PREDICTING AND IDENTIFYING THE REAL ESTATE OPPORTUNITIES USING MACHINE LEARNING

[1]Ayush Sharma, [2]Chandan Singh, [3]Priyank Rawat, [4]Krish Rawat,

Ms. Arushi Gupta

1,2,3,4 Student, Department of Information Technology, Dr. Akhilesh Das Gupta Institute of Technology & Management, Delhi, India.

Assistant Professor, Department of Information Technology, Dr. Akhilesh Das Gupta Institute of Technology & Management, Delhi, India.

ABSTRACT

The real estate market fluctuates a lot regarding pricing as real estate price depends upon many factors. So, it becomes one of the best prime fields to use the ideas of machine learning to enhance and foresee the prices with high accuracy. The purpose of the paper is that the prediction of the market value of a real estate property. This system helps us to find a starting price for a property based on the geographical factors. By breaking down past market patterns and value ranges, and coming advancements future costs are going to be anticipated. This technique helps us to predict house prices in a city with Decision tree regressor. It'll help clients to place resources into an estate without moving towards a broker. The results acquired by this research proves that the decision tree regressor gives an accuracy of 88%.

KEYWORDS: Decision tree regressor, machine learning, house price prediction, real estate.

I.   INTRODUCTION

In modern world, applications of artificial intelligence (AI) are found everywhere. It's prevalent to an extent that one could be using it and not even remember it. Machine Learning is one of the most widespread applications of AI where devices and software perform human brain like cognitive capabilities. Every organization in today's real estate business is working effectively to attain a competitive edge over other competitors. There's a desire to simplify the method for a normal person while providing the best results. This paper aims to develop a system that will predict the price of houses using regression machine learning algorithm. This model is built not only for the houses that are for sale but also for the houses that are under construction.

Regression is a machine learning apparatus that encourages you to form expectations by taking in from the current measurable information. A house's cost relies upon parameters, as an example, the quantity of rooms, living region, area etc. On the off chance that we apply counterfeit determining a way to these parameters, we are able to compute house valuations in a particular land region.

In this model the target feature of the house property is the price and also the independent features are: no. of bedrooms, no. of bathrooms, built-up area, the floor, zip code and age of the property. Aside from those of the mentioned features, which are generally required for predicting the house prices, we've included two other features - air quality and rate. These features provide a valuable contribution towards predicting property prices since the upper values of those features will result in devaluation of house prices. The whole implementation of this model is done using the python language. To construct the predictive model, a Decision tree regressor is used from the Scikit-learn machine learning library. Grid Search CV helps to find the best max-depth value to construct the decision tree.

II.   METHODOLOGY

A. Collection of data

Data processing techniques and processes are numerous. We collected data for 's real Delhi estate properties from various real estate websites. The data would be having attributes such as Location, carpet area, built-up area, age of the property, zip code, etc. We must collect the quantitative data which is structured and categorized. Data collection is needed before any kind of machine learning research is carried out. Dataset validity is a must otherwise there is no point in analyzing the data.

B. Data preprocessing

Data preprocessing is a vital process which involves initial investigations on data, performed so as to check hypothesis, discover patterns, detect anomalies, or check any assumptions using graphical representations and statistical summaries. There might be missing values or outliers in dataset. These outliers occur due to existence of errors of some sort during data collection, and are needed to be eliminated for improving the model's performance. It can be handled by data cleaning. So, the data which we collected must be prepared and refined before we apply ML techniques to it for analysis. Age and floor parameters were handled for their missing values. The target attribute is also dropped off from the training dataset. Pandas library is used for this purpose. For statistical visualization of the dataset, the min, max, standard deviation, mean of the target attribute were found out. We split the dataset into a training set (80%) and a test set (20%).

C. Studied algorithms

Various regression algorithms were studied in the process of developing this model such as SVM, Random Forest, Linear regression, Decision Tree Regressor, KNN, all were tested upon the training dataset. Although, the decision tree regressor showed the highest accuracy in terms of predicting the house prices. The decision by which we choose the algorithm highly depends upon the dimensions and type of data used. The decision tree algorithm suited best for our dataset.

D. Decision tree regressor

The decision tree regressor perceives features of an attribute and trains a model within the style of a tree to predict future prices to provide meaningful output. The max depth and min depth of a graph helps the decision tree regressor to learn and it analyzes the data according to the system. Methods of regression are categorized under supervised ML. Based on a prior dataset, regression methods facilitate in explaining or predicting the computed numerical value, like a property's price prediction based on data of prices for similar properties.

E. Max Depth

As mentioned earlier grid search cv helps to search out max depth for the tree. We've used matplotlib to visualize the various max-depths and complexity performance. By using Grid Search CV we can deal with parameter tuning that will efficiently manufacture and assess a model for for each combination of calculation parameters indicated in a grid. The best value for max-depth is assessed using Grid Search CV, using which the decision tree is constructed.

F. Fitting the model

Since the data is split down into two modules: A Training set and Test set, we must initially train the model. To train the dataset decision tree regressor algorithm is applied on it. The Decision tree builds a regression model in the form of a tree structure. The trained model is applied to check dataset and house prices are predicted. We have used the predict function to predict the test set results.
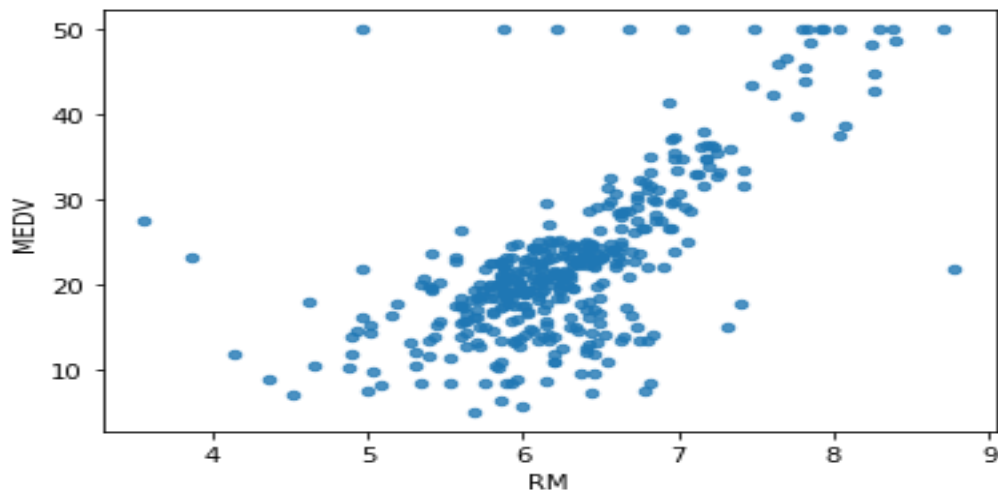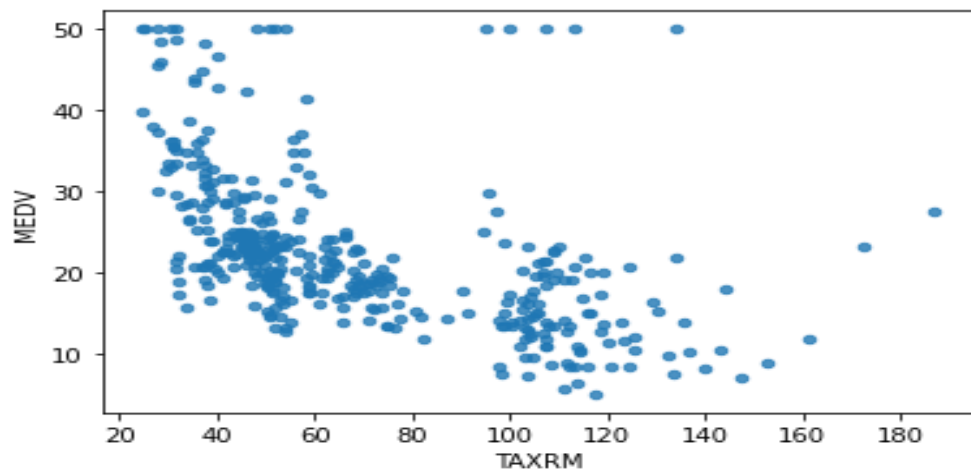
Fig 1. The generic flow of development

III.   RESULTS

The following shows the plot of predicted vs actual prices with the accuracy of prediction:

Accuracy is nothing but the r2 score of the regression model.

IV. CONCLUSION

In this paper, the Decision tree algorithm is used to construct a prediction model to predict potential selling prices for any real estate property. Extra features like air quality and crime rate were included within the dataset that will assist in predicting the costs even better. In most of the datasets of other prediction systems these features are not included, which makes this system different. These kind of features influence people's decision while purchasing a property, so why not include it in predicting house prices. The system provides 89% accuracy while predicting the prices for the real estate. Based on the results, it can be concluded that ML-driven predictions are easily understandable and significant from a data-analytics point of view. When correctly implemented, a high rate of accuracy may be achieved, and thus ML techniques find applications across a wide range of fields.

V. REFERENCES

1. Theobald, O. (2017). Machine learning for absolute beginners.

2. Kelleher, J. D., Mac Namee, B., &D'arcy, A. (2015). Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies. MIT press.

3. Minaie, N. (2019). A Beginner's Guide to Selecting Machine Learning Predictive Models in Python, https://towardsdatascience.com/the-beginners-guide-to-selecting-machine-learning-predictive-models-in-python-f2eb594e4ddc

4. Park, B., & Bae, J. K. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data. Expert Systems with Applications, 42(6), 2928-2934.

5. Satish, G. N., Raghavendran, Ch. V., Rao, M. D. S., Srinivasulu, Ch. (2019). House Price Prediction Using Machine Learning.International Journal of Innovative Technology and Exploring Engineering, 8(9), 2278-3075.

6. Baldominos, A., Blanco, I., Moreno, A. J., Iturrarte, R., Bernárdez, Ó., & Afonso, C. (2018). Identifying Real Estate Opportunities Using Machine Learning. Applied Sciences, 8(11), 2321.

7. Sarkar, S. (Sep, 2019). Predicting House prices using Classical Machine Learning and Deep Learning techniques, https://medium.com/analytics-vidhya/predicting-house-prices-using-classical-machine-learning-and-deep-learning-techniques-ad4e55945e2d.