ISSN 2348-0831

Determining House Price Using Regression

Anirban Chakraborty[1]

[1]MTECH, Artificial Intelligence
Lovely Professional University, Phagwara, Punjab, India

DOI: 10.29322/10.5281/zenodo.3956571
https://zenodo.org/record/3956571#.Xy_M-4ozbIU

Abstract- *The purpose of this article is to estimate the purchasing and sale opportunities of houses on the market by Machine learning techniques. For financial stability, the housing sector is quite critical. People seeking to purchase a new house appear to be more cautious in their expectations and sales tactics analyzing historical industry patterns and pricing levels, as well as potential changes. The index of our method consists of the price of the house and its efficiency metrics, such as the amount of renovation, the distance from the city center, the construction programs, the height of the property, the floor and the location of the apartment in the home, and there are several other criteria. Service includes a database that recognizes the preferences of its clients and then integrates machine learning software. The program will enable consumers invest in real estate without approaching brokers. It therefore reduces the uncertainties inherent with the deal. The program has a login ID and a pin. At the same time, when the user searches for an attribute, the value of the original attribute and the value of the predicted attribute are displayed.*

Keywords*: House prices, machine learning algorithms-random forest, decision trees, price prediction, market trends.*

1. INTRODUCTION

The basic needs in life are food, shelter and safety. This ensures that the efficiency and happiness of people are improved. Therefore households are a critical factor in human capital growth in any economy. Society growth is a pillar of the relentless rise in household demand. The accurate forecast of home prices has also helped both buyers and sellers.

Machine learning develops and uses algorithms to predict new data and develop computer models. The main difference between standard algorithms is that the model is not just running a series of instructions but is based upon input data. It involves a reference data set and a test data set. Supervised learning uses secret data, while unattended learning uses anonymous data. There are a number of common algorithms, including regression, the classification of wild forests, and a decision book.

This topic brings together the latest research on forecasting markets to further enhance their use by economic forecasters. It offers a overview of the future markets, as well as the actual markets, which are helpful in analyzing[1] the sector, helping to make accurate forecasts. This project utilizes machine learning algorithms to forecast prices by evaluating existing house prices, thus forecasting potential prices according to consumer requirements.

2. METHODOLOGY

2.1 Data Gathering and Processing

Many databases and other online portals of the housing are obtained from the results. For each search, the area of the city varies greatly[2]. This work covers all information's of housing.

The analysis did not include the features of unimaginably low or high value because they represent a relatively small part of the remaining data collection. The following statistics shows around 1,000 apartments created by the promoter in the area. Any of these houses must be made manually to make correct predictions. As a consequence, certain features are necessary. For example, some of these houses have costly interior appliances, furniture, electricity and many other pricing requirements[3].

2.2 Feature Extraction and Selection

A lot of different features were extracted from all the data. Many of the characteristics of the commodity is included in the house such as pricing, parking, furniture, etc. Incorrectly recorded, outliers have been removed by manual inspection, such as a room, a hall and a kitchen, which represent only a small part of the total data. It also involves certain functions that have an effect on the Indian industry, such as Vastu[4], interior design and furniture. This work uses techniques for selecting features such as information value, main component analysis and data conversion techniques. Used for the selection of features and the subsequent conversion process. Such methods shall be employed in the following manner:

	area	bathroom	balconies	Lift Available	Car Parking	Power Backup	24 X 7 Security	Children's play area	Vaastu Compliant	Club House	...
0	1611	3	2	1	1	1	0	0	0	0	...
1	1279	2	0	1	1	1	1	0	1	1	...
2	670	1	0	1	1	1	1	1	1	1	...
3	956	2	0	1	1	1	1	1	1	1	...
4	1706	3	1	1	1	1	1	1	0	0	...
5	1015	2	2	1	1	1	1	1	0	1	...
6	3725	5	0	1	1	1	1	1	1	1	...
7	450	1	0	1	1	1	1	1	1	1	...
8	1650	3	2	1	1	1	1	0	1	1	...
9	449	1	1	1	0	1	1	1	0	1	...
10	625	1	1	1	1	1	1	1	1	1	...
11	900	1	1	1	1	1	1	1	1	1	...
12	1100	2	1	0	1	1	1	1	0	0	...
13	3300	5	3	1	1	1	1	1	1	1	...
14	1305	2	2	1	1	1	1	0	0	0	...
15	1500	2	1	1	1	1	1	1	1	1	...

Figure 1: Features selection for CART

2.3 Proposed Architecture4

The domestic market prediction model[5] presented in this research helps buyers and sellers forecast house prices accurately by offering 15 functions of CSV data. In this case, we will use a range of features as input parameters in accordance with the owner or buyer's preferences. In order to make forecasts, we used decision trees and random algorithms. The first goal of this study is to identify the necessary quality parameters. Its a bit of knowledge. For example, square foot, location, room, power, balcony. Secondly, the relation between parameters across the training data set and the assessment data set will be examined. Third, machine learning algorithms are used to construct models and forecast outcomes.

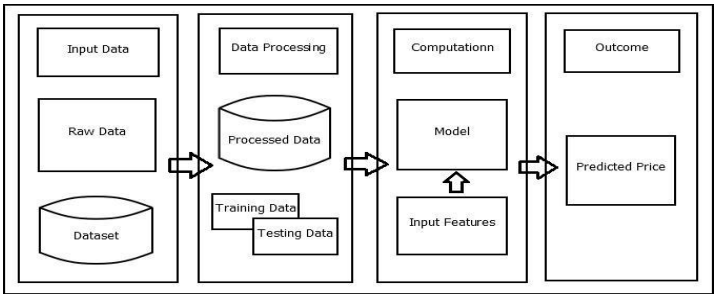

Figure 2: System Architecture

2.4 Algorithm 1. Decision Tree

Decision trees[6] are a subset of technology such as linear regression for data mining. The tree of judgment includes network nodes. A main node tree with several leaf nodes. It is a tree with a node identified as a root that does not have incoming but only outgoing corners. The remaining nodes have their inputs. The outbound and inbound node calls the central node. Each intermediate node is divided or split into two or more sub trees, depending on the importance of the input variable. Each check considers a single attribute for elementary and other instances for sub trees to be divided by the value of the attributes. A range is used in the case with quantitative characteristics.

That variable is assigned a class which represents the most significant target value. In other words, a leaf may provide a function of chance to measure the likelihood of a specific target value. Objects are determined as part of the route checks from the tree parent node to the child node.

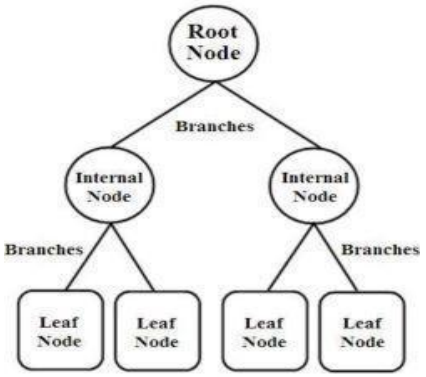

Figure 3: Decision Tree

ISSN 2348-0831

2.5 Algorithm 2. Random Forest

Random Forests are ensemble classifier built from the Decision Trees set. The value of the classifier are the decision Trees' value mode. The marked **Bremen principle** fusion with a random collection of applications. Random forests The Random Trees algorithms were developed by **Leo Bremen** and **Adele**. Plant trees for regression, generating a random forest is done.

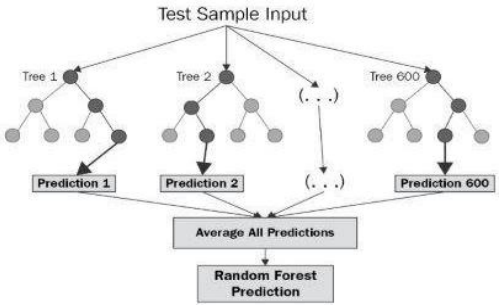

Figure 4: Random Forest Algorithm

2.6 Theory and Working

Additionally, in the CSV file and also in the archive, we have obtained and stored data from other reliable sources and web portals. This is then transferred to **kimchi disk** (unstructured data file). Which is used for ordered and unstructured material processing. In the Python file configuration, we can pick up any object in python for storage on the disk.

The kimchi file is read and stored in the piece using pandas. **Pandas** is an open source library offering high-performance, user-friendly data models and resources like Python 's language data analysis programming.

2.7 Used Sklearn-learning for data analysis

Sklearn software is a simple and efficient tool for data processing and data extraction (formerly Scikits training are now called sklearn). This is a free Python programming language computer research library. This is built in library with the Python Virtual and Scientific Library of the **NumPy** and **SciPy**. This provides a variety of classifications, regressive regressions, clusters comprising vector support facilities, random forestry, gradient boosters, k-medians, and DBSCAN.

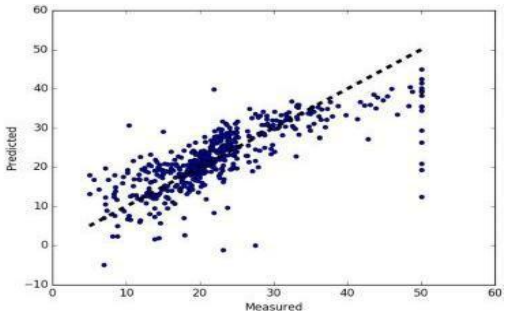

Figure 5: Sklearn for random forest

We then used the analysis of the Sklearn Decision Tree. It built Tree and a model decision system. The model is selected for additional testing and the model is then evaluated against the collection of reference data. Horoscope proves the function of the random regression of forests. The increase of a group of trees and their recruitment to vote for the most popular class rescued in a change in the accuracy of the classification.

We used Sklearn's Random Forest Regression to compile and construct a random model for forestry regression. The model was then analyzed using the data set and the results were recorded and discussed in the results section.

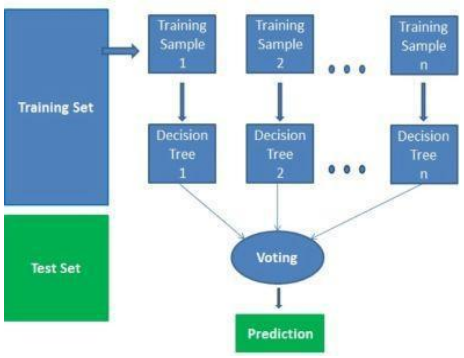

Figure 6: Decision Tree and Random Forest Regression Block Diagram

Ultimately, by using the interface built in **PyQt5** and giving this question to our platform, this question is taken from the customers. The model predicts its features, the price of the house after any estimation.

It is **Figure 7** that shows the predicted result by using split data simulation by way of the algorithm. The red line is the optimal line, and the blue line reflects the datasets in the table. The difference between the perfect match and the real match is not clear.

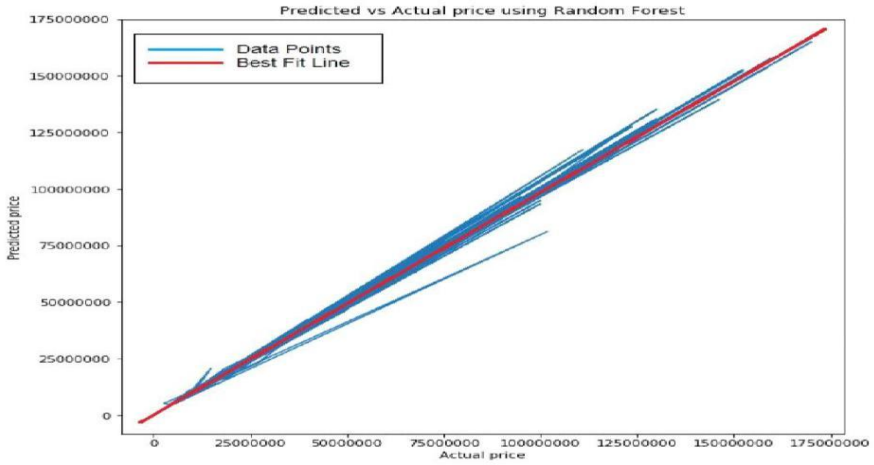

Figure 7: Predicted Price vs. Actual Price using Random Forest

3. RESULTS AND DISCUSSION

Home prices will be accurately forecast from the point of view of sellers and investors, because it is on the immovable market. Several machine learning algorithms are used in this analysis for the development of a predictive model for house price prediction. By measuring how specifically a technique can determine whether the selling price is above or below the listing price, we check the efficiency of those techniques. Using the previous data analysis, we estimate the house price. Two separate machine learning algorithms are chosen for which the algorithm generates the highest accuracy values, namely the Decision Tree and Random Forest. In the analysis of previous research, we find that the random forest efficiency is better than that of the Decision Tree model. Random Forest beats the other type of house price estimation. The decision-tree sort as well as the use of sklearn for data analysis and data mining during both experiments were carried out.

4. CONCLUSION

We evaluate our solution to the issue of house pricing, in which current algorithms are mixed. For the reliability of these techniques, other algorithms calculate, in fact, how a technique decides whether the sale price is higher or lower than the listing price. Decision Tree and Random Wood, two special algorithms for machine learning gives much advanced results. The highest precision score is given by Random Wood algorithm.

5. REFERENCES

[1] Alexander, K. L., Dauber, S. L., & Entwisle, D. R. (1996). Children in motion: School transfers and elementary school performance. The Journal of Educational Research, 90(1), 3–11.

[2] Diamond, K. E., Reagan, A J., & Bandyk, J. E. (2000). Parents' conceptions of kindergarten readiness: Relationships with race, ethnicity, and development. The Journal of Educational Research, 94(2), 93–100.

[3] Machamer, A. M., & Gruber, E. (1998). Secondary school, family, and educational risk: Comparing American Indian adolescents and their peers. The Journal of Educational Research, 91(6), 357–369.

[4] McNeal, R. B., Jr. (1998). High school extracurricular activities: Closed structures and stratifying patterns of participation. The Journal of Educational Research, 91(3), 183–191.

[5] Meisels, S. J., & Liaw, F.-R. (1993). Failure in grade: Do retained students catch up? The Journal of Educational Research, 87(2), 69–77.

[6] Rush, S., & Vitale, P. A. (1994). Analysis for determining factors that place elementary students at risk. The Journal of Educational Research, 87(6), 325–333.

[7] Smith, J. B. (1997). Effects of eighth-grade transition programs on high school retention and experiences. The Journal of Educational Research, 90(3), 144–152.

[8] Trusty, J. (2000). High educational expectations and low achievement: Stability of educational goals across adolescence. The Journal of Educational Research, 93(6), 356–365.

ISSN 2348-0831

BIOGRAPHIES

Mr. Anirban Chakraborty

MTECH, Artificial Intelligence, Lovely Professional University, Punjab, India