Mini Project Report on

Housing Price Prediction using ML

Submitted in Partial Fulfillment of the requirement for the Award of the Degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

Submitted by:

Student Name-   University Roll No.-
KRITI AGRAWAL   2018893

Under the Mentorship of
Dr. Vikas Tripathi

Department of Computer Science and Engineering
Graphic Era (Deemed to be University)
Dehradun, Uttarakhand
January 2023

DECLARATION

I hereby certify that the work which is being presented in the project report entitled **"Housing Price Prediction using ML"** in partial fulfillment of the requirements for the award of the degree of **Bachelor Of Technology in Computer Science and Engineering** of the **Graphic Era Deemed to be University, Dehradun** shall be carried out by the under the mentorship of **Dr. Vikas Tripathi**, Professor, Department of Computer Science and Engineering, Graphic Era(Deemed to be University), Dehradun.

Kriti Agrawal   2018893   **kritiagrawal**

Table of Contents

Chapter 1

INTRODUCTION

1.1 GOAL – The main goal of the project is to predict the price of Houses using different Regression Techniques taking into account of various features of the dataset on which housing prices depend.

1.2 MOTIVATION – Being extremely interested in everything having a relation with the Machine Learning and Data Science, the independent project was a great occasion to give me the time to learn and confirm my interest for this field. The fact that we can estimate, predict and give machines the ability to learn on their own is powerful and limitless in terms of applications. Machine learning can be used in everything from finance to medicine. So I decided to do a project on machine learning

1.3 DOMAIN INTRODUCTION – Housing is critical to family well-being, and homeownership is perceived as integral to both the stability of family life and the creation of wealth. Housing prices reflect the quality of life in urban environments and it is a key element to any sustainable, productive, and resilient city.

Many houses are sold every day. There are a few questions every buyer asks: What is the actual price this house is worth? Am I paying a fair price? This project proposes a machine learning model for predicting house prices based on data about houses (size, the year it was built in, etc.).

1.4 SCOPE OF THE PROJECT – India is a dense country with a lot of population the demand for land and property is gone sky rocketing in recent years this proposed model serves the purpose of finding the essential values and prices of the properties. This model makes it easier to maintain transparency between customers and make comparisons. If a customer finds a higher price for a house on a particular website than the price predicted by the model, he can reject the house.

This project develops and evaluates the performance and predictive power of models trained and tested using data collected from homes. This study uses machine learning algorithms as a research method. Home price forecasts help determine home sales prices and determine the right time to buy a home. There are several factors that affect the price of a home, including physical conditions, concept, and location.

1.5 SYSTEM REQUIREMENTS –

1.5.1 Operating System – Windows 10 / Ubuntu / Mac OS

1.5.2 System Architecture – 64-bit System Required

1.5.3 Software Required – Anaconda (Jupyter Notebook)

1.5.4 Additional Software Required – Web Browser

Chapter 2

LITERATURE SURVEY

[1] Real estate price prediction with Regression and Classification

In this paper, we use logical factors covering many parts of private homes to predict house costs. House costs are projected with different fallback methods. Lasso, Ridge, SVM Fallback, and Random Forest are just a few examples. In this article, the best performing model for a relapse issue is SVR (Gaussian bit), which has an RMSE of 0.5261. In any case, representation for SVR was trouble because of its very high dimensions. As per examination, residing regional rectangular feet, the material of rooftop, and locality have best measurable importance in seeing a family deal cost.

[2] House Price Prediction Using Machine Learning and Neural Networks

In the paper means to make assessments in light of each essential boundary that is considered during the decision of the cost. This model contains various fallback strategies and the results are still up in the air with a single strategy. Instead, it's a weighted average of the various procedures for the most reliable results. The outcome demonstrated is that this method yields the smallest mistake and most extreme precision than separate calculations useful.

[3] Vision-based real estate price estimation

The subject of automatic calculation of real estate market prices has received greater attention due to the development of online real estate database companies such as Zillow, Trulia and Redfin. many real estate websites use proprietary methodology to offer such estimations. These predictions are often close to the actual selling price, but can be downright wrong in other situations. The inside and outside look of a house are important aspects that are not taken into account while creating computerised value estimations. This study examines the impact of a home's aesthetic quality on its market value. We develop a method to predict the grade level of real estate photos using a deep convolutional neural network on a large dataset of real estate interior and exterior photos. We create a new framework for automatic value evaluation that incorporate above images as well as property parameter(such as size, proposed price, and number of bedrooms). Finally, we show that the proposed price estimation method outperforms Zillow's estimates by applying it to a new dataset of property photos and information.

[4] House Price Prediction Using Machine Learning Algorithms

The demand for renting and owning homes has increased as a result of increased urbanization. As a result, finding better techniques for calculating real estate prices that truly represent market prices has become a popular topic. This research focuses on applying machine learning techniques such as simple linear regression (SLR), multiple linear regression (MLR), and neural networks to correctly determine house prices (NN). The algorithm with the least Mean Square Error (MSE) is picked as the best for predicting the price of a property. This helps both sellers and buyers determine the best price for a home.

[5] Housing Price Prediction Based on CNN

A novel prediction model based on CNN is proposed for prediction of housing price as well as the process of feature selection. Compared with other traditional methods are work can get a better performance through experiments using real data of property transaction.

Author: An sheng Chen, Zhendong Shang, Yong Piao

Keywords: housing price prediction; feature selection; convolutional neural network

Algorithm: Convolutional neural network

Year of publishing :2019

[6] An SVR based forecasting approach for real estate price prediction

The help support vector machine (SVM) has been applied successfully to order, bunch, and gauge. The support Vector Relapse (SVR) method is proposed in this paper to estimate land costs in China. The point of this newspaper was to analyze the attainability of SVR in land cost expectation. Test results were determined based on total mean error (MAE), total mean error (MAPE), and root mean squared error (RMSE), making the SVR-based method an effective tool for measuring land costs.

Chapter 3

METHODOLOGY

3.1 LIBRARIES USED –

1. NumPy
2. Pandas
3. Matplotlib
4. Seaborn
5. Plotly
6. Sklearn

3.2 DATA –

A key component of any machine learning task that clearly requires special attention is the data. In fact, the results are heavily influenced by the data, based on where it is found, how it is formatted, whether it is consistent, whether it has outliers, etc. Many questions need to be answered in this step to ensure that the learning algorithm is efficient and accurate.

	price	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterheating	airconditioning	parking	prefarea	furnishingstatus
0	13300000	7420	4	2	3	yes	no	no	no	yes	2	yes	furnished
1	12250000	8960	4	4	4	yes	no	no	no	yes	3	no	furnished
2	12250000	9960	3	2	2	yes	no	yes	no	no	2	yes	semi-furnished
3	12215000	7500	4	2	2	yes	no	yes	no	yes	3	yes	furnished
4	11410000	7420	4	1	2	yes	yes	yes	no	yes	2	no	furnished

Table 3.1

This is how the first 5 rows of our dataset looks like.

Our Dataset has 545 Rows and 13 Columns in total.

3.3 DESCRIPTION OF ALL COLUMNS –

Table 3.2

Column	Datatype	Description
Price	Integer	Sale Price of the House
Area	Integer	Area of the house
Bedrooms	Integer	Number of Bedrooms
Bathrooms	Integer	Number of Bathrooms
Stories	Integer	Number of Floors
Main Road	String	Is house on Main Road
Guestrooms	String	Does it have a Guestroom
Basement	String	Does it have a Basement
Hotwater Heating	String	Is it Installed
Airconditioning	String	Is AC Installed
Parking	Integer	Number of Parking
Prefarea	String	Is it buyer's preferred area
Furnishing Status	String	Furnishing status of house

3.4 NULL VALUES IN EACH COLUMN –

```
price                 0
area                  0
bedrooms              0
bathrooms             0
stories               0
mainroad              0
guestroom             0
basement              0
hotwaterheating       0
airconditioning       0
parking               0
prefarea              0
furnishingstatus      0
dtype: int64
```

As we can observe, there are no NULL values in any of the Columns so no Data Cleaning is required.

But there are string values in our dataset and the Regression Algorithms and most in-built methods work only on numerical values. So, we have to encode the string values to some

numerical values. We do this by encoding 'yes' as 1 similarly 'no' as '0' in all columns with string datatype except furnishing status.

In Furnishing Status column, we apply One Hot Encoding.

3.5 Checking the Co-relation of Price with different columns-
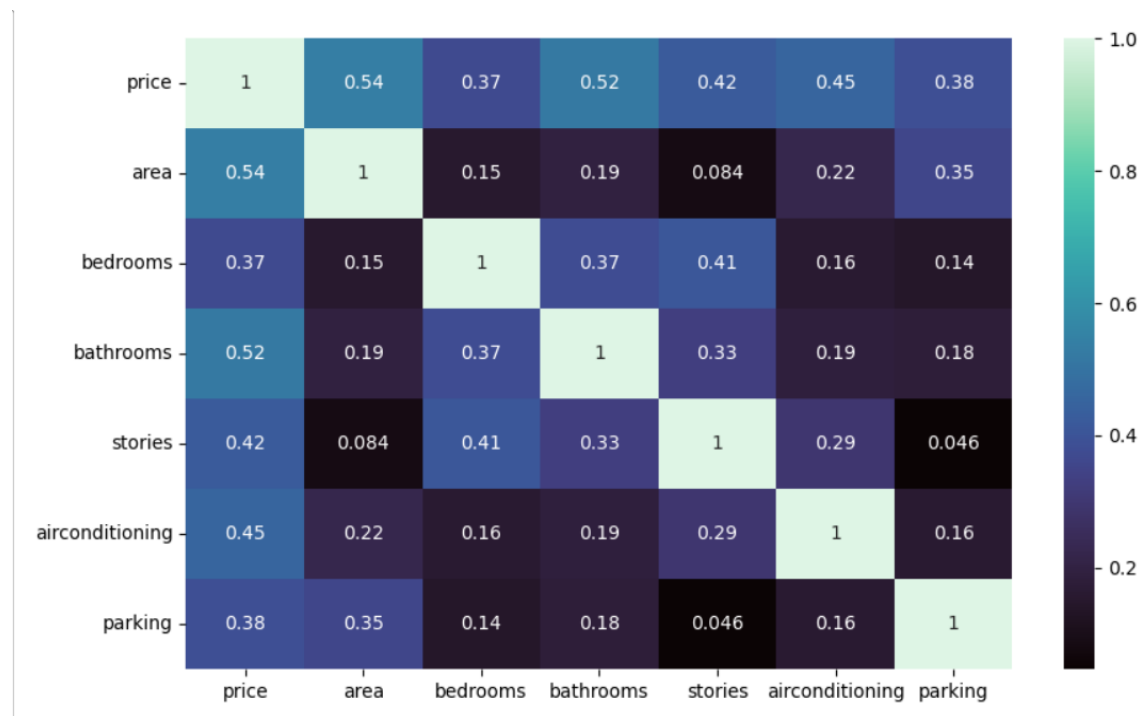
There are more columns but we are mostly interested with columns which are at-least 35% correlated with the Price Column.

3.6 PREPARING DATA AND PREDICTING THE HOUSE PRICES –

3.6.1 Removing the outliners of the data– We removed the data where number of bathrooms were 4 and number of bedrooms were 6 as they were very less in number and could result in wrong prediction of the prices.

3.6.2 Splitting the Dataset– We first separate our target variable ('price') from the rest of our variables, then we split the data in the ratio of 80:20 (80% for training our model and 20% for evaluating our model) keeping the ratio of 'airconditioning' variable same (Stratified Split) as it is highly correlated to our target variable.

3.6.3 Using Regression Techniques to predict the value of House Prices –

Regression Technique – Regression Analysis is a form of predictive modelling technique which investigates the relationship between dependent variables (X) and an independent variable (y).

In the Project I have used **Multiple Linear Regression, Gradient Boosting Regressor and Random Forest Regressor** to predict values and evaluate our model.

Checking the goodness of our Prediction–

R-Squared Method – It is a statistical measure of how close the original data points are to the fitted regression line. It is also called as coefficient of multiple determination.

Chapter 4

RESULT AND DISCUSSION

Final R-Squared Scores of our Models – In the project, I have resampled the
data 5 times and recorded the R-Square values of all three Regression algorithms
which are as follows –

```
In [34]: scores_linear

Out[34]: [0.6194255914743709,
          0.7586729468155239,
          0.6517894278649636,
          0.693102186241648,
          0.6371033513698846]
```

```
In [35]: scores_gradientboost

Out[35]: [0.612180789717477,
          0.6872618869550211,
          0.648350297818416,
          0.6375132581571249,
          0.6369906350166048]
```

```
In [36]: scores_rf

Out[36]: [0.6023294147199476,
          0.7010488487087518,
          0.5970773371686655,
          0.6296030928099962,
          0.6478180345601888]
```

From this, we can conclude that **Multiple Linear Regression** is the best Regression
Algorithm for our dataset as it has produced the maximum **R2 score of 75%.**

Chapter 5

Conclusion and Future Work

In this project, we built several regression models to predict the price of some house given some of the house features. We evaluated and compared each model to determine the best performing model. We also examined how some models rank features by importance. In this, we followed the data science process starting with getting the data, then processing the data, followed by exploration of the data and building the models, then finally evaluating the results. We tried various machine learning algorithms to get the best model. When compared to all other algorithms, Multiple Linear Regression is the best Regression Algorithm. Future direction of this may consider incorporating additional property transaction data from a larger geographical location with more features, or analyzing other property types beyond housing development. In the future, the model will be enhanced on the accuracy rate of the forecasted price. Future work will focus on the data preprocessing by including sentiment data before the testing and training experiments.

References

[1] A. Varma et al. [House Price Prediction Using Machine Learning And Neural Networks],2018 Second International Conference on Inventive Communication and Computational Technologies、S. 1936 – 1939、1936.

[2] Yu, H., and J. Wu.[Real estate price vaticination with retrogression and bracket CS 229 Afterlife 2016 Project Final Report 1 – 5.] - (2016).

[3] N.N. Ghosalkar and S.N. Dhage, "Real estate value prediction using Linear Retrogression, "2018 Fourth International Conference on Computing Communication Control and Robotization (ICCUBEA), Pune, India, 2018, pp. 1-5.

[4] Banerjee, Debanjan, and Suchibrota Dutta. "Predicting the case price direction using machine literacy ways."2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI). IEEE, 2017.

[5] "House price prediction using Machine Learning and Neural networks", Computer Engineering Department from KJ Somaiya College of Engineering, Vidyavihar(Mumbai) .

[6] "Predicting public housing prices using delayed Neural Networks" - Lipo Wang, Fung Foong Chan, Yaoli Wang, and Qing Chang