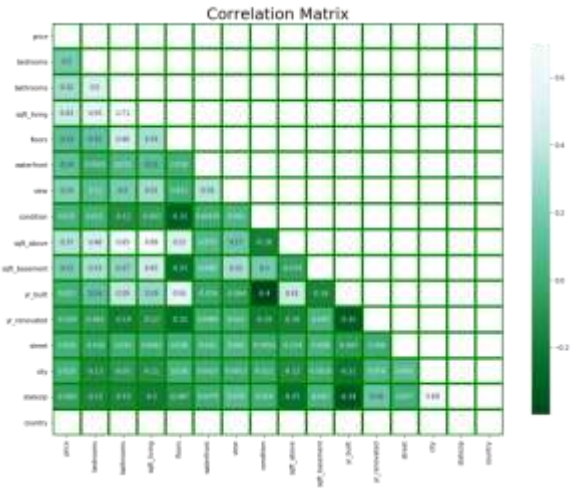
TRAINING

We get data from the data splitting process from which the training data is extracted and thus the extracted data is trained using the machine learning algorithm such as**Light GBM & XG Boost.**

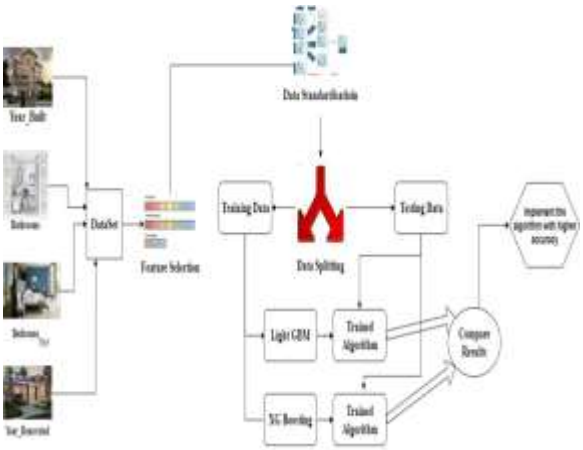
TESTING

Thus, the trained algorithms are tested by comparing the high accuracy in the results while implementing the algorithm, also the model is tested.

HEATMAP



BLOCK DIAGRAM



ALGORITHMS

LIGHTGBM

LightGBM(L-GBM) makes use of frameworks such as gradient boosting which relies on tree- based learning algorithms. Unlike usual trees which grows horizontally which are usually leaf-wise methods, vertical trees are grown in L-GBM and hence it archives level-wise method. The maximum valued leaf will be chosen so that the loss could be reduced

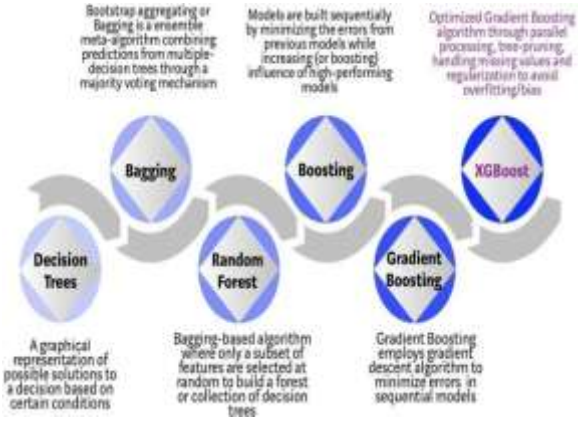
during growing. Leaf-wise growing have shown more reduction in loss when compared to level-wise algorithm. Due to gradual increase in the size and amount of data, traditional data science algorithms fails to show up better results with expected accuracy. The algorithm used here is capable of handling a very large amount of data (knowledge) with comparatively less amount of memory allocation and also achieved higher speed when compared to other algorithms. On behalf of above reasons, this algorithm comes with name as L-GBM where the "L" here represents Light. It popularity further increased due it its ability in providing results of massive data with improved accuracy. Another notable advantage of L-GBM is that it can also be used along with GPU learning which makes it to find it's position in wide data science applications development.

XGBOOST

XGboost algorithm is a boosting algorithm which creates multiple decision trees consecutively with one decision tree leadin to the creation of another.

In this project, the algorithm creates decision tree for each of the column in the dataset by segregating each to an average value and splitting the column under maximum of two conditions.

After the column is separated a value for each condition is assigned which is the average value of the column data under this particular condition.



In this way the algorithm proceeds with each step updating the value of the conditions by considering the current condition of the column and also the previous one.

HARDWARE AND SOFTWARE CONSTRUCTION

HARDWARE:RAM: 8GB and above,Processor: i5 and above.

SOFTWARE: OS: linux OS or Windows 10,**TOOL**: Jupyter NoteBook.

CONCLUSION

We have overseen a model that could give any real-estate person or clients, a novel methodology with higher accuracy for the

prediction of house prices. A couple of backslide methodologies have been explored. Furthermore, boosting algorithms are used for the improvement of accuracy in predictions. Another advantage of this system is that we make use of various processors parallelly instead of formal sequential processing which ultimately leads to faster performance, reduction in time required for certain initial computations and enhanced accuracy in results even with massive amounts of data.