INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 4 - V7I4-1498)
Available online at: https://www.ijariit.com

House Price Prediction using Machine Learning

Mayur Manohar Patil
iammayur.515@gmail.com
Shrama Sadhana Bombay Trust's College of Engineering and Technology, Jalgaon, Maharashtra

Bhushan Ramkrushna Upasani
bhushanupasani3@gmail.com
Shrama Sadhana Bombay Trust's College of Engineering and Technology, Jalgaon, Maharashtra

Anup Sanjay Patil
anupp0584@gmail.com
Shrama Sadhana Bombay Trust's College of Engineering and Technology, Jalgaon, Maharashtra

Janhavi Kailas Chaudhari
janhavikchaudhari@gmail.com
Shrama Sadhana Bombay Trust's College of Engineering and Technology, Jalgaon, Maharashtra

Priti R. Sharma
pritisharma@gmail.com
Shrama Sadhana Bombay Trust's College of Engineering and Technology, Jalgaon, Maharashtra

ABSTRACT

Machine learning plays a major role from past years in image detection, spam reorganization, normal speech com- mand, product recommendation and medical diagnosis. Present machine learning algorithm helps in enhancing security alerts, ensuring public safety and improve medical enhancements. Ma- chine learning system also provides better customer service and safer automobile systems. Designing an effective machine learning model for prediction of regression and classification problem is a tedious endeavor. Significant time and expertise are needed to customize the model for a specific problem. A significant way to reduce the complicated design is by using Automated Machine Learning (AML) that can intelligently optimize the best pipeline suitable for a problem or dataset. This study utilizes machine learning algorithms as a research method that develops housing price prediction models. In that point a housing cost prediction model to support a house vendor or a real estate agent for better information based on the valuation of house is recommended.

Keywords—*Machine learning, Regression and Classification, Automated Machine Learning.*

1. INTRODUCTION

Proposed system aims to make a machine learning model which could predict the real estate house prices. Respective chapter has in view of the content focusing on scope, mo- tivation, objectives and selection of life cycle model of the projected system. It also states the actual problem definition which is to be solved using machine learning models.

A. Problem Definition

In India, there are multiple real estate classified websites where properties are listed for sell/buy/rent purposes such as 99acres, housing, common floor, magic bricks and more. How- ever, in each of these websites we can see lot of inconsistencies in terms of pricing of an apartment and there are some cases when similar apartments are priced differently and thus there is lot of in-transparency. Sometimes the consumers may feel the pricing is not justified for a particular listed apartment but there no way to confirm that either. Proper and justified prices of properties can bring in a lot of transparency and trust back to the real estate industry, which is very important as for most consumers especially in India the transaction prices are quite high and addressing this issue will help both the customers and the real estate industry in the long run. We propose to use machine learning and artificial intelligence techniques to develop an algorithm that can predict housing prices based on certain input features.

B. Scope of Machine Learning based solutions

Scope of Machine Learning (ML) is vast, and in the near future, it will deepen its reach into various fields like medical, finance, social media, facial and voice recognition, online fraud detection, and real estate. Gartner predicts that 30% of Government and large enterprise contracts will require AI-fueled solutions by 2025. ML shall also fuel areas that are highly dependent on data. For example, the real estate, machine learning can analyse data and evaluate it for better results.

2. LITERATURE SURVEY

Machine learning is a form of artificial intelligence which compose available computers with the efficiency to be trained without being veraciously programmed. Machine learning interest on the extensions of computer programs which is capable enough to modify when unprotected to new-fangled data. Machine learning algorithms are broadly classified into three divisions, namely; Supervised learning, Unsupervised learning and Reinforcement learning.

A. Supervised Learning

Supervised learning is a learning in which we teach or train the machine using data which is well labelled that means some data

is already tagged with correct answer. After that, machine is provided with new set of examples so that supervised learning algorithm analyses the training data and produces a correct outcome from labelled data.

B. Unsupervised Learning

Unsupervised learning is the training of machine using information that is neither classified nor labelled and allowing the algorithm to act on that information without guidance. Here the task of machine is to group unsorted information according to similarities, patterns and differences without any prior training of data. Unlike, supervised learning, no teacher is provided that means no training will be given to the machine. Therefore, machine is restricted to find the hidden structure in unlabelled data by our-self.

C. Reinforcement Learning

Reinforcement learning is an area of Machine Learning. Reinforcement. It is about taking suitable action to maximize reward in a particular situation. It is employed by various software and machines to find the best possible behaviour or part it should take in a specific situation.

Reinforcement learning differs from the supervised learning in a way that in supervised learning the training data has the answer key with it so the model is trained with the correct answer itself whereas in reinforcement learning, there is no answer but the reinforcement agent decides what to do to perform the given task. In the absence of training dataset, it is bound to learn from its experience.

3. PROPOSED SOLUTION

As a proposed solution to the problem stated above and based on the literature survey the linear regression model comes in picture to be a solution.

Linear Regression is a machine learning algorithm based on supervised learning. Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output).

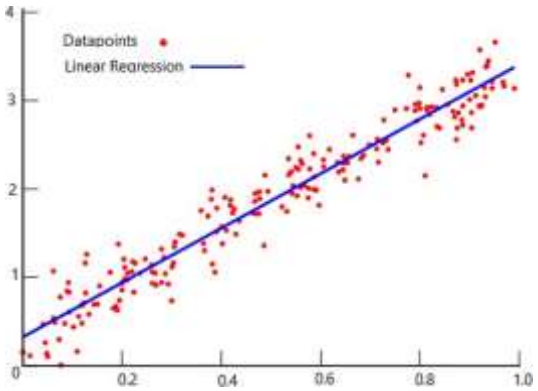

Fig. 1. Linear Regression Graph

A. Equations of Linear Regression

The formula for linear regression equation is given by:

$$y = a + bx \qquad (1)$$

now, ‘a’ and ‘b’ can be computed using following formulas:

$$b = \frac{n\sum xy - [(\sum x)(\sum y)]}{n\sum x2 - (\sum x)2} \qquad (2)$$

$$a = \frac{\sum y - b\ (\sum x)}{n} \qquad (3)$$

where,

y = Dependant variable (Price of the house)
x = Independent variable
b = Slope of the line
a = Y-intercept of the line

Finally, substituting the value of ‘b’ from equation(2) and value of ‘a’ from equation(3) in equation(1) will return the value of dependant variable i.e ‘y’.

B. Data Flow Diagram

Fig. 2. Level 2 Data Flow Diagram

C. Algorithm for the model building

1) Problem Definition.
2) Hypothesis Generation.
3) Data Extraction/collection.
4) Data Exploration and Transformation.
5) Predictive Modeling.
6) Model Deployment.

D. Hypothesis Generation

- Location where the house is.
- Does the property has a parking facility?
- What type of property is it?
- The type of streets leading to the house.
- The total sqft area of the house.
- No of bedrooms in the house.
- No of bathrooms in the house.
- No of rooms excluding the bedrooms and bathrooms in the house.
- The quality score given to the rooms.
- The quality score given specifically to the bathrooms.
- The quality score given specifically to the bedrooms.
- The quality score of the property.

E. Data Exploration

1) Reading the data.
2) Variable Identification.
3) Univariate Analysis. (Bar-plots, Histograms)
4) Bi-variate Analysis. (Scatter-plots)
5) Missing Value Treatment. (Mean, Mod, Median)

6) Outlier Treatment.
7) Variable Transformation.

F. Algorithm for working of the software

- **Step 1:** START
- **Step 2:** User Login through Google or Github.
- **Step 3:** Authenticating and conforming credentials through API.
- **Step 4:** User inputs required Information.
- **Step 5:** The Information is then sent to the Backend.
- **Step 6:** The Information obtained is processed by trained Machine Learning Model.
- **Step 7:** The Model prepares an answer as output.
- **Step 8:** Output is displayed to the User.
- **Step 9:** STOP decision.

More factors like recession that affect the house prices shall be added. In-depth details of every property will be added to provide ample details of a desired estate. This will help the system to run on a larger level.

4. RESULTS

After training the dataset on three different machine learning model the outcome that has been extracted is as follows, the linear regression model performed the best with the score of approximately 91%, followed by the lasso regression with an approximate score of 91%, lastly the decision tree score almost 90% score when trained on a dataset.

As the final decision to choose over these machine learning models the optimal choice had to be the linear regression model and hence, that is the reason of using the same in the proposed solution.

Table I: Different Machine Learning Models With Their Individual Scores

Model	Scores in %	Best Parameters
Linear Regression	90.686856	*{*'normalize': True*}*
Lasso Regression	90.684626	*{*'alpha': 2, 'selection': 'cyclic'*}*
Decision Tree	89.400316	*{*'criterion': 'mse', 'splitter': 'random'*}*

5. CONCLUSION

In today's real estate world, it has become tough to store such huge data and extract them for one's own requirement. Also, the extracted data should be useful. The system makes optimal use of the Linear Regression Algorithm. The system makes use of such data in the most efficient way. The linear regression algorithm helps to fulfill customers by increasing the accuracy of estate choice and reducing the risk of investing in an estate. A lot's of features that could be added to make the system more widely acceptable. One of the major future scopes is adding estate database of more cities which will provide the user to explore more estates and reach an accurate

6. REFERENCES

[1] Edna Dias Canedo and Bruno Cordeiro Mendes. *Regression Machine Learning*. Department of Computer Science, University of Bras´ılia (UnB), P.O. Box 4466, Bras´ılia 70910-900, Brazil.

[2] Atharva chogle, priyanka khaire, Akshata gaud and Jinal Jain4 *House Price Forecasting.* [*House price forecasting using machine learning*]. Bachelor of Engineering, Dept. of Computer Engineering, RGIT, Mum- bai, Maharastra, India

[3] Suraya Masrom, Thuraiya Mohd, Nur Syafiqah Jamil and Norhayati Baharun *Automated Machine Learning*. IEEE 2020.

