A Project Report

on

House Price Prediction

Submitted in partial fulfillment of the

requirement for the award of the degree of

Master of Computer Applications

Under The Supervision of
Tarun Kumar
Assistant Professor
Department of Computer Science and Engineering

Submitted By

19SCSE1010519-TANISHQ PUNDIR

19SCSE1010284 – RITIK KUMAR

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING /
DEPARTMENT OF COMPUTERAPPLICATION
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA
DECEMBER-2021

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"House Price Prediction"** in partial fulfillment of the requirements for the award of the **BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING** submitted in the **School of Computing Science and Engineering of Galgotias University**, Greater Noida, is an original work carried out during the **July-2021** to **December-2021**, under the supervision of **MR. Tarun Kumar**, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

19SCSE1010519-Tanishq Pundir

19SCSE1010284-Rititk Kumar

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Supervisor Name

Designation

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination **19SCSE1010519 Tanishq Pundir 19SCSE1010284 Ritik Kumar** has been held on _____ and his/her work is recommended for the award of **BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING.**

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date: November, 2013

Place: Greater Noida

Abstract

Machine Learning is the study of Computer algorithm that can work on naturally through experience and by the utilization of data. It is viewed as a part of Artificial Intelligence. Machine Learning algorithm builds a model dependent on sample data, known as "training data", to settle on expectations or choices without being unequivocally customized to do as such. Machine Learning algorithms can be classified into three types supervised, unsupervised and reinforcement Learning algorithms. Machine Learning algorithms are utilized in a wide variety of applications, for example, in medication, email filtering, speech recognition, and Computer vision, where it is difficult or impossible to make algorithms to play out the required task. Traditional programming contrasts essentially from machine learning. In traditional programming, a developer code every one of the guidelines in interview with a specialist in the business for which programming is being created. Each standard depends on an intelligent establishment; the machine will execute a yield following the legitimate assertion. At the point when the framework develops intricate, more principles should be composed. It can immediately become impractical to keep up with.

House price prediction is commonly used to estimate the price of the house according to the given condition. In our project we are going to predict the price of the given house according to it’s the input given by the user. We will write our program in python to implement machine learning as it is the one of the best languages to use machine learning. To read and load the data into the program we will be using pandas and for implementing the learning we will be using sklearn library. We will also write the research paper for our project and will also publish it in Scopus journal or present our paper in Scopus conference.

Table of Contents

Chapter1

Introduction

House is one of human existence's most fundamental requirements, alongside other essential necessities like food, water, and considerably more. Interest for houses became quickly throughout the years as individuals' expectations for everyday comforts improved. While there are people who make their home as a speculation and property, yet the vast majority all throughout the planet are purchasing a house as their shelter. Our project idea is to make model to predict the price of their house which them to predict the best house according to their requirement.

An increase in house demand occurs each year, indirectly causing house price increases every year. The problem arises when there are numerous variables such as location and property demand that may influence the house price, thus most stakeholders including buyers and developers, house builders and the real estate industry would like to know the exact attributes or the accurate factors influencing the house price to help investors make decisions and help house builders set the house price.

This project brings together the latest research on prediction markets to further their utilization by economic forecasters. Thus, there is a need to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. This project efficiently analyses previous market trends and price ranges, to predict future prices. This topic brings together the latest research on prediction markets to further their utilization by economic forecasters. It provides a description of prediction markets, and also the current markets which are useful in understanding the market which helps in making useful predictions. There is a need to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. This project uses data mining algorithm to predict prices by analysing current house prices, thereby forecasting the future prices according to the user requirements.

After examining data, we find that the data quality is a key factor to predict the house prices. Data input feature density estimation is important for regression. Hence, normality test for each feature is to confirm whether it is well-modeled by a normal distribution and to explore possible transformation to a normal distribution.
The aim is to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. By analyzing previous market trends and price ranges, and also upcoming developments future prices will be predicted.

There are many factors which determine the houses prices. If we look into real estate in general, then an increase in the real estate market is explained by the rise of the particular area's inhabitants' income. However, careful analysis suggests that we can only temporarily suggest that the prices of real estate are increasing due to these factors, such as demand-oriented variables and others. Therefore, we can conclude that the factors can be changed from time to time. The house prices are based on income of the inhabitants of the area, house stock supply and the payment system, whether accept installment or require cash payment.

House price prediction can be done by using multiple prediction models (Machine Learning Model) such as support vector regression, artificial neural network, and more. 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. Meanwhile, this model can help potential buyers decide the characteristics of a house they want according to their budget [5]. Previous studies focused on analyzing the attributes that affect house price and predicting house price based on the model of machine learning separately. However, this article combines such a both predicting house price and attributes together.

Other essential variables can include whether the price is the affordable, unemployment rate, demographics, and others, but we can explain house prices as a general income function. IN this study do not consider all the possible variables that can be used to predict housing prices. In this study, we use only the housing data available from the property websites to predict the housing prices by looking at the recent trends.
Prediction house prices are expected to help people who plan to buy a house so they can know the price range in the future, then they can plan their finance well. In addition, house price predictions are also beneficial for property investors to know the trend of housing prices in a certain location. Our project is a machine learning app, based on certain specifications of your future home it will try to guess the most accurate price.

Information such as state, city, area, stores etc. To avoid generation of meaningless and unwanted rules in Decision Trees, tree should not be deeper which results in over fitting. Such a tree with over fitting works more accurate with training data and less accurate with test data. Pre pruning and Post pruning are the techniques used in Decision Tree to reduce the size of the trees and avoid over fitting. In Post Pruning the Decision Tree branches and hence the level (depth) of the tree is reduced after completely building the tree. In Pre-Pruning, care is taken to avoid over fitting while building the tree itself.
Prediction house prices are expected to help people who plan to buy a house so they can know the price range in the future, then they can plan their finance well. In addition, house price predictions are also beneficial for property investors to know the trend of housing prices in a certain location.

However, careful analysis suggests that we can only temporarily suggest that the prices of real estate are increasing due to these factors, such as demand-oriented variables and others. Therefore, we can conclude that the factors can be changed from time to time. The house prices are based on income of the inhabitants of the area, house stock supply and the payment system, whether accept installment or require cash payment.

We are using Jupyter notebooks for the making the model. Jupyter notebooks are popular among data scientist because they are easy to follow and show your working steps.

There are frameworks in python to handle commonly required tasks. I Implore any budding data scientists to familiarize themselves with these libraries:

Pandas — For handling structured data

Scikit Learn — For machine learning

NumPy — For linear algebra and mathematics

Seaborn — For data visualization

1. **Jupyter Notebook-** The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. Its uses include data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more. Jupyter Notebook is a web-based interactive computational environment for creating Jupyter notebook documents**.**

2. **Pandas-** Pandas is an open-source Python package that is most widely used for data science/data analysis and machine learning tasks. It is built on top of another package named NumPy, which provides support for multi-dimensional arrays. As one of the most popular data wrangling packages, Pandas works well with many other data science modules inside the Python ecosystem, and is typically included in every Python distribution.

3. **NumPy-** NumPy, which stands for Numerical Python, is a library consisting of multidimensional array objects and a collection of routines for processing those arrays. Using NumPy, mathematical and logical operations on arrays can be performed.

4. **Scikit Learn-** scikit-learn is an open-source Machine Learning Python package that offers functionality supporting supervised and unsupervised learning. Additionally, it provides tools for model development, selection and evaluation as well as many other utilities including data pre-processing functionality.

5. **Seaborn-** Seaborn is an open-source Python library built on top of matplotlib. It is used for data visualization and exploratory data analysis. Seaborn works easily with data frames and the Pandas library. The graphs created can also be customized easily.

Chapter-2

Literature Survey

Several papers were read and assessed in order to capture all of the factors that determine the price of a home. According to several articles, the most important criteria in forecasting the price of a property is the square footage of the unit, followed by the number of bathrooms and bedrooms. Aside from that, the report claims that increasing the floor space of a house by 100 square feet raises its value by 2.6 percent.

In several instances, the location was cited as a key factor in determining the price of a home. From first-class residential districts to fifth-class residential districts, the impact of location attribute on house price decreases as expected. Hospitals, schools, campuses, and recreational parks, which can be included in the locational qualities, are the four items that have the greatest impact on housing prices, according to some papers.

Several attributes are evaluated by previous studies in their work to anticipate property prices, according to a survey of numerous articles. Locational, structural, neighborhood, and economic qualities can be used to categories all of these characteristics. The locational attribute is made up of values that describe shopping mall accessibility, school accessibility, hospital accessibility, restaurant accessibility, and public transportation accessibility. The structural attribute, on the other hand, includes elements such as the number of bedrooms, bathrooms, floor area, garage and patio, property age, and lot size. The neighborhood, on the other hand, described socioeconomic characteristics, crime rates, places of worship, beautiful scenery, and a peaceful ambiance, all of which are subjective variables. Finally, the economic aspects include the revenue factor and the material cost factor. Previous study has found that locational and structural qualities, as well as neighborhood and economic variables, are the most commonly employed attributes in house price prediction.

House price index (HPI) is used to measure price changes of residential housing in many countries, for example, the US Federal Housing Finance Agency HPI, S&P/Case-Shiller price

index, UK National Statistics HPI, UK Land Registry's HPI, UK Halifax HPI, UK Rightmove HPI and Singapore URA HPI. Since HPI is a summary figure for all transactions, it isn't enough to use it to predict a specific house price. A few documents explore the correlation among house price and local amenities, local area and renovation. After investigation the correlation between house price volatility, returns and local amenities, and proves that high amenity areas experience greater price volatility.

Using State level data in USA, examine the correlation among house prices and real per capita disposable income, common shocks, macroeconomic and local disturbances, net borrowing cost, state level population growth and spatial factors. Based on the administrative data from the Netherlands, find that higher income and wealth buyer leads to higher purchase price, while higher income and wealth seller leads to lower selling price. In Singapore property market, the main factors are private or HDB property, new or resale, freehold, potential embolic (collective sale), distance to MRT, location, size, and high-level floors, and economic cycle. Comparing with the 79 variables provided in Kaggle competition, we find that the information is incomplete, and a few features are unnecessary for prediction, there are implicit conditions in house price.

We try to find those implicit characteristics among those features, then transform those features to normal distribution and transform for increasing linearity. We also explore regression algorithms with parameters adjustment and consider the coupling effect of different algorithms to achieve better test result. The sale price of the houses using various machine learning algorithms like, lasso, SVR, Logistic regression and decision tree and compared the accuracy. the absent level problems in Random Forests, Decision Trees, and Categorical Predictors. Using three real data sets, the authors have illustrated how the absent levels affect the performance of the predictors.

The decision tree approach for finding the resale prices of houses based on their significant characteristics. In this paper, hedonic based regression method is employed for identifying the relationship between the prices of the houses and their significant characteristics.
Hedonic pricing is a price prediction model based on the hedonic price theory, which assumes that the value of a property is the sum of all its attributes value. IN the implementation, hedonic

pricing can be implemented using regression model will show the regression model in determining a price.

The data tabulation offer information of the houses includes: home id, address (street name), longitude-latitude, year, building area, land area, NJOP building price (IDR/m2), NJOP land price (IDR/m2), distance from city center(km), amount number of campuses, amount number of restaurants, amount number of health facilities, amount number of playground, amount number of schools, amount number of traditional markets or malls, amount number of worship places, and also easiness access to public transportation.

The prediction model used in this research is hedonic pricing, the suitable model using regression, with the standard formula as shown in. The dependent variable symbolized as Y is NJOP price and independent variables with symbol x1- x14 consist of year, building area, land area, NJOP land price (IDR/m 2), NJOP building price (IDR/m 2), distance to center of the city, amount number of campuses, amount number of restaurants, amount number of health facilities, amount number of amusement parks, amount number of educational facilities, amount number of traditional markets, amount number of worship places, and easiness to public transportations is shown in.

Location is an important factor in shaping the price of a house. This is because the location determines the prevailing land price. In addition, the location also determines the ease of access to public facilities, such as schools, campus, hospitals and health centers, as well as family recreation facilities such as malls, culinary tours, or even offer a beautiful scenery. The experimental process examines the parameters used on particle swarm optimization such as particle test, iteration test, and also inertia weight combination test.

As per real estate markets emphatically effect on a country's currency, which is a significant public economy scale. Mortgage holders will buy merchandise, for example, furniture and family hardware for their home, and homebuilders or project workers will buy natural substance to construct houses to fulfill house interest, which means that the financial wave impact made by the new house supply. Other than that, customers have funding to make an enormous venture,

and the development business is in acceptable condition can be seen through a nation's significant degree of house supply.

The paradigm of evaluating the house demand can be classified into two classes which are the traditional method and the advanced valuation method. The traditional valuation scheme, including multiple regression method and stepwise regression process, whilst hedonic pricing tool, artificial neural network (ANN) and spatial analysis framework are advancing valuation method. The model selection to be used to predict house price is quite critical as varieties of models are available.

Various worldwide associations and basic freedoms have accentuated house significance. House is significantly established in the monetary, monetary, and political construction of every country.

Regression analysis is a model used to determine the relationship between variables. In order to evaluate the correlation of the variables, the correlation coefficient or regression equation can be used. Multiple regression models can determine which characteristics are the most important to explain the dependent variable. Multiple regression analysis also allows certain price predictions by capturing independent and dependent variable data. In, the power of the multiple regression model can be seen when the value of the relationship between dependent and independent variables is measured. Use multiple regression modelling to describe improvements to an independent variable with a dependent variable. This model can be achieved using the house price projection as separate and dependent variables like house prices, house size, property sort, number of bedrooms, and many more. Therefore, the house price is set as a target or dependency variable, while other attributes are set as independent variables to determine the main variables by identifying the correlation coefficient of each attribute.

An increment in house demand happens every year, in a roundabout way causing house cost expands each year. The issue emerges when there are various factors, for example, area and property request that might impact the house value, in this manner

most partners including purchasers and engineers, house developers and the land business might want to know the specific credits or the exact variables affecting the house cost to assist financial backers with deciding and help house developers set the house cost.

House price prediction should be possible by utilizing a various prediction model (Machine Learning Model, for example, support vector regression, artificial neural network, and that's only the tip of the iceberg. There are many advantages that home purchasers, property financial backers, and house developers can procure from the house-value model. This model will give a great deal of data and information to home purchasers, property financial backers and house developers, like the valuation of house costs in the current market, which will assist them with deciding house cost.

Support vector regression is a predictive model based on SVM, a neural network that usually has three layers, a powerful form of supervised learning. The model is based on a subset of training data. The advantages of support vector regression are that it is capable of processing non-linear results, provides only one possible optimal solution, and able to overcome a small sample learning issue.

The potential to produce market predictions in several markets, including real estate, shows that this model can overcome the non-linear regression problems and small sample learning problems. Moreover, as this model did not depend on probability distribution assumptions, and the ability of mapping the input attribute, either linear or non-linear, this model was commonly used at house price modelling.

Support vector regression offers huge benefits in so many aspects as this model can avoid over-fitting problems, while ensuring a single optimum solution by minimizing structural risks and empirical risks. IN this field of study, support vector regression is used to collect details on neighborhood, structural and locational attributes.

The artificial neural network model has always been selected when a non-linear attribute is involved. The analysis of home price estimation should also use this model as a spatial consideration for the price of housing is also non-linear. Therefore, produces a good result, thus it is promising to provide an exact predictive model utilizing the artificial neural network algorithm. This system, however, has very limited performance. ANN can model complex non-linear relationships as house price predictions involve many non-linear variables.

In addition to the locational and structural attributes, many researchers use the attributes of neighborhood to determine house prices. This can be seen in, where he claimed that neighborhood influences affected the house price, because citizens are likely to select a better neighborhood today. The neighborhood attributes also include low crime rates, pleasant scenery and quiet atmosphere. These factors will determine whether the price of house is high or low.

However, four popular prediction models, or more known as classifiers, used by researchers to construct this predictive model which are multiple linear regression, supporting vector regression, artificial neural network and classifier gradient booster.
The hedonic price model was typically used for classification of important variables together with other regression model, such as support vector regression, multiple regression analysis and other models. Meanwhile, research by selects XGBoost as the best model since it provides the lowest RMSE value in contrast with other models in his study. Such an analysis is related to the second research question of this study.

Based on reviewing numerous papers, there are several attributes used by researchers in their work to forecast house prices. All of these attributes can be divided into 4 main categories which are locational, structural, neighborhood and economic attributes. The locational attribute consists of variables which described the accessibility to shopping mall, accessibility to school, accessibility to hospital, restaurants and availability of public transport. Meanwhile, the structural attribute consists variables of number of bedrooms, number of bathrooms, floor area, garage and patio, property age housing and lot size. The neighborhood on the other hand described the socioeconomic variables, crime rates, place of worship, pleasant landscape, and

quiet atmosphere which the variables is quite subjective. And lastly the economic attributes consist of income factor and cost of material factor.

The most common used attributes by previous research regarding house price prediction are locational and structural attributes, as neighborhood and economic attributes are difficult to define and measure. The classification of attributes made it easier to analyze the effects of different attributes on different models. The tabulation of finding based on evaluation of previous study. Looking at the table above, it can be seen that support vector regression has the smallest RMSE value, 0.0047.

Technically, the RMSE value of a model is highly dependent on the attributes used during the prediction process. Most of the model that are using the same attributes (locational attributes) will generate a very low RMSE value indicating the best model. However, this cannot simply show the best model because several previous studies rarely provide the RMSE value to justify their model being the best model. On the basis of the analysis table, the locational attribute may be assumed to be the main attribute used by several models such as support vector regression and artificial neural network support.

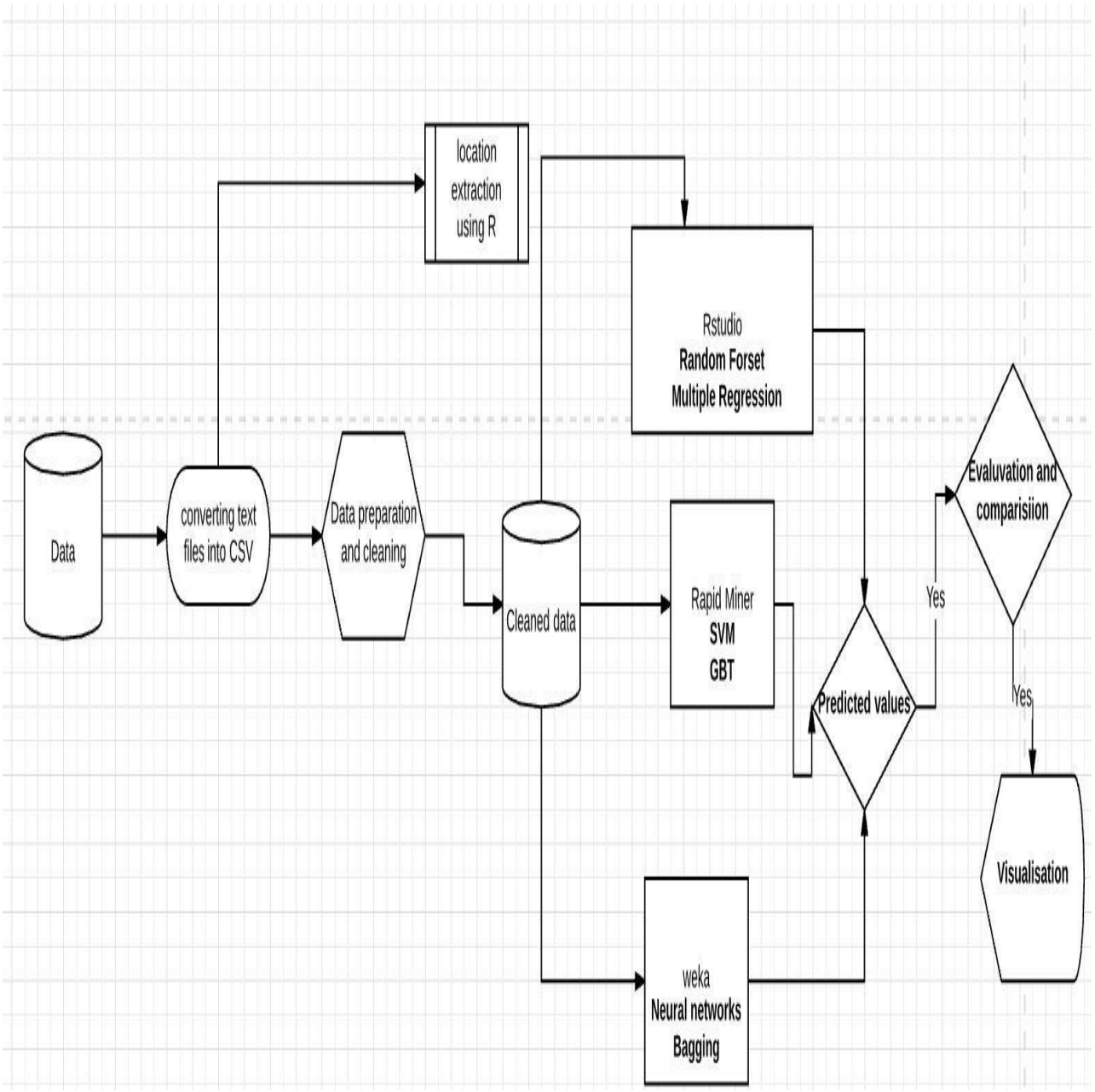
The RMSE value is very low with the presence of the locational attribute only, however, the RMSE value is quite high when the structural attribute is combined with the locational attribute for the input to make a prediction. Next, ANN provides the second lowest of RMSE value which is 0.0581. Findings revealed that locational attributes are indeed the relevant attributes used in the ANN model to forecast house price. Similar to SVR, the RMSE value is fairly high when locational and structural attributes are used together. This indicates that the locational attribute should be used alone to achieve low RMSE values by ANN model. In the meantime, the XGBoost model also yields the lowest RMSE while only the structural attribute is involved.

However, research on model with structural attributes alone is very limited as previous research focused mainly on locational attributes or the combination of locational and structural attributes in order to predict house prices. In general, our analysis suggests that SVR, ANN and XGBoost

are the most efficient models compared to other models, whereas locational attributes are the main attribute in predicting house price.

Although only a few researchers chose economic attribute, including individual income and the expenses of constructing a house, as the factor in determining a house price, we agree that economic attributes do have a major impact on house prices. Stated in his study that a house price can be determined based on an individual income because the government plays a role in setting a house price dependent on individual financial conditions. Supports this study in saying that the relationship between house price and income is important to describe the affordability of a house. This is one of the factors leading to each person's affordability to own or rent a house.

An evaluation of the main attributes that impact the house price is crucial and related to the first research question of this study. After evaluating the main attributes impacting house price determination, the data mining methodology (which is in the context of this study is a predicting model) can be used to estimate house price. In order to forecast house price, predictive modeling such as support vector regression or artificial neural network were used. Predictive modeling uses data mining to forecast what it observed during the study phase.

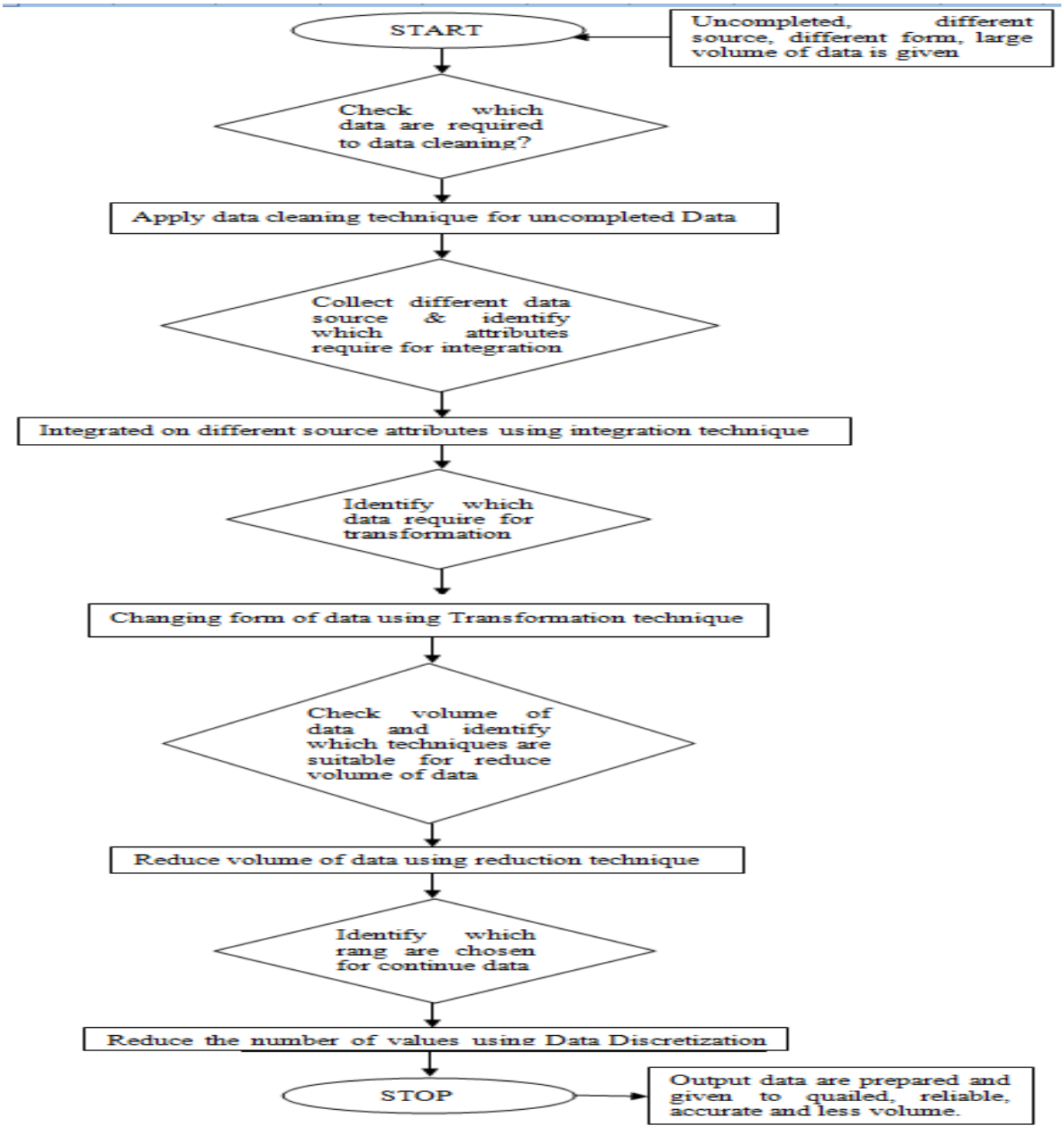


The architecture flow diagram

Chapter-3

Functionality/Working of Project

Data Preprocessing



"Delhi House Price Prediction" is a dataset with over 1260 records and nine variables that indicate house prices transacted between 2009 and 2018. These factors, which functioned as dataset attributes, were then utilized to forecast each house's average price per square meter.

The data tabulation offer information of the houses includes: home id, address (street name), longitude-latitude, year, building type, land area, NJOP building price (IDR/m2), location, BHK, area, transicption, year, per_sqft.

The location of a home has a significant impact on its pricing. This is because the current land price is determined by the location. Furthermore, the location influences the ease of access to public services such as schools, campuses, hospitals, and health centers, as well as family amusement facilities such as malls, gourmet tours, and even scenic views.

The next step was to look into the data that was missing. Variables with a missing data rate of greater than 50% would be excluded from the dataset.

Due to their ambiguity, attributes specifying the number of kitchens, baths, and drawing rooms should be removed.

• Choose a number between 1 and 4 for the number of living rooms (bedrooms were mistranslated as living rooms).

• Set minimum values for the "price" and "area" properties.

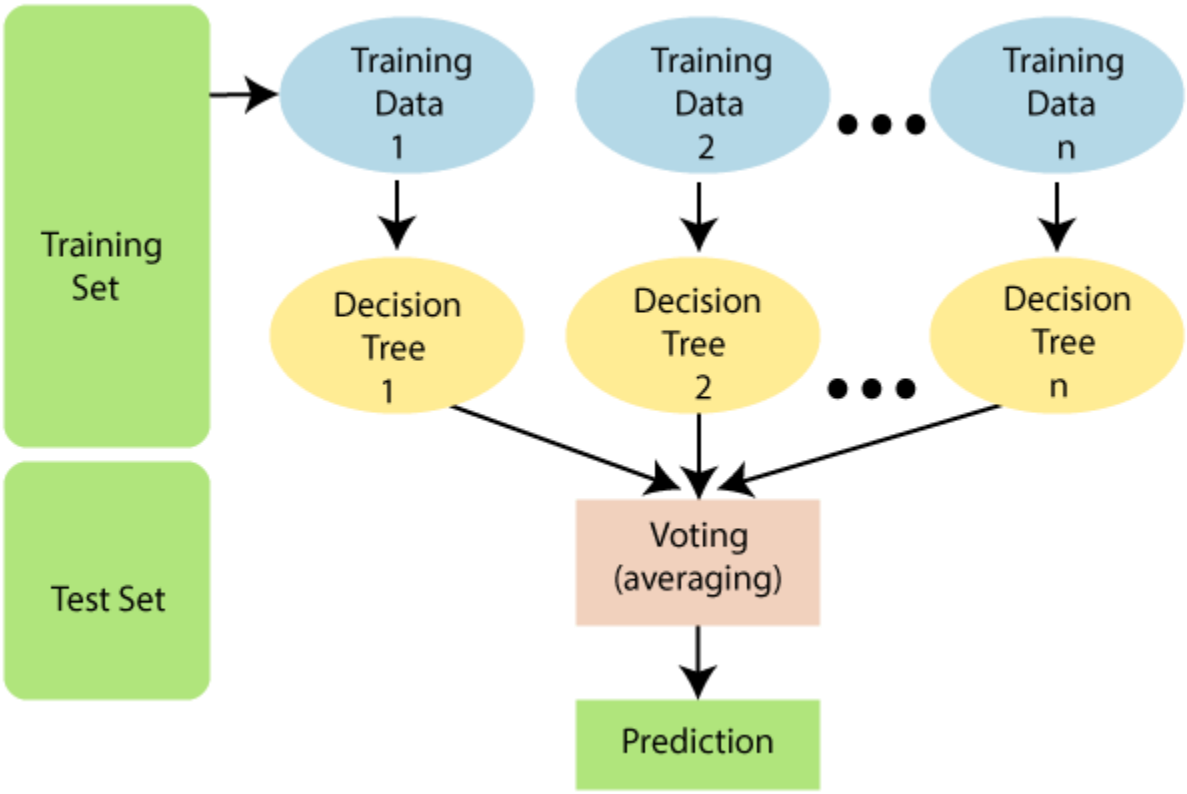
The dataset was evaluated for outliers after feature engineering. An outlier x can be found using Inter-Quartile Range (IQR) if:

$x < Q1 - 1.5 \cdot IQR$ OR $Q3 + 1.5 \cdot IQR < x$ (1)

where:

Q1 = 25th percentiles Q3 = 75th percentiles $IQR = Q3 - Q1$

Random Forest



Random Forest is an ensemble model that aggregates the predictions of numerous decision trees to produce a more precise final prediction. Random Forest has been proven to be a powerful technique in prior studies. The random forest algorithm can be broken down into the steps below:

1. Make an n-sample random bootstrap sample (randomly choose n samples from the training set with replacement).

2. From the bootstrap sample, create a decision tree with the following nodes:

(a) Pick d features at random and leave them alone.

(b) Split the node using the feature that delivers the best split in terms of the objective function, such as maximizing information gain.

3. Perform the steps a total of 1-2 thousand times. 4. By using a majority vote, assign the class label based on the predictions made by each tree.

We utilized the Random Forest Classifier class from sklearn in this paper. The n estimators' argument in the Random Forest Classifier lets you to specify how many trees to build, which we set to 900. While increasing the number of trees in the random forest improves accuracy, it also

increases the model's overall training time. The bootstrap parameter, which we set to True, is also included in the class.

However, only a limited collection of attributes will be used to introduce variation within the trees in random forest subsets. We improved the performance even further by iterating the model numerous times and adding a few parameters when we initialized the RandomForestClassifier:

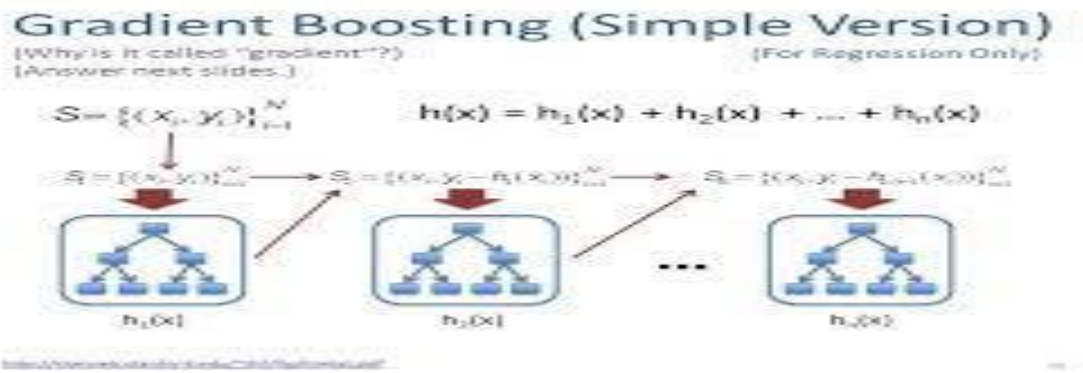
• Set max depth = 20, limiting the depth of these trees to 20.

• Set min samples split = 10, which means that a node can only be split if it has at least ten rows.

The Random Forest classifier is built utilizing an 80 percent split of the training dataset from the main dataset. After the decision tree model has been built, it is tested for accuracy by using it to categories the remaining 20% of the dataset, known as test data.

Gradient Boost

Gradient boosting is a machine learning technique that was developed in 1999 and is widely utilized due to its performance, consistency, and interpretability. Gradient boosting is cutting-edge in a number of machine learning tasks, including multistage classification, click prediction, and ranking. Gradient boosting has faced additional hurdles in recent years, particularly in terms of finding a compromise between accuracy and performance. Gradient boosting has a limited set of parameters. The following procedures can be performed to ensure a dynamic balance between fit and regularity when selecting parameters: determining regularization parameters (lambda, alpha), lowering the learning rate, and determining those ideal parameters once more.



Testing

Several approaches, including Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE), were used to test the model proposed in this study (RMSE). The average percentage of the absolute error of each anticipated result is used to calculate MAPE. As a result, MAPE may be used to determine the amount of prediction error.

MAE calculate the average of absolute error for each predicted result. MAE is useful when measuring errors in certain units. MAE values can be calculated using.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\boldsymbol{x}_i - \boldsymbol{x}|$$

RMSE is used to calculate predicted performance by considering the prediction error of each data. RMSE formula can be seen there.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \|y(i) - \hat{y}(i)\|^2}{N}},$$

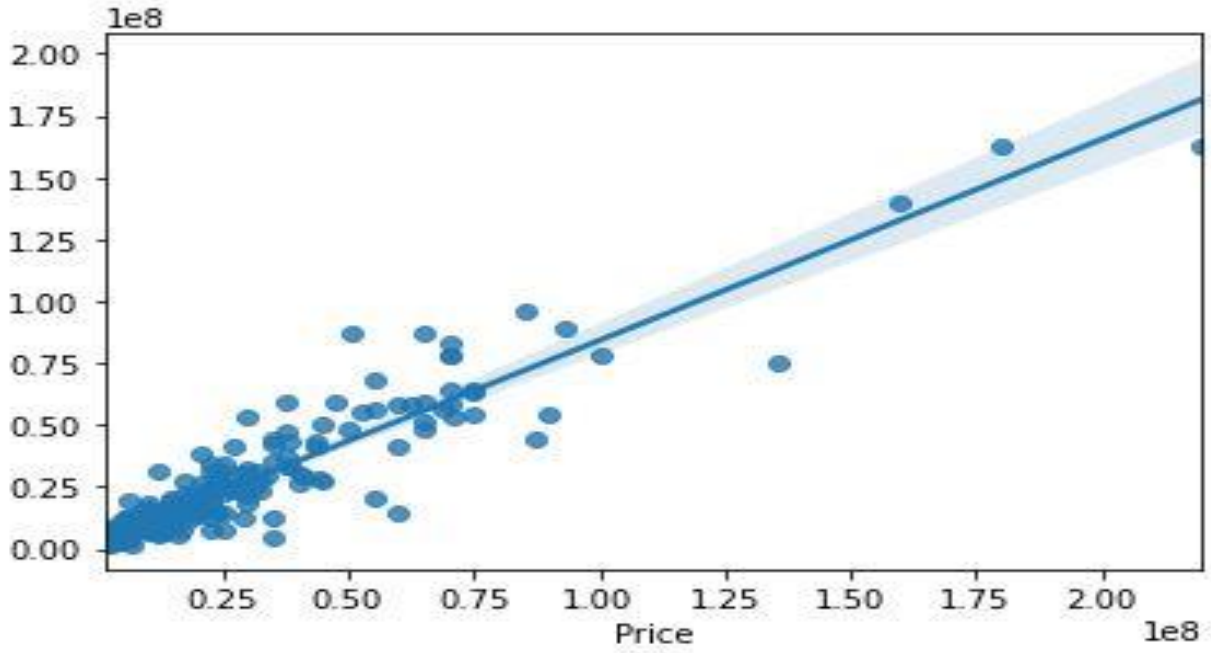
$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left| \frac{V(t) - P(t)}{V(t)} * 100 \right|$$

Chapter-4

Results

The categorization of qualities made analyzing the effects of various variables on various models much easier. The categorization of qualities made analyzing the effects of various variables on various models much easier. The best model will have a very low RMSE score since it uses the same features (locational attributes) as the majority of the models. However, because several previous studies rarely provide the RMSE value to establish their model as the best model, this cannot simply indicate the best model. The locational attribute may be presumed to be the major attribute employed by numerous models, such as support vector regression and artificial neural network support, based on the analysis table. When only the locational attribute is present, the RMSE value is very low; however, when the structural attribute is combined with the locational attribute for the input, the RMSE value is fairly high to make a prediction.

However, past research relied mostly on locational qualities or a combination of locational and structural attributes in predicting house prices, therefore research on models with structural attributes alone is sparse.


1e8
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Price
1e8

OUTPUT

Chapter-5

Conclusion and Future Scope

5.1 Conclusion

The most basic machine learning techniques, such as decision tree classifier, decision tree regression, and multiple linear regression, are used in this article. Scikit-Learn, a machine learning programmed, is used to carry out the work. This project assists users in predicting the availability of houses in the city as well as their costs.

The accurate prediction model would enable investors and home buyers to estimate a house's actual price, as well as house developers to select an affordable house price. The attributes employed by previous researchers to forecast a house price using various prediction models were examined in this work.

This model was created using a variety of input variables, all of which are highly correlated with the price of a home. Finally, the goal of this study was to aid and assist other researchers in the development of a true model that can reliably anticipate housing values. More study on an actual model is required, as well as the application of our findings to corroborate them.
In future research, alternative methodologies that match time-series data will be employed to acquire smaller error prediction values and more data to reach a better outcome.

References

[1] A. S. Temür, M. Akgün, and G. Temür, “Predicting Housing Sales in Turkey Using Arima, Lstm and Hybrid Models,” J. Bus. Econ. Manag., vol. 20, no. 5, pp. 920–938, 2019, doi: 10.3846/jbem.2019.10190.

[2] A. Ebekozien, A. R. Abdul-Aziz, and M. Jaafar, “Housing finance inaccessibility for low-income earners in Malaysia: Factors and solutions,” Habitat Int., vol. 87, no. April, pp. 27–35, 2019, doi: 10.1016/j.habitatint.2019.03.009.

[3] A. Jafari and R. Akhavian, “Driving forces for the US residential housing price: a predictive analysis,” Built Environ. Proj. Asset Manag., vol. 9, no. 4, pp. 515–529, 2019, doi: 10.1108/BEPAM-07-2018-0100. [4] Choong Wei Cheng, “Statistical Analysis of Housing Prices in Petaling,” Universiti Tunku Abdul Rahman, 2018.

[5] F. S. Gharehchopogh, T. H. Bonab, and S. R. Khaze, ―A Linear Regression Approach to Prediction of Stock Market Trading Volume: A Case Study,‖ Int. J. Manag. Value Supply Chain., vol. 4, no. 3, pp. 25– 31, 2013.

[6] 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.

[7] T. D. Phan, “Housing price prediction using machine learning algorithms: The case of Melbourne city, Australia,” Proc. - Int. Conf. Mach. Learn. Data Eng. iCMLDE 2018, pp. 8–13, 2019, doi: 10.1109/iCMLDE.2018.00017.

[8] Y. Y. S. Song, T. Zhou, H. Yachi, and S. Gao, “Forecasting house price index of China using dendritic neuron model,” PIC 2016 - Proc. 2016 IEEE Int. Conf. Prog. Informatics Comput., pp. 37–41, 2017, doi: 10.1109/PIC.2016.7949463.

[9] R. Aswin Rahadi, S. K. Wiryono, D. P. Koesrindartoto, and I. B. Syamwil, “Factors Affecting Housing Products Price in Jakarta Metropolitan Region,” Int. J. Prop. Sci., vol. 6, no. 1, pp. 1–21, 2016, doi: 10.22452/ijps.vol6no1.2.

[10] A. Nur, R. Ema, H. Taufiq, and W. Firdaus, “Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization Case Study : Malang, East Java, Indonesia,” Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 10, pp. 323–326, 2017, doi: 10.14569/ijacsa.2017.081042.

[11] A. Yusof and S. Ismail, “Multiple Regressions in Analysing House Price Variations,” Commun. IBIMA, vol. 2012, pp. 1–9, 2012, doi: 10.5171/2012.383101.

[12] A. Osmadi, E. M. Kamal, H. Hassan, and H. A. Fattah, “Exploring the elements of housing price in Malaysia,” Asian Soc. Sci., vol. 11, no. 24, pp. 26–38, 2015, doi: 10.5539/ass.v11n24p26.

[13] T. L. Chin and K. W. Chau, “A critical review of literature on the hedonic price model,” Int. J. Hous. Sci. Its Appl., vol. 27, no. 2, pp. 145–165, 2003.

[14] M. J. Ball, “Recent Empirical Work on the Determinants of Relative House Prices,” Urban Stud., vol. 10, no. 2, pp. 213–233, 1973, doi: 10.1080/00420987320080311.

[15] M. Rodriguez, “Managing Corporate Real Estate: Evidence from the Capital Markets.” Journal of Real Estate Literature, 1996.