NAME OF THE PROJECT

Housing: Price Prediction

Submitted To:

Ms. Khushboo Garg

SME

Submitted By:

MONIKA LATHAR

Intern-34

ACKNOWLEDGMENT

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to **Ms KHUSBOO GARG** for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

I would like to express my gratitude towards my parents & member of **Flip Robo Technology** for their kind co-operation and encouragement which help me in completion of this project.

I would like to express my special gratitude and thanks to industry persons for giving me such attention and time.

My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.

I have studied lot of research papers and references for completing this project of Housing Price Prediction.

List of all the research papers as given below:

1) Forecasting Australian Real House Price Index: A Comparison Study of Machine Learning and Time Series Methods.

https://www.researchgate.net/publication/334388950_Forecasting_Australian_Real_House_Price_Index_A_Comparison_Study_of_Machine_Learning_and_Time_Series_Methods

2) House Price Prediction Using Regression Techniques: A Comparative Study

https://ieeexplore.ieee.org/document/8882834

3) Housing Price Prediction via Improved Machine Learning Techniques

https://www.sciencedirect.com/science/article/pii/S1877050920316318

4) House Price Prediction Using Machine Learning and Neural Networks

https://ieeexplore.ieee.org/document/8473231

INTRODUCTION:

In this report, we propose our system “House price prediction”. House is one of human life's most essential needs, along with other fundamental needs such as food, water, and much more. Demand for houses grew rapidly over the years as people's living standards improved. House price prediction can be done using multiple prediction models (Machine Learning Model) such as support vector regression, artificial neural network, etc. There are many benefits that home buyers, property investors, and house builders can reap from the house-price model. This model will provide a lot of information and knowledge to home buyers, property investors, and house builders, such as the valuation of house prices in the present market, which will help them determine house prices. The target feature in this proposed model is the price of the real estate property and the independent features are: MSZoning, MSSubClass: Identifies the type of dwelling involved in the sale,LotFrontage: Linear feet of street connected to property,LotArea: Lot size in square feet,Street: Type of road access to property,no. of bedrooms, Exterior covering on house, Exterior covering on house carpet area, the floor, car parking, and Lot configuration etc. The whole implementation is done using the python programming language. In recent years, due to the growing trend towards Big Data, machine learning has become a vital prediction approach because it can predict house prices more accurately based on their attributes, regardless of the data from previous years.

Data mining is now commonly used in the real estate market. Real Estate is a clear industry in our ecosystem. The ability to extract data to extract relevant information from raw data makes it very useful to predict house prices, important housing features, and much more. Housing prices continue to change from day to day and are sometimes raised rather than based on calculations. Research has shown that fluctuations in housing prices often affect homeowners and the housing market. Literature research is done to analyze the relevant factors and the most effective models for predicting housing prices. The findings of this analysis confirmed the use of Artificial Neural Network, Support Vector Regression, and Linear Regression as the most efficient models compared to others. In addition, our findings also suggest that spatial and real estate agents are key factors in predicting house prices. This study will be of great benefit, especially to housing developers and researchers, to find the most important criteria for determining housing prices and identify the best machine learning model used to conduct research in this field.

PROBLEM STATEMENT

As We know that A US-based housing company named Surprise Housing has decided to enter the Australian market. The company uses data analytics to purchase houses at a price below their actual values and flip them at a higher price. For the same purpose, the company has collected a data set from the sale of houses in Australia. The data is provided in the CSV file. The general and standardized real estate characteristics are often listed separately from the asking price and general description. Because these characteristics are separately listed in a structured way, they can be easily compared across the whole range of potential houses. Because every house also has its unique characteristics, such as a particular view or type of sink, house sellers can provide a summary of all the important features of the house in the description. All given real estate features can be considered by the potential buyers, but it is nearly impossible to provide an automated comparison of all variables due to the large diversity. This is also true in the other direction: house sellers have to estimate the value based on its features in comparison to the current market price of similar houses. The diversity of features makes it challenging to estimate an adequate market price. Apart from providing a summary of the important features of the house, the house description is also a means of raising curiosity in the reader, or in other words persuading the person. Housing prices are an important reflection of the economy, and housing price ranges are of great interest to both buyers and sellers. In this project, house prices will be predicted given explanatory variables that cover many aspects of residential houses. The goal of this project is to create a regression model that can accurately estimate the price of the house given the features

LITERATURE REVIEW

i. Survey Existing System: Trends in housing prices indicate the current economic situation and also are a concern to the buyers and sellers. Many factors have an impact on house prices, such as the number of bedrooms and bathrooms. House price depends upon its location as well. A house with great accessibility to highways, schools, malls, and employment opportunities, would have a greater price as compared to a house with no such accessibility. Predicting house prices manually is a difficult task and generally not very accurate, hence there are many systems developed for house price prediction. Sifei Lu, Zengxiang Li, Zheng Qin, Xulei Yang, and Rick Siow Mong Goh had proposed an advanced house prediction system using linear regression. This system aimed to make a model that can

give us a good house price prediction based on other variables. They used the Linear Regression for Ames dataset and hence it gave good accuracy.

Limitation Existing system or research gap

In the existing literature, a limited amount of work has been focused on the housing price prediction model, particularly, to solve the problem using machine learning approaches. A few identified papers were reported above. In addition, most of the past research considered the housing market problem as a classification problem to develop a classification model instead of a regression model. Therefore, the objective of the study is to predict the housing price valuation using machine learning techniques and considering competitive regression models. An improved ML-based algorithm is proposed, which includes the predicted target price binning variable as features in the model and improves the model accuracy significantly. More precisely, the model accuracy is increased by 10 percent compared to other contemporary machine learning techniques.

METHODOLOGY:

The proposed system face recognition-based attendance system can be divided into three main modules. The modules and their functions are defined as follows.

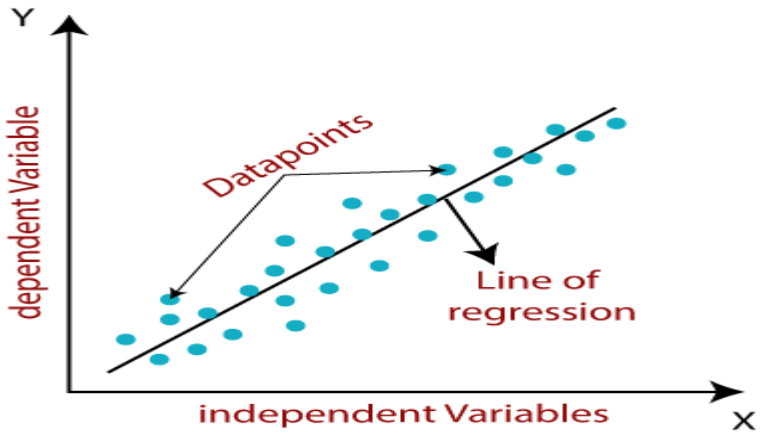
a)Data Collection: For doing machine learning projects we need a vast amount of data which we have provided by the Flip Robo Technology in the form of CSV Files.The data consists of several factors in a house of a given locality such as the 'MSSubClass', 'MSZoning', 'LotFront age', 'LotArea', 'Street','Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig','LandSlope', ' Neighborhood', 'Condition1', 'Condition2', 'BldgType', etc. The dataset that we will be using will be batch i.e., static data and not dynamic.

b. Train the Dataset and Feature Extraction: The dataset will further be trained and processed using Linear Regression Algorithm which will predict the house price accurately. In applied machine learning, we seek a model that learns the relationship between the input and output variables using the training dataset. The hope and goal is that we learn a relationship that generalizes to new examples beyond the training dataset.

Linear regression is one of the easiest and most popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as **sales, salary, age, product price,** etc.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.

The linear regression model provides a sloped straight line representing the relationship between the variables. Consider the below image:



The algorithm used will be a regression since regression helps to get a value by analyzing a dataset, in this condition the value is the price of the house.

c. Displaying the output: The project uses a web application to work with the users and give them accurate price predictions. The user will input features accordingly which they want on the frontend and the data will be processed . “Housing Price in Australia” is a dataset containing Data contains 1460 entries each having 81 variables representing housing prices as shown below:

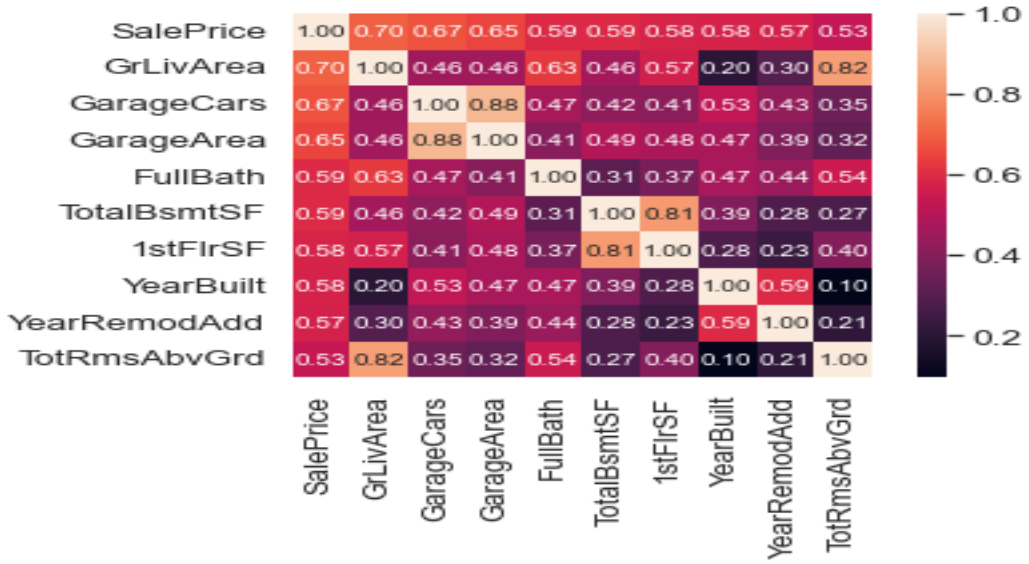
‘Id', 'MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street','Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig','LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgType', 'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd','RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType','MasVnrArea', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual', 'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1', 'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heating','HeatingQC', 'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF','LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath','HalfBath', 'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual', 'TotRmsAbvGrd', 'Functional', 'Fireplaces', 'FireplaceQu', 'GarageType', 'GarageYrBlt', 'GarageFinish', 'Ga

rageCars', 'GarageArea', 'GarageQual','GarageCond', 'PavedDrive', 'WoodDeckSF', 'OpenPorchSF', 'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'PoolArea', 'PoolQC', 'Fence', 'MiscFeatu, ' MiscVal', 'MoSold', 'YrSold', 'SaleType','SaleCondition', 'SalePrice'],

These variables, which served as features of the dataset, were then used to predict the average price per square meter of each house. The next step was to investigate missing data. Variables with more than 50% missing data would be removed from the dataset. The variable

PoolQC 99.400685
MiscFeature 96.232877
Alley 93.407534
Fence 79.708904
FireplaceQu 47.174658
LotFrontage 18.321918
GarageYrBlt 5.479452
GarageFinish 5.479452
GarageType 5.479452
GarageQual 5.479452
GarageCond 5.479452
BsmtExposure 2.654110
BsmtFinType2 2.654110
BsmtQual 2.568493
BsmtCond 2.568493

were removed because of missing data. Any observation which had missing values were also removed from the dataset and Treaming missing values by imputing for columns with missing values less than or equal to 40%.By Visualising the variables with missing values all the missing values has been treated successfully. Below are a few feature engineering processes which were done to cleanse the dataset on the backend using Linear Regression and the predicted output will be shown to the respective user.



The vaiables are highly correlated which are in the lighter shades.GarageCars and GarageArea are having optimal values in between 0.88 to 1.00.

ALGORITHM:

1) Importing the required packages into our python environment

2) Importing the house price data

3) Data Visualization of the house price data

4) Modelling the data using the algorithms

5) Take inputs from the user and display the analysed result

6. RESULT AND OUTPUT:

The result of our project has an accuracy of 90.75, where the user has to enter the LotFrontage', 'LotArea','MasVnrArea', 'BsmtFinSF1', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', '1stFlrSF', '2ndFlrSF', 'LowQualFinSF', 'GrLivArea', 'BsmtFullBath','BsmtHalfBath', 'FullBath', 'HalfBath', 'BedroomAbvGr', 'KitchenAbvGr', 'TotRmsAbvGrd', 'Fireplaces','GarageCars', 'GarageArea','WoodDeckSF', 'OpenPorchSF', 'EnclosedPorch', '3SsnPorch','ScreenPorch', 'PoolArea', 'MiscVal' to get the desired output.

In this project work, I have taken you through the Ridge and Lasso Regression in Machine Learning and how to implement it by using the Python Programming Language. As We have

showed from the above achieved results in case of Regression for both Ridge and Lasso regression got a very satisfied results, as required in the project work.

Ridge regression is a regularized version of linear regression. This forces the training algorithm not only to fit the data but also to keep the model weights as small as possible.

Ridge : Train :91.72 Test :75.84

Least absolute shrinkage and selection operator regression (usually just called lasso regression) is another regularized version of linear regression: just like peak regression, it adds a regularization term to the cost function.

Lasso : Train :90.9 Test :74.0

Top 5 most significant variables in Ridge are:

('SaleCondition_Partial', 0.115)

('SaleCondition_Others', 0.95)

('SaleCondition_Normal', 0.091)

('GarageFinish_Unf', 0.070)

('GarageFinish_RFn', 0.064)

Top 5 most significant variables in Lasso are:

('SaleCondition_Partial', 0.168)

('SaleCondition_Others', 0.101)

('SaleCondition_Normal', 0.094)

('GarageFinish_Unf', 0.077)

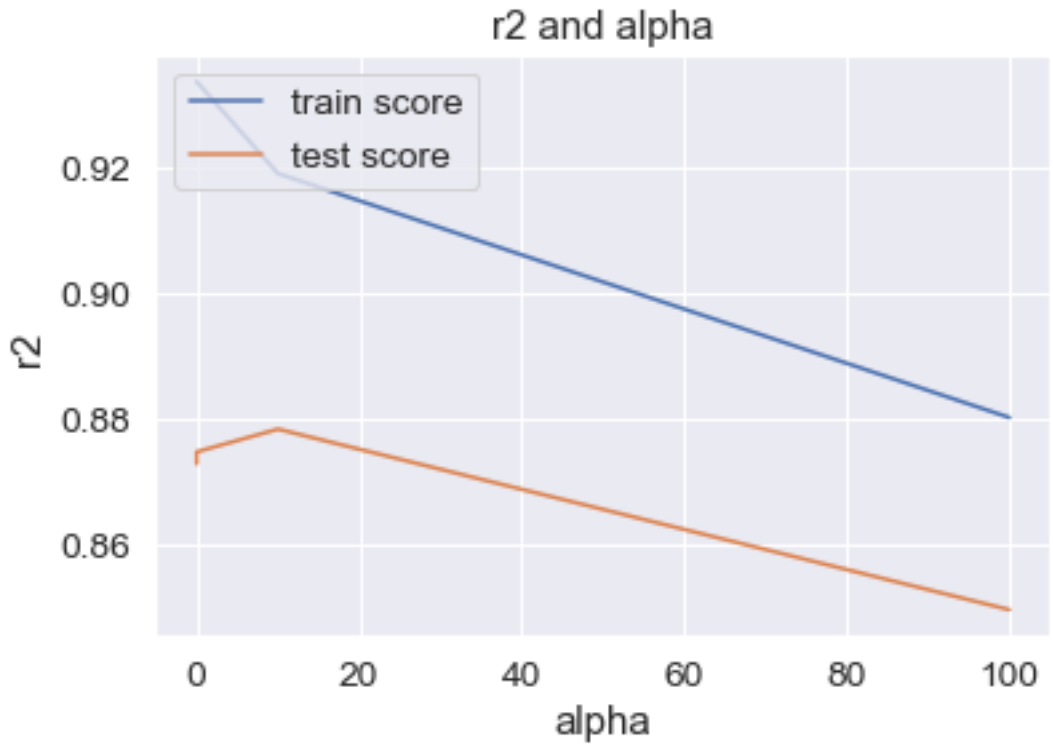
('GarageFinish_RFn', 0.066)

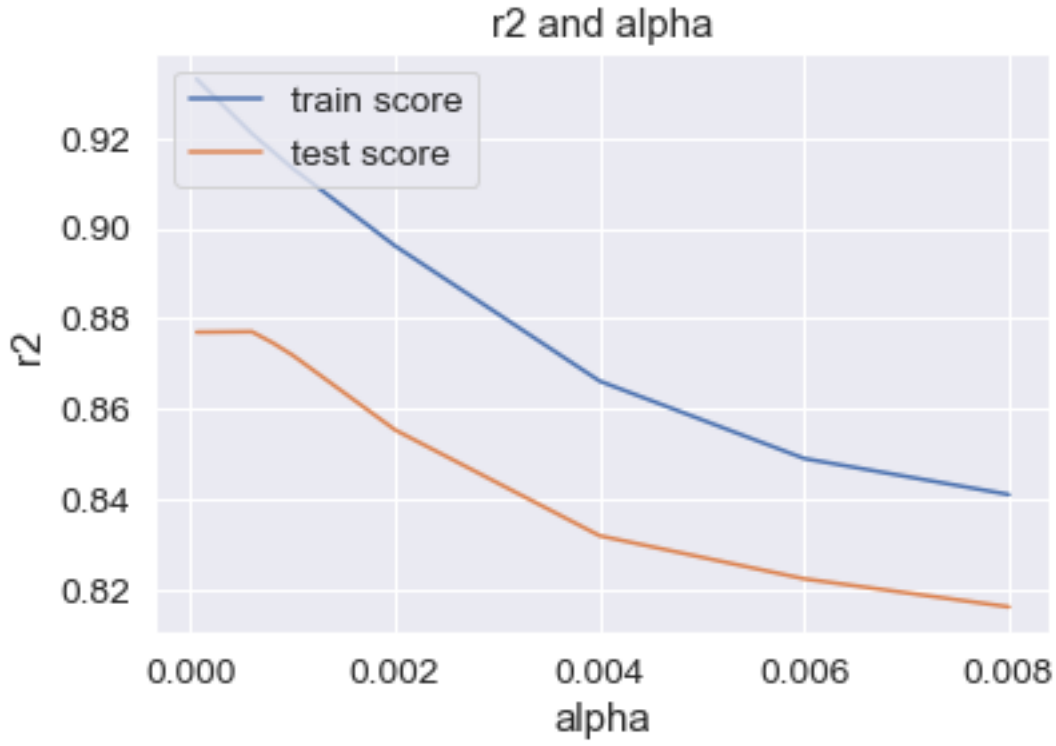
These Varaiables are directly proportional to each other.

Optimal Value of lamda for ridge : 10

Optimal Value of lamda for Lasso : 0.001

Because of Feature selection as well we can choose Lasso regression in this case.





Conclusion:

This paper examined and analysed the current research on the significant attributes of house prices and analysed the data mining techniques used to predict house prices. The accurate prediction model would allow investors or house buyers to determine the realistic price of a house as well as the house developers to decide the affordable house price. This paper discusses an overview of the concept of machine learning and its various applications. Taking the sample dataset for houses, and considering its various attributes, the prices for houses have been predicted by employing machine learning methods of regression for predicting the price of the estate using prior data.

References:

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

[2] Mueller, J. P., & Massaron, L. (2016). Machine learning for dummies. John Wiley & sons [3] G. Gao et al., "Location-Centered House Price Prediction: A Multi-Task Learning Approach," pp. 1–14, 2019, [Online]. Available: http://arxiv.org/abs/1901.01774

[3] Y. Zhou, "Housing Sale Price Prediction Using Machine Learning Algorithms," 2020.

[4] T. Mohd, S. Masrom, and N. Johari, "Machine learning housing price prediction in petaling Jaya, Selangor, Malaysia," Int. J.Recent Technol. Eng., vol. 8, no. 2 Special Issue 11, pp. 542– 546, 2019, Doi: 10.35940/ijrte. B1084.0982S1119.

[5] A. Varma, A. Sarma, S. Doshi, and R. Nair, "House Price Prediction Using Machine Learning and Neural Networks," Proc. Int.Conf. Inven. Commun. Comput. Technol. ICICCT 2018, pp. 1936–1939, 2018, Doi: 10.1109/ICICCT.2018.8473231.

[6] House Price Index. Federal Housing Finance Agency. https://www.fhfa.gov/ (accessed September 1, 2019). [2] Fan C, Cui Z, Zhong X. House Prices Prediction with Machine Learning Algorithms. Proceedings of the 2018 10th International Conference on Machine Learning and Computing - ICMLC 2018. doi:10.1145/3195106.3195133. [3] Phan TD. Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia. 2018 International Conference on Machine Learning and Data Engineering (ICMLDE) 2018. doi:10.1109/icmlde.2018.00017. [4] Mu J, Wu F, Zhang A. Housing Value

Forecasting Based on Machine Learning Methods. Abstract and Applied Analysis 2014;2014:1–7. doi:10.1155/2014/648047.