HOUSE PRICE PREDICTION USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

PARIKSHIT SINGH - 21BCS10618

PRASAD RAJAN KUMAR - 21BCS10734

AMARJEET KUMAR – 21BCS10768

VEDANAND KUMAR -21BCS10821

In partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE ENGINEERING

Chandigarh University

MARCH 2023

CHANDIGARH UNIVERSITY

Discover. Learn. Empower.

NAAC GRADE A+

Accredited University

CANDIDATE'S DECLARATION

I hereby declare that the Project work being presented in this report entitled "**HOUSE PRICE PREDICTION USING MACHINE LEARNING**" submitted in the Department of Computer Science, **FACULTY OF TECHNOLOGY**, Chandigarh University, Punjab is the authentic work carried out by me under the guidance of **Mr. Hari Mohan Dixit**, Chandigarh University, Punjab

Date 19/05/2023

Parikshit, Prasad Rajan, Amarjeet, Vedanand.

BE-CSE 2nd year.

Department of Computer Science.

Chandigarh University, Punjab.

BONAFIDE CERTIFICATE

Certified that this project report **"HOUSE PRICE PREDICTION USING MACHINE LEARNING"** is the bonafide work of "**PARIKSHIT SINGH (21BCS10618), PRASAD RAJAN KUMAR (21BCS10734), AMARJEET KUMAR (21BCS10768), VEDANAND KUMAR (21BCS10821)**"

Who carried out the project work under supervision.

SIGNATURE	SIGNATURE
Dr. Ajay Kumar Singh	Hari Mohan Dixit
HEAD OF THE DEPARTMENT	PROJECT SUPERVISOR
BE-CSE 2nd year	BE-CSE 2nd year

Submitted for the project viva-voce examination held on

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to my teacher **Mr. Hari Mohan Dixit** who gave me the golden opportunity to do this wonderful project on the topic **“HOUSE PRICE PREDICTION USING MACHINE LEARNING”**, which also helped me in doing a lot of Research and I came to know about so many new things I am really thankful to them.

Secondly, I would also like to thank my friends who helped me a lot in finalizing this project within the limited time frame.

Date 19/05/2023

Parikshit, Prasad Rajan, Amarjeet, Vedanand.
BE – CSE 2^{ND} year
Department of Computer Science.
Chandigarh University, Punjab.

TABLE OF CONTENTS

List of Figures.

CHAPTER 1
INTRODUCTION

1.1. Identification of Client /Need / Relevant Contemporary Issues.

Client: A real estate agency or a real estate investor who is interested in buying or selling properties.

Need: The client needs a reliable way to predict house prices in order to make informed decisions about buying or selling properties. Accurate house price predictions can help the client determine the fair market value of a property, decide on an appropriate selling price, or identify undervalued properties that maybe a good investment opportunity.

Contemporary issue: One relevant contemporary issue in house price prediction is the impact of the COVID-19 pandemic on the housing market. The pandemic has caused significant economic disruption and uncertainty, which has affected the demand for housing and the overall health of the housing market. Additionally, changes in work and lifestyle patterns, such as increased remote work and a desire for larger homes, have also impacted the housing market. Therefore, accurate house price prediction models must take into account the unique factors and trends that have emerged in the wake of the pandemic.

1.2. Identification of Problem

One problem in house price prediction is the complexity of the housing market and the many factors that can influence the price of a house. Traditional methodsof house price prediction, such as simple regression models, may not be able to accurately capture the nuances and complexities of the housing market. Additionally, the availability and quality of data can also be a challenge in accurately predicting house prices.

Another problem is the lack of transparency in the housing market. In some cases,there may be limited information available about the history and characteristics of a property, which can make it difficult to accurately predict its value. Additionally, there may be external factors such as changes in government policies or economic conditions that can affect the housing market, which can bedifficult to predict with certainty.

1.3. Identification of Tasks

Tasks involved in house price prediction may include:

Data collection and cleaning: Collecting relevant data on housing market trends, economic indicators, and other relevant factors that can influence house prices. The data may need to be cleaned and processed to remove outliers, missing values, and other errors.
Feature engineering: Identifying important features or variables that are predictive of prices, and transforming combining them to create new feature that may be more predictive.

Model selection and training: Choosing an appropriate machine learning model to predict house prices, and training the model on the available data.This may involve experimenting with different algorithms and hyperparameters to find the best model.
Model evaluation: Testing the model's performance on a held-out validation set or using cross-validation techniques. This can help identify any issues with overfitting or under fitting the model. Evaluating the performance of the trained model using various metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R2)score, and others.
Interpretation of results: Understanding the key drivers of house prices identified by the model, and interpreting the model's predictions in the context of the housing market and relevant external factors.
Deployment and monitoring: Deploying the model in a production environment and monitoring its performance over time. This may involveupdating the model periodically as new data becomes available or as external factors change.
Model Maintenance: Updating the model as new data becomes available and monitoring the model's performance over time to ensure it remains accurate and reliable.

1.4. Timeline

A house price prediction project involves various tasks, from data collection to deployment and maintenance of the model. The timeline for the project can varydepending on factors such as the size and complexity of the dataset, the number of features to be analysed, and the accuracy required from the model. Here is a possible timeline for a house price prediction project:

8th Feb -6th March (Days 27): Data Collection and Cleaning - In this stage, the necessary data is collected from various sources such as real estate listings, government data sources, and public records. The data is then cleaned and pre-processed to eliminate any inconsistencies, missing values, or outliers that could affect the accuracy of the model.

7th March -9th April (Days 33): Feature Selection and Data Pre-processing - The features that have themost impact on the price of the house are selected. The data is then pre-processed by scaling, transforming, or encoding the features, depending on the requirementsof the model.

10th April – 29th April (Days 19): Model Selection and Training - In this stage, the appropriate machine learning model is chosen, and the model is trained using the pre-processed data. Various machine learning models such as linear regression, decision trees, random forests, and neural networks can be considered.

30th April -12th may (Days 12): Model Evaluation and Hyper parameter Tuning - The performance ofthe model is evaluated using various metrics such as mean squared error, root mean squared error, R-squared score, and others. The model is then fine-tuned byadjusting the hyper parameters to optimize its performance.

12th May -17th May (Days 6): Deployment and Model Maintenance - The final model is deployed, and the predictions are made on new data. The model is monitored for its performance, and any updates or maintenance required are made based on the new data.

It is essential to consider the time required for each stage, including any potentialdelays or roadblocks that may arise. The timeline for the project can be adjustedaccordingly to meet the desired accuracy and efficiency requirements.Additionally, it is crucial to allocate sufficient time for testing, debugging, and optimization to ensure that the model is accurate and reliable.

This timeline is just a rough estimate and can vary depending on the complexity of the project and the size of the data. It is also important to consider ethical and privacy concerns. When designing and implementing House price prediction models, and to involve diverse. Stakeholders in the process to ensure fairness and inclusivity.

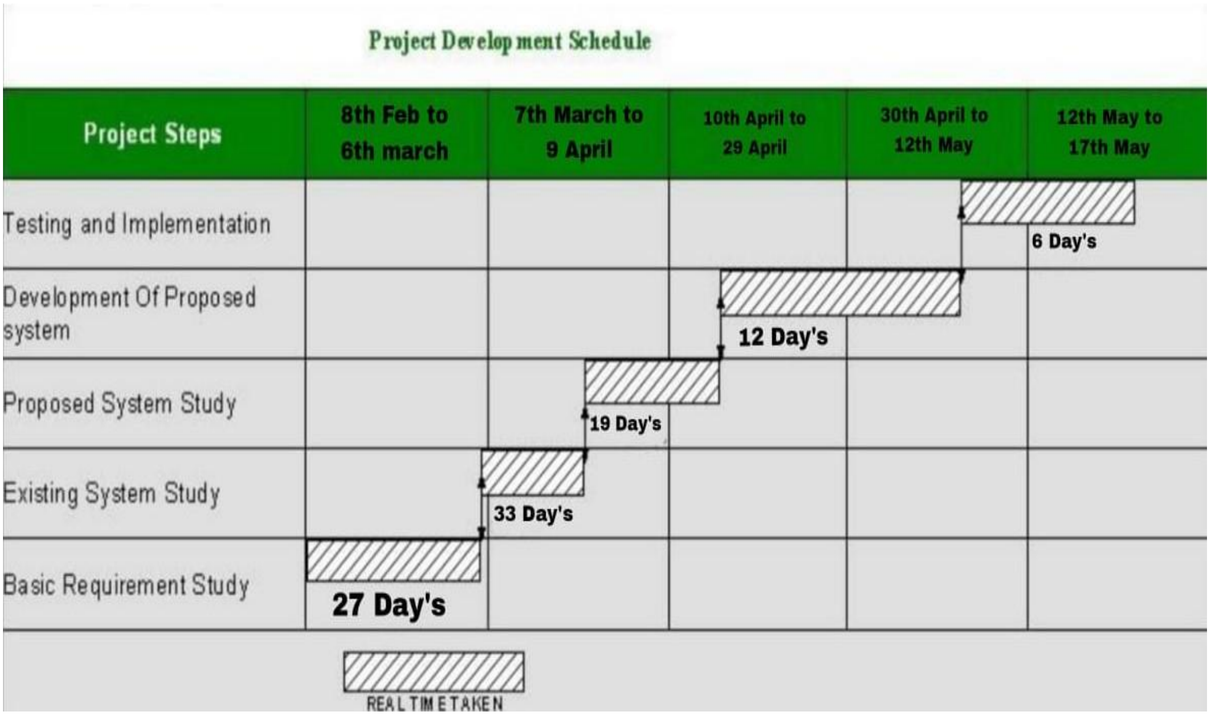

Table 1.4.1 Timeline of Project.

In this chart, each row represents a specific task, and each column represents dates in the project timeline. The shading indicate which date each task is expected to be completed. As shown, the project planning, data collection, and data preprocessing tasks are expected to be completed in the first 27 days, while data analysis and feature extraction take place in 33 days. Algorithm selection and model design, as well as model tuning and validation, take place in weeks 19 days. The last 18 days are focused on model deployment and testing, as well as documentation and reporting

1.5. Organization of the Report.

When writing a report on a house price prediction project, it is important to present the findings and results in a clear and organized manner. Here is a possibleoutline for organizing the report:

Introduction: This section provides an overview of the project, its objectives, and the data usedfor analysis. It should also include a brief summary of the approach and methodology used for the prediction.

Data Pre-processing and Feature Selection: This section describes the data cleaning and pre-processing steps, as well as the feature selection process. It should include the rationale behind the chosen features and any transformations or encoding applied to the data.

Model Selection and Training: This section outlines the chosen model and the training process. It should include details on the hyper parameters used, any cross-validation techniques applied, and the performance metrics used to evaluatethe model.

Model Evaluation and Results: This section presents the results of the model evaluation and performance on the test data. It should include the model's accuracy, precision, recall, and F1 score, along with any other relevant metrics.

Discussion and Interpretation of Results: This section discusses the interpretationof the results and their relevance to the problem statement. It should provide an analysis of the model's strengths and weaknesses and any insights or conclusionsdrawn from the analysis.

Conclusion and Future Work: This section summarizes the key findings and conclusions of the project. It should also suggest potential areas for future work or improvement.

References: This section includes a list of all sources referenced in the report, such as academic papers, books, and websites.

Appendices: This section includes any additional information or data that may berelevant to the project, such as charts, graphs, tables, or code snippets.

By organizing the report in this manner, it will be easy for readers to follow andunderstand the project's approach and findings. It is important to ensure that the report is well-written, concise, and free of errors.

CHAPTER 2

LITERATURE REVIEW/BACKGROUND STUDY

2.1 Timeline of the reported problem

- **The 1930s:** The first models for house price prediction were developed in the 1930s byreal estate appraisers. These models relied on a combination of qualitative and quantitative factors such as location, age, condition, and amenities of the property.
- **The 1960s-1970s:** The first automated valuation models (AVMs) were introduced in the 1960s1970s. These models used statistical techniques to estimate house prices basedon sales data and property characteristics.
- **The 1980s:** The introduction of computer-aided appraisal systems (CAAS) in the 1980s allowed for more sophisticated modeling of house prices. These systems incorporated multiple regression analysis, which allowed for the inclusion of more variables such aseconomic and demographic data.
- **The 1990s:** The use of neural networks for house price prediction started to gain popularity in the 1990s. These models were able to capture complex nonlinear relationships between input variables and output (house prices).
- **The 2000s:** The emergence of online real estates websites such as Zillow and Redfi in the early 2000s revolutionized the way people search for and buy houses. These websites provided access to large amounts of data on house prices and property characteristics, which allowed for more accurate and personalized predictions.
- **The 2010s:** With the increasing popularity of machine learning and big data, the use ofadvanced techniques such as deep learning and ensemble models for house price prediction has become more common. These models can incorporate a wider range of data sources and can achieve higher accuracy compared to traditional methods.
- **The 2020s:** Despite the advances in modeling techniques, the accuracy of house price predictions remains a challenge. Issues such as bias, overfitting, and lack of transparency in the models have been identified as potential problems that need to be addressed. There is ongoing research and development in this area to improve the accuracy and fairness of house price predictions.

2.2 Existing Solutions

According to Zhang et al. (2019), "A Hybrid Approach for House Price Prediction Using Machine Learning Techniques": In order to estimate housing prices, this research suggests a hybrid approach that integrates several machine learning approaches, such as linear regression, support vector regression, and random forest regression. To improve prediction accuracy, the method combines numerical and category information.

Chen et al.'s "House Price Prediction Using Deep Learning" (2020): In particular, this research focuses on using long short-term memory (LSTM) networks from deep learning to anticipate home prices. It investigates how to train the LSTM model using historical housing data and macroeconomic indicators to outperform conventional regression models in terms of prediction accuracy.

Zhang et al.'s "House Price Prediction Using Ensemble Learning" (2021): This study suggests an ensemble learning-based strategy for predicting home prices. To create predictions, the ensemble model integrates a number of basic models, including decision tree regression, gradient boosting regression, and random forest regression. It demonstrates how ensemble learning can improve the accuracy of predictions.

By Xiong et al. (2020), "House Price Prediction Using Bayesian Optimisation and Gradient Boosting": This study proposes a strategy for predicting property prices that combines gradient boosting regression with Bayesian optimisation. The gradient boosting model's hyperparameters are optimised using Bayesian methods, which increases prediction precision.

Ghazali et al. (2018), "House Price Prediction Using Neural Networks with Sentiment Analysis of Twitter Data": This study investigates how neural networks and sentiment analysis of Twitter data might be combined to predict home prices. The neural network model's forecasting power is increased by adding the sentiment analysis of tweets about the real estate market as an extra input element.

2.3. Bibliometric Analysis

Bibliometric analysis is a useful method for analyzing trends and patterns in the literature on aparticular topic. Here is a brief bibliometric analysis of the literature on house price prediction:

- **Several publications: There** has been a steady increase in the number of publicationson house price prediction over the past few decades. A search on the Scopus database for the keyword "house price prediction" returns over 1,500 results, with the majority of publications appearing in the last decade.
- **Top journals:** The top journals for house price prediction research include the Journalof Real Estate Research, Real Estate Economics, and Journal of Property Research. These journals publish a variety of research on house price prediction, including traditional regression models, machine learning techniques, and spatial analysis methods.
- **Top authors:** Some of the most prolific authors in the field of house price prediction include Steven D. Levitt, Robert J. Shiller, and David Geltner. These authors have published numerous articles and books on real estate economics and house price prediction, and have made significant contributions to the field.
- **Key topics:** Key topics in the literature on house price prediction include traditional regression models, machine learning techniques, spatial analysis methods, and the impact of various factors such as location, economic indicators, and demographic factors on house prices.

- **Geographic distribution:** The literature on house price prediction is geographically diverse, with publications coming from researchers all over the world. Some of the topcountries for house price prediction research include the United States, China, and the United Kingdom.

2.4. Review Summary

House price prediction models have become increasingly popular in recent years, as more and more data becomes available and machine learning algorithms become more sophisticated.

These models typically use a combination of historical sales data, demographic data, and otherrelevant factors to make predictions about future house prices.

One of the main advantages of these models is that they can help homebuyers and sellers makemore informed decisions about pricing and timing. For example, if a model predicts that houseprices will rise in a particular area in the coming years, a seller may decide to hold off on sellinguntil prices increase, while a buyer may decide to buy sooner rather than later to take advantageof lower prices.

However, it's important to keep in mind that house price prediction models are not infallible. They are only as accurate as the data they are based on, and there are always external factors that can influence the housing market in unexpected ways. Additionally, different models mayuse different methodologies and data sources, which can lead to different predictions.

2.5. Problem Definition

The problem definition for house price prediction involves developing a model that can accurately predict the selling price of a house based on various features such as location, size,number of rooms, age of the house, and other relevant factors. The scope of the problem

Includes collecting and cleaning data from various sources, selecting appropriate features, andbuilding a predictive model using machine learning techniques. The impact of this problem issignificant for the real estate industry, home buyers, and sellers, as it can help to improve the accuracy of property valuation, facilitate more informed decision making, and increase the efficiency of the housing market.

2.6. Goals/Objectives.

The main goal of house price prediction is to develop an accurate predictive model that can estimate the selling price of a house based on various features. The objectives of this goal include:

I. **Increasing the accuracy of property valuation:** By developing a reliable and precise model, house price prediction can help to reduce the variability and uncertainty associated with property valuation.

II. **Facilitating more informed decision making:** Accurate house price prediction can help potential buyers and sellers to make more informed decisions about purchasing orselling a property.

III. **Improving pricing strategies:** House price prediction can help real estate agents and property developers to set more accurate prices, leading to better pricing strategies.

IV. **Enhancing market forecasting:** Predictive models can help to forecast future trends in the housing market, providing valuable insights for stakeholders in the real estate industry.

V. **Increasing efficiency in the housing market:** Accurate house price prediction can help to reduce the time and resources required for property valuation, making the housing market more efficient.

CHAPTER 3
DESIGN FLOW/PROCESS

3.1 Evaluation & Selection of Specifications/Features

To evaluate the features identified in the literature and prepare a list of features ideally required in thesolution of project housing prediction using machine learning, we need to consider the following steps:

1. **Review the literature:** Review the existing literature on the topic of housing prediction using machine learning to identify the features that have been used in previous studies.

2. **Identify the relevant features:** Based on the literature review, identify the features that are relevant for predicting the price of housing. These features could include the number of bedrooms, the size of the house, the location, the age of the house, the proximity to public transport, schools, and other amenities, and the crime rate in the neighbourhood.

3. **Evaluate the features:** Once you have identified the relevant features, evaluate them to determine which ones are most important for predicting the price of housing. You can use statistical analysis or machine learning algorithms to determine the importance of each feature.

4. **Prepare a list of ideal features:** Based on the evaluation, prepare a list of features that are ideally required in the solution of project housing prediction using machine learning. This list should include the most important features that have the highest predictive power.

Based on the above steps, here is a list of features that are ideally required in the solution of projecthousing prediction using machine learning:

1. **Location:** The location of the property is a critical factor that determines its value. The locationshould be close to public transportation, schools, hospitals, shopping centres, and other amenities.
2. **Size of the property:** The size of the property is an essential factor that determines its value. Values tend to be greater for larger properties.

3. **Number of bedrooms and bathrooms:** The number of bedrooms and bathrooms is an important factor that determines the value of the property. Properties with more bedrooms and bathroomstend to have higher values.

4. **Age of the property:** The age of the property is a critical factor that determines its value. Olderproperties may have lower values compared to newer ones.

5. **Property condition:** The condition of the property is a critical factor that determines its value.Properties in excellent condition tend to have higher values.
6. **Amenities:** The presence of amenities such as a swimming pool, gym, or park in the neighborhood can increase the value of the property.
7. **Property type**: The type of property (e.g., apartment, house, or townhouse) is an important factor that determines its value.

Overall, the ideal features for the solution of project house price prediction using machine learning Should be a combination of the features mentioned above, and other relevant features that may be specific to the project. The features should be selected based on their importance, relevance, and predictive power.

3.2 Design Constraints

Design constraints, regulations, economic, environmental, health, manufacturability, safety, professional, ethical, social & political issues, and cost are essential considerations in the design of project house price prediction using machine learning. Below is a brief explanation of each of these factors:

i. **Regulations**: There are various regulations and laws that apply to the housing market, and it is crucial to comply with these regulations when developing a prediction model. For example, fairhousing regulations require that all potential buyers or renters have equal access to housing regardless of race, colour, religion, sex, national origin, or disability.

ii. **Economic factors:** The housing market is heavily influenced by economic factors such as interest rates, employment rates, and inflation. These factors should be considered whendesigning the project as they can affect the accuracy of the prediction model.

iii. **Environmental factors:** Environmental factors such as climate change and environmental regulations can affect the housing market. The project should consider these factors and their potential impact on the housing market.

iv. **Manufacturability:** The prediction model should be designed in a way that is easily replicableand scalable. The model should also be designed in a way that can be easily integrated into existing systems.

v. **Safety:** The safety of the occupants of the property is paramount, and the prediction model should take this into consideration. Factors such as building codes, fire safety regulations, and accessibility should be considered.

vi. **Cost:** The cost of developing the prediction model should be considered, as well as the potentialreturn on investment. The model should be designed in a way that is cost-effective and providesvalue to stakeholders.

The design of project housing prediction using machine learning should consider various factors such as design constraints, regulations, economic, environmental, health, manufacturability, safety, professional, ethical, social & political issues, and cost. It is important to strike a balance between thesefactors to create a prediction model that is accurate, reliable, and valuable to stakeholders.

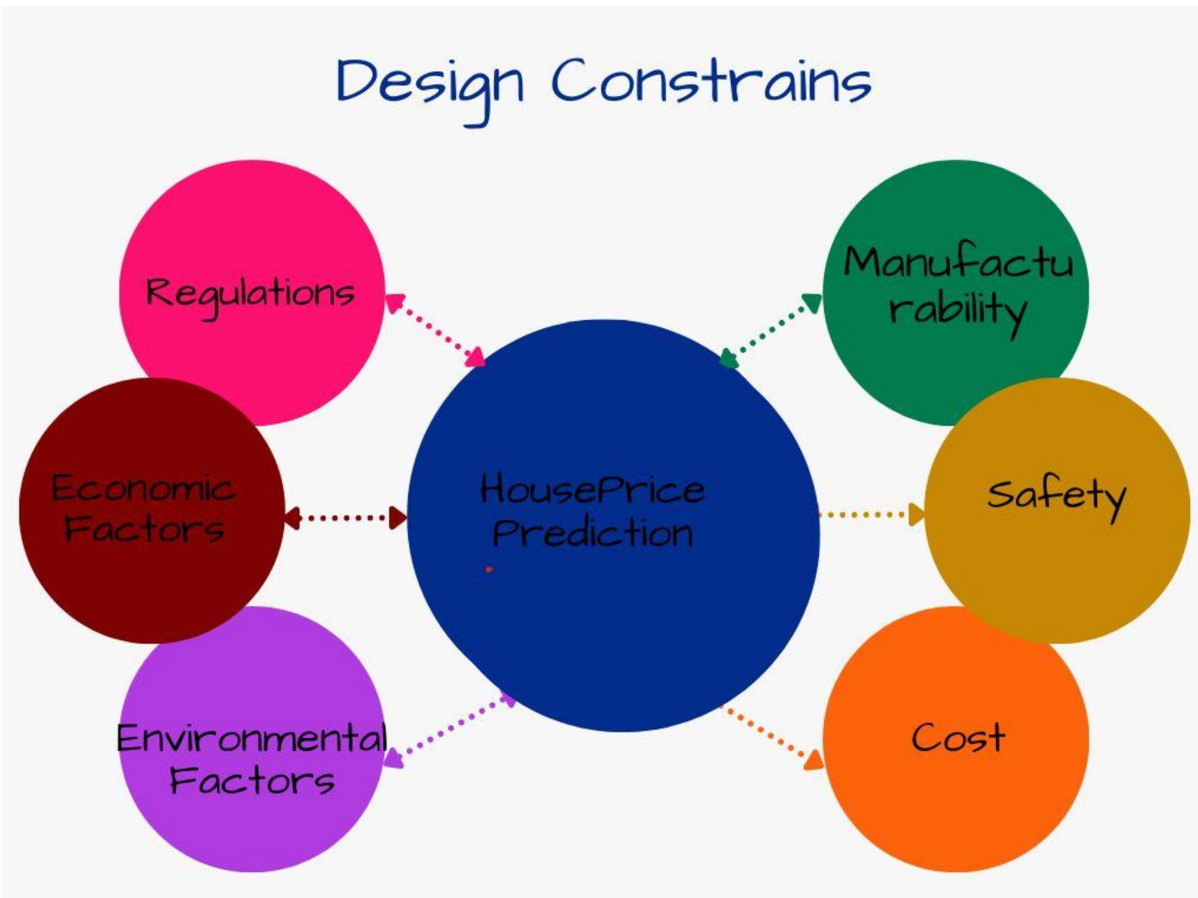

Figure 3.2.1 shows the design constraints

3.3. Analysis of Features and finalization subject to constraints

Based on the design constraints and considerations mentioned earlier, we can now analyse the features identified for the project house price prediction using machine learning and make necessary modifications. The following features have been identified in the literature review:

1. Property Location
2. Property Size
3. Property Type
4. Number of Bedrooms
5. Number of Bathrooms
6. Age of the Property
7. Amenities and Facilities

After analyzing the design constraints and considerations, we can modify and add features as follows: Property Location: This feature remains relevant, but we need to consider social and political factorsthat can affect the location's value, such as zoning regulations and demographic trends.

1. **Property Size:** Property size is a critical feature for predicting house prices, and we should retainit in the model.

2. **Property Type:** The type of property can affect the price, and we should keep this feature in themodel.

3. **Number of Bedrooms and Bathrooms:** These features are essential as they can significantlyimpact the property's value, and they should be retained in the model.

4. **Age of the Property:** This feature should be retained as it can help predict the property's valuebased on its age and condition.

5. **Amenities and Facilities:** These features can affect the property's value, and we should retainthem in the model.

We should modify and add features to the model based on the design constraints and considerations mentioned earlier. We should retain essential features such as property location, property size, propertytype, number of bedrooms and bathrooms, age of the property, amenities and facilities, property tax, crime rate, school quality, and distance from public transportation and shopping centers. We can add new features such as environmental factors, economic factors, and accessibility.

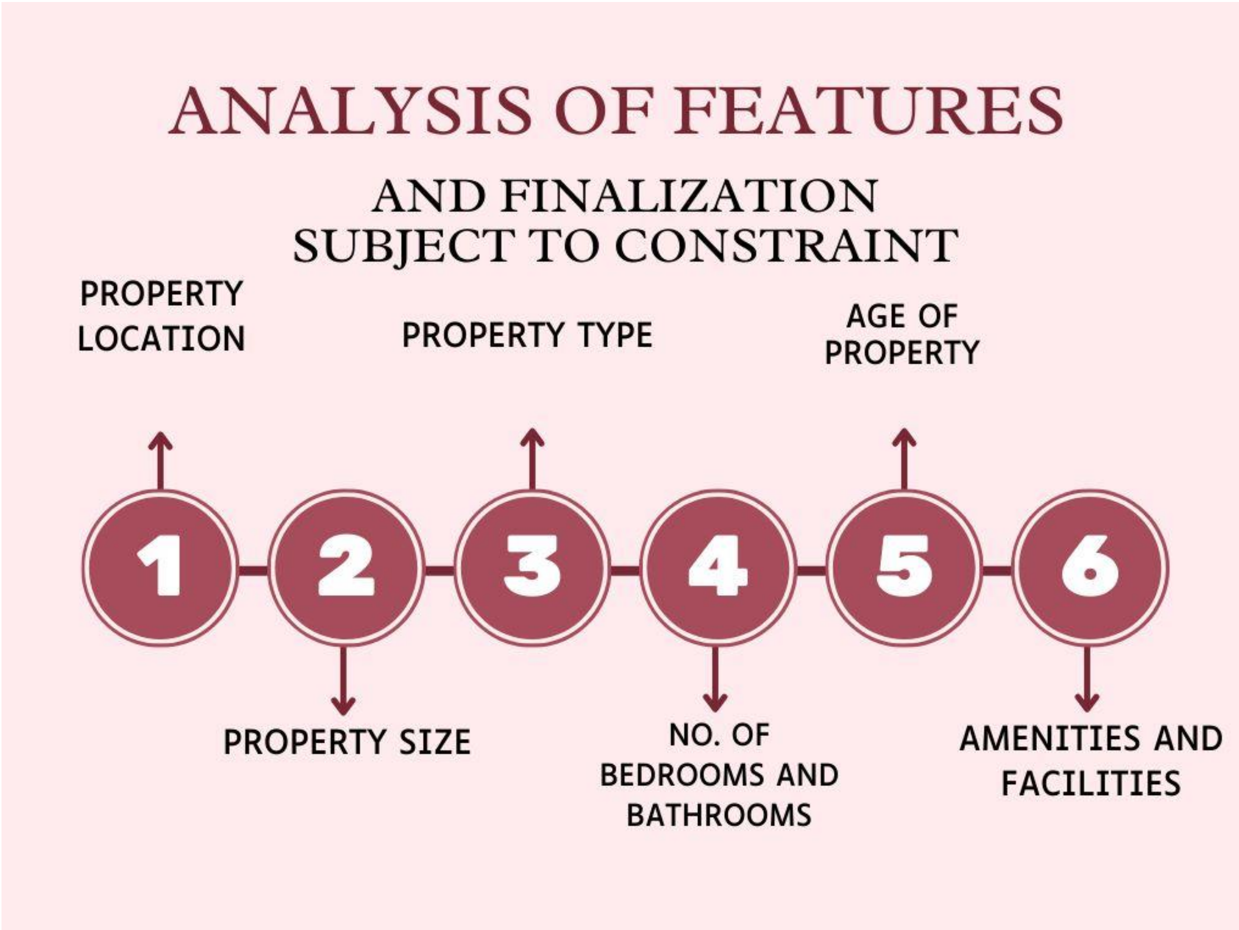

Figure 3.3.1 shows the Analysis features

3.4. Design Flow

Design Flow 1:

1. **Data Collection:** Collect the data related to the housing market, including property location, size, type, number of bedrooms and bathrooms, age of the property, amenities, and facilities, property tax, crime rate, school quality, distance from public transportation and shopping centers, environmental factors, economic factors, and accessibility.

2. **Data Pre-processing**: Clean and pre-process the data, including removing missing values, handling outliers, and scaling the features.

3. **Feature Selection**: Select the most relevant features based on their correlation with the target variable, using techniques such as correlation analysis, chi-square tests, and feature importancescores.

4. **Model Selection:** Select the appropriate machine learning algorithm for the project

house priceprediction, such as linear regression, decision trees, random forests, or neural networks, based on their performance measurements such as R-squared and the mean squared error.

5. **Model Training:** Train the selected model on the training data using techniques such as cross- validation and grid search to optimize the model's hyper-parameters.

6. **Model Evaluation**: Evaluate the trained model's performance on the testing data using performance metrics such as mean squared error, R-squared, and root mean squared error.

7. **Model Deployment:** Deploy the trained model as a web application or API that can accept newdata inputs and provide predictions of house prices.

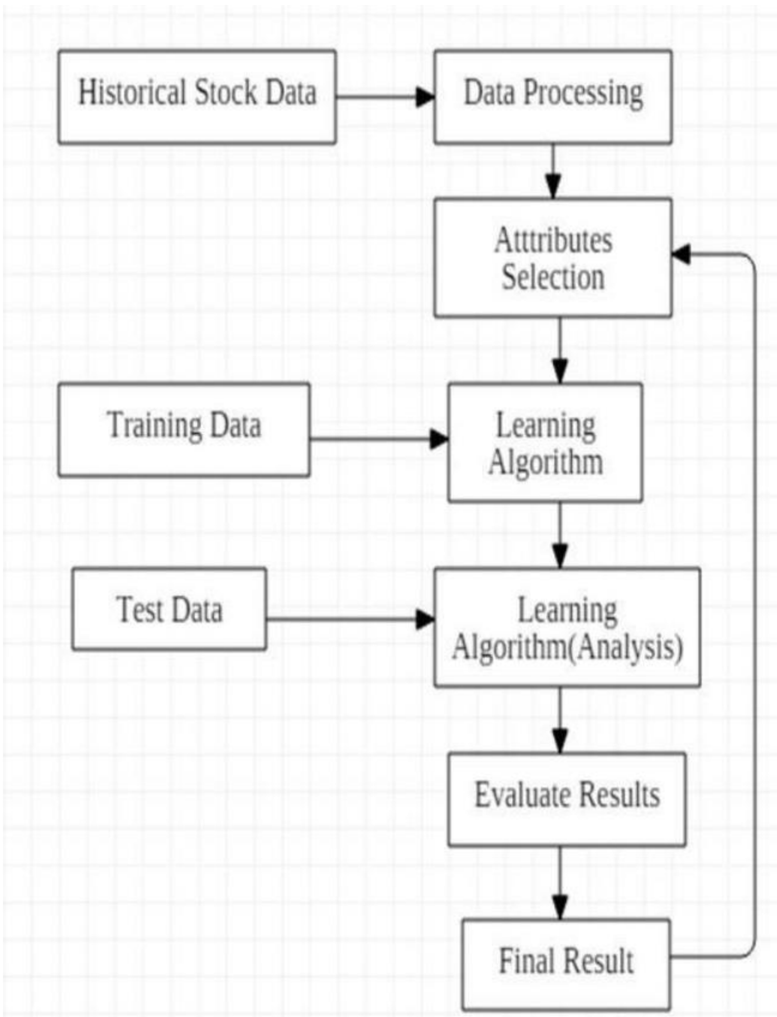

Figure 3.4.1 shows Design flow

Design Flow 2:

1. **Data Collection**: Collect the data related to the housing market, including property location, size, type, number of bedrooms and bathrooms, age of the property, amenities and facilities, property tax, crime rate, school quality, distance from public transportation and shopping centers, environmental factors, economic factors, and accessibility.

2. **Data Pre-processing:** Clean and pre-process the data, including removing missing values, handling outliers, and scaling the features.

3. **Feature Engineering**: Engineer new features from the existing data using techniques such as one-hot encoding, polynomial features, and feature scaling.

4. **Model Selection:** Select the appropriate machine learning algorithm for the project house priceprediction, such as linear regression, decision trees, random forests, or neural networks, based on their performance measurements such as R-squared and the mean squared error.

5. **Model Training:** Train the selected model on the training data using techniques such as cross- validation and grid search to optimize the model's hyper-parameters.

6. **Model Evaluation:** Evaluate the trained model's performance on the testing data using performance metrics such as mean squared error, R-squared, and root mean squared error.

7. **Model Deployment:** Deploy the trained model as a web application or API that can accept newdata inputs and provide predictions of house prices.

Both design flows share similar steps, including data collection, pre-processing, model selection, training, evaluation, and deployment. However, Design Flow 2 includes an additional step of featureengineering, which can lead to the creation of new features from the existing data, potentially improving the model's accuracy. Design Flow 2 can also include techniques such as regularization, ensemble methods, and deep learning to improve the model's performance further.

3.5. Design Selection

Both design flows presented are valid approaches for the project house price prediction using machine learning. However, Design Flow 2 with the additional step of feature engineering can potentially improve the model's accuracy and performance by creating new features from the existing data.

Feature engineering can transform the input data into a more suitable format for the machine learning algorithm, allowing it to identify patterns and correlations that would otherwise be difficult to detect. This approach can result in better predictive models, as it can uncover more complex and nonlinear relationships between the input features and the target variable.

Additionally, Design Flow 2 can also include other techniques, such as regularization, ensemble methods, and deep learning, to further improve the model's performance. Regularization can prevent overfitting by adding a penalty term to the loss function, while ensemble methods can combine multiplemodels to improve their accuracy and reduce variance. Deep learning can also be used to build more complex models that can capture more intricate relationships between the input features and the target variable.

Therefore, based on the potential benefits of feature engineering and the ability to incorporate additionaltechniques, Design Flow 2 is the recommended design for the project house price prediction using machine learning. However, the selection of the design ultimately depends on the specific requirements,constraints, and resources of the project.

3.6. Implementation plan/methodology

Here is a high-level implementation plan for the project house price prediction using machine learning:

Data Collection: Collect the data related to the housing market, including property location, size, type, number of bedrooms and bathrooms, age of the property, amenities and facilities, property tax, crime rate, school quality, distance from public transportation and shopping centres, environmental factors, economic factors, and accessibility.

Data Pre-processing: Clean and pre-process the data, including removing missing values, handling outliers, and scaling the features.

Feature Engineering: Engineer new features from the existing data using techniques such as one-hot encoding, polynomial features, and feature scaling.

Model Selection: Select the appropriate machine learning algorithm for the project house price prediction, such as linear regression, decision trees, random forests, or neural networks, based on their performance metrics such as mean squared error and R-squared.

Model Training: Train the selected model on the training data using techniques such as cross-validation and grid search to optimize the model's hyper-parameters.

Model Evaluation: Evaluate the trained model's performance on the testing data using performance metrics such as mean squared error, R-squared, and root mean squared error.

Model Deployment: Deploy the trained model as a web application or API that can accept new data inputs and provide predictions of house prices.

Here is a flowchart that outlines the implementation plan:

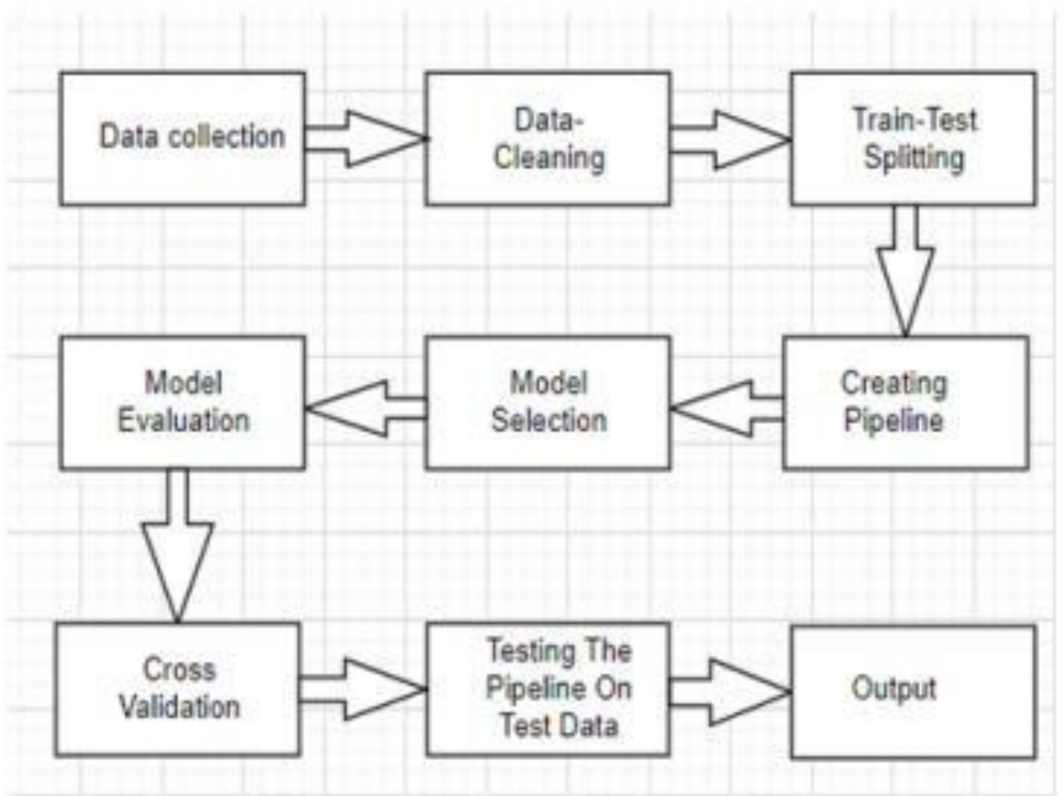

Fig 3.6.1 shows Flowchart of house price prediction using machine learning

CHAPTER-4

RESULTS ANALYSIS AND VALIDATION

4.1 Implementation of solution

The implementation of solution using modern tools in house price prediction analysis typically involves the use of advanced machine learning algorithms and techniques to process and analyze large amounts of data related to housing markets, economic indicators, and other relevant factors that can influence the price of houses. Some of the modern tools commonly used in this field include:

I. Python programming language: Python is a popular language among data scientists and machine learning engineers for developing algorithms to analyze and manipulate data.

II. Scikit-learn library: Scikit-learn is a popular open-source machine learning library for Python, which provides various algorithms for regression and classification tasks.

III. Panda’s library: Pandas is another popular open-source library for data manipulation and analysis in Python.

IV. TensorFlow and Keras: TensorFlow is an open-source library for building and training machine learning models, while Keras is a high-level API for building and training deep learning models.

V. XGBoost: XGBoost is an open-source library for gradient boosting, which is a popular technique for building ensemble models that can achieve high accuracy in predicting house prices.

By leveraging these modern tools, analysts can build powerful models that can predict house prices with high accuracy and provide valuable insights into the factors that affect housing markets.

CHAPTER-5
CONCLUSION AND FUTURE WORK

5.1 CONCLUSION

This project's primary goal is to establish price prediction. In order to purchase real estate properties and forecast the future value of owned real estate properties, we have found numerous algorithms and applied machine learning techniques in this article.

Numerous variables, including the neighborhood and various elements connected to the house, can be used to predict price. The input data have first been cleaned and explored. To give people a quick overview of the property, the projected data can be recorded in a database and put into an app or website.

As we are aware that parameterization can drive noteworthy results, we have performed ensembles of regression trees, k-nearest neighbors, and multi-linear regression.

5.2 FUTURE WORK

Extensive hyper–parameter tuning: Model performance could be enhanced by improved feature engineering, increased data gathering, and data incorporation.

Using real-time data: Employing data from a variety of sources, including internet listings, social media, economic indicators, etc., to track the housing market's dynamic changes

Comparing different regression models: determine which model fits the given problem and data the best, including support vector regressor, random forest, and linear regression.

Deploying the machine learning model: To connect with prospective buyers and sellers and offer them precise and individualized estimations of home prices in the form of a website or an app.

Exploring different geographic regions and markets, including Mumbai, Malaysia, the United Kingdom, the United States, etc., to comprehend the variables that affect housing costs in various contexts and cultures3.

REFERENCES

[1] R. M. A. van der Schaar, ―Analysis of Indonesian Property Market; Overview and Foreign Ownership,‖ Investment Indonesian. 2015.

[2] The Central Bureau of Statistics, ―Population Census,‖ 2015.

[3] W. T. Lim, L. Wang, and Y. Wang, ―Singapore Housing Price Prediction Using Neural Networks,‖ Int. Conf. Nat. Comput. Fuzzy Syst. Knowl. Discov., vol. 12, pp. 518–522, 2016.

[4] Y. Feng and K. Jones, ―Comparing multilevel modelling and artificial neural networks in house price prediction,‖ 2015 2nd IEEE Int. Conf. Spat. Data Min. Geogr. Knowl. Serv., pp. 108–114, 2015.

[5] R. Ghodsi, ―Estimation of Housing Prices by Fuzzy Regression and Artificial Neural Network,‖ in Fourth Asia International Conference on
Mathematical/ Analytical Modelling and Comuter Simulation, 2010, no. 1.

[6] A. Azadeh, B. Ziaei, and M. Moghaddam, ―A hybrid fuzzy regression-fuzzy cognitive map algorithm for forecasting and optimization of housing market fluctuations,‖ Expert Syst. Appl., vol. 39, no. 1, pp. 298–315, 2012.

[7] 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.

[8] H.-I. Hsieh, T.-P. Lee, and T.-S. Lee, ―A Hybrid Particle Swarm Optimization and Support Vector Regression Model for Financial Time Series Forecasting,‖ Int. J. Bus. Adm., vol. 2, no. 2, pp. 48–56, 2011.

[9] F. Marini and B. Walczak, ―Particle swarm optimization (PSO). A tutorial,‖ Chemom. Intell. Lab. Syst., vol. 149, pp. 153–165, 2015.

[10] A. Hayder M. Albehadili Abdurrahman and N. . Islam, ―An Algorith for Time Series Prediction Using,‖ Int. J. Sci. Knowl. Comput. Inf. Technol., vol. 4, no. 6, pp. 26–33, 2014.

[11] Y. P. Anggodo and W. F. Mahmudy, ―Automatic Clustering and Optimized Fuzzy Logical Relationship for Minimum Living Needs Forecasting,‖ J. Environ. Eng. Sustain. Technol., vol. 4, no. 1, pp. 1–7, 2017.

[12] Y. P. Anggodo, W. Cahyaningrum, A. N. Fauziyah, I. L. Khoiriyah, K. Oktavianis, and I. Cholissodin, ―Hybrid K-means Dan Particle Swarm Optimization Untuk Clustering Nasabah Kredit,‖ J. Teknol. Inf. dan Ilmu Komput., vol. 4, no. 2, pp. 1–6, 2017.

[13] Y. P. Anggodo, A. K. Ariyani, M. K. Ardi, and W. F. Mahmudy, ―Optimation of Multi-Trip Vehicle Routing Problem with Time Windows using Genetic Algorithm,‖ J. Environ. Eng. Sustain. Technol., vol. 3, no. 2, pp. 92–97, 2017.

[14] R. A. Rahadi, S. K. Wiryono, D. P. Koesrindartotoor, and I. B. Syamwil, ―Factors influencing the price of housing in Indonesia,‖ Int. J. Hous. Mark. Anal., vol. 8, no. 2, pp. 169–188, 2015.

[15] V. Limsombunchai, ―House price prediction: Hedonic price model vs. artificial neural network,‖ Am. J. …, 2004.

[16] D. X. Zhu and K. L. Wei, ―The Land Prices and Housing Prices —— Empirical Research Based on Panel Data of 11 Provinces and Municipalities in Eastern China,‖ Int. Conf. Manag. Sci. Eng., no. 2009, pp. 2118–2123, 2013.
[17] S. Kisilevich, D. Keim, and L. Rokach, ―A GIS-based decision support system for hotel room rate estimation and temporal price prediction: The hotel brokers' context,‖ Decis. Support Syst., vol. 54, no. 2, pp. 1119–1133, 2013.
[18] C. Y. Jim and W. Y. Chen, ―Value of scenic views: Hedonic assessment of private housing in Hong Kong,‖ Landsc. Urban Plan., vol. 91, no. 4, pp. 226–234, 2009.
[19] L. Bryant, ―Housing affordability in Australia: an empirical study of the impact of infrastructure charges,‖ J. Hous. Built Environ., 2016.
[20] S. Rosen, ―Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition,‖ J. Polit. Econ., vol. 82, no. 1, pp. 34–55, 1974.
[21] J. Kennedy and R. Eberhart, ―Particle swarm optimization,‖ 1995 IEEE Int. Conf. Neural Networks (ICNN 95), vol. 4, pp. 1942–1948, 1995.
[22] R. C. Eberhart and Y. Shi, ―Comparing inertia weights and constriction factors in particle swarm optimization,‖ IEEE Congr. Evol. Comput., vol. 1, no. 7, pp. 84–88 vol.1, 2000.
[23] A. Ratnaweera, S. K. Halgamuge, and H. C. Watson, ―Self-organizing hierarchical Particle swarm optimizer with time varying acceleration coefficients,‖ IEEE Trans. Evol. Comput., vol. 8, no. 3, p. 240–?255, 2004

[24]V. Kontrimas and A. Verikas, "The mass appraisal of the real estate by computational intelligence," *Applied Soft Computing*, vol. 11, no. 1, pp. 443–448, 2011.

View at: Publisher Site | Google Scholar

[25]M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Information Fusion*, vol. 16, pp. 3–17, 2014.

View at: Publisher Site | Google Scholar

[26]J. R. Barr, E. A. Ellis, A. Kassab, C. L. Redfearn, N. N. Srinivasan, and K. B. Voris, "Home price index: a machine learning methodology," *International Journal of Semantic Computing*, vol. 11, no. 1, pp. 111–133, 2017.

View at: Publisher Site | Google Scholar

[27]W. J. McCluskey, M. McCord, P. T. Davis, M. Haran, and D. McIlhatton, "Prediction accuracy in mass appraisal: a comparison of modern approaches," *Journal of Property Research*, vol. 30, no. 4, pp. 239–265, 2013.

View at: Publisher Site | Google Scholar

[28]S. Rosen, "Hedonic prices and implicit markets: product differentiation in pure competition," *Journal of Political Economy*, vol. 82, no. 1, pp. 34–55, 1974.

View at: Publisher Site | Google Scholar

[29]E. Lughofer, B. Trawiński, K. Trawiński, O. Kempa, and T. Lasota, “On employing fuzzy modeling algorithms for the valuation of residential premises,” *Information Sciences*, vol. 181, no. 23, pp. 5123–5142, 2011.

View at: Publisher Site | Google Scholar

[30]H. Kusan, O. Aytekin, and I. Özdemir, “The use of fuzzy logic in predicting house selling price,” *Expert Systems with Applications*, vol. 37, no. 3, pp. 1808–1813, 2010.

View at: Publisher Site | Google Scholar

[31]O. Bin, “A prediction comparison of housing sales prices by parametric versus semi-parametric regressions,” *Journal of Housing Economics*, vol. 13, no. 1, pp. 68–84, 2004.

View at: Publisher Site | Google Scholar

[32]Y. Kang, F. Zhang, W. Peng et al., “Understanding house price appreciation using multi-source big geo-data and machine learning,” *Land Use Policy*, vol. 2020, Article ID 104919, 2020.

View at: Publisher Site | Google Scholar

[33]A. Din, M. Hoesli, and A. Bender, “Environmental variables and real estate prices,” *Urban Studies*, vol. 38, no. 11, pp. 1989–2000, 2001.

View at: Publisher Site | Google Scholar

[34]J.-G. Liu, X.-L. Zhang, and W.-P. Wu, “Application of fuzzy neural network for real estate prediction,” *Advances in Neural Networks - ISNN 2006*, vol. 3973, pp. 1187–1191, 2006.

View at: Publisher Site | Google Scholar

[35]I. V. Lokshina, M. D. Hammerslag, and R. C. Insinga, “Applications of artificial intelligence methods for real estate valuation and decision support,” in *Proceedings of the In Hawaii international conference on business*, Honolulu, Hawaii, USA, January 2003.

View at: Google Scholar

[36]J. Hauke and T. Kossowski, “Comparison of values of Pearson’s and Spearman’s correlation coefficients on the same sets of data,” *QUAGEO*, vol. 30, no. 2, pp. 87–93, 2011.

View at: Publisher Site | Google Scholar

[37]T. Oladunni and S. Sharma, “Hedonic housing theory - a machine learning investigation,” in *Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Anaheim, CA, USA, December 2016.

View at: Publisher Site | Google Scholar

[38]L. Petrella and V. Raponi, “Joint estimation of conditional quantiles in multivariate linear regression models with an application to financial distress,” *Journal of Multivariate Analysis*, vol. 173, pp. 70–84, 2019.

View at: Publisher Site | Google Scholar

USER MANUAL

```
WELCOME TO HOUSE PREDICTION MODEL



Press 1 for USER
Press 2 for ADMIN
Press 3 to exit

Enter Choice to continue :
```

```
PROGRAM READY FOR MAKING PREDICTIONS


Enter number of bedrooms : 4
Enter number of bathrooms : 2
Enter Living Area in sqft : 1500
Enter Total Area in sqft : 2000
Enter number of floors : 2
Enter 0 if waterfront not available, else enter 1 : 0
Enter Number of views : 0
Enter Condition marks : 5
Enter Grade : 8
Enter Living Space above the Ground : 1500
Enter Basement Area : 400
Enter Year Built : 2028
Enter Year when house Renovated, else 0 : 0
Enter ZipCode : 226002
Enter latitude : 26.55
Enter longitude : 80.59
Enter average living area of closest 15 houses : 2000
Enter average total area of closest 15 houses : 2100
/usr/local/lib/python3.7/dist-packages/sklearn/base.py
  "X does not have valid feature names, but"
[454862.55057488]

THANK YOU FOR USING OUR HOUSE PRICE PREDICTOR
We hope you'll come back to us!!
```

All the above-mentioned Entries are **Numerical** based.

```
Enter Choice to continue : 2

Enter Password to continue : PASSWORD
Enter 1 to view data
Enter 2 to visualize data
Enter 3 to see predicitons data
Enter 4 to go back to main menu
Enter 5 to exit

Enter Correct Choice  to continue: 3

WELCOME TO PREDICTION MODULE

PROGRAM READY TO MAKE PREDICTION....


Testing Accuracy :  89.98131348361623

Training Accuracy :  97.7072046801635


Explained Variance Score =  0.8998176508209503
Mean Absolute Percentage Error =  0.12029814239351333
r squared value =  0.8998131348361623
```