

REAL ESTATE PRICE PREDICTION USING MACHINE LEARNING

Tanish Dalal

10TH September 2023

Abstract

The study explores real estate price prediction with machine learning, considering the perspectives of both buyers and sellers. Recognizing property prices is a skill that is not commonly found among people today. Here we will use big data processing to solve the given problem. We will use Linear regression to bring in positive results from the dataset. We will check the need for REPP, how it affects us, and what are the laws and regulations for REPP. What are the factors that are considered for any real estate, and what are the various types of internal and external constraints for a particular REPP? We will also study some past studies regarding REPP. We will study the expectations of buyers and sellers. We will do research by covering design problems and valuable resources available over the internet.

1. Introduction

In today's economy, the demand for ideal housing continues to grow. Therefore, we aim to create a real estate price prediction algorithm. Keeping in mind both seller and buyer. When purchasing a residential property, it's important to consider various factors such as rights, benefits, interests, and liabilities that arise after ownership. Sometimes we habitually purchase overpriced properties that do not match their value due to location, condition, growth potential, or distance from services.

Real estate price prediction is the process of using historical data and various statistical methods to forecast the future value of a property. It is a complex task, as many factors, such as economic conditions, interest rates, demographics, and local market trends, can influence real estate prices. However, real estate price prediction can be a valuable tool for both buyers and sellers, as it can help them to make more informed decisions about their transactions.

According to a 2017 survey conducted in the United States housing market, the number of houses sold in 2016 was 5.42 million. The survey also revealed that the inventory of starter homes decreased by 10.7% compared to 2015(1). Many players in the commercial center, such as land agents, appraisers, brokers, and developers, make a fair profit from property holdings.

1.1 Data preprocessing

Data preparation is crucial for machine learning to provide precise and insightful results. The more reliable the data, the more dependent we are on the outcomes. Large real-world datasets

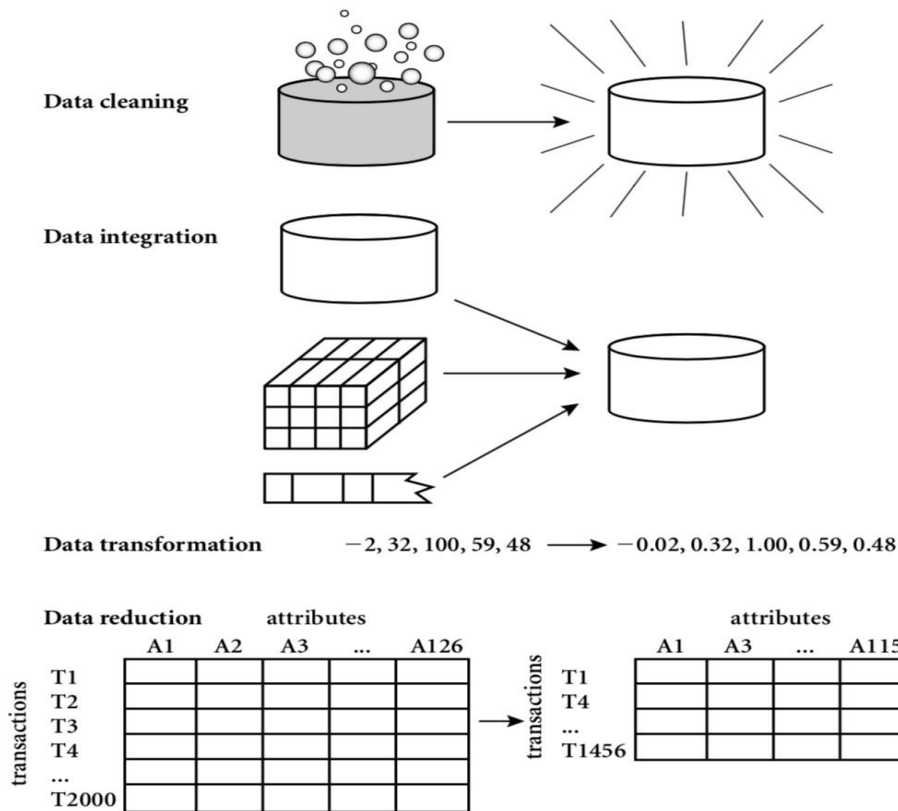
typically have inconsistent, noisy, and incomplete data. By completing missing or partial data, reducing noise, and addressing inconsistencies, data preprocessing improves the data quality.

Numerous factors can lead to incomplete data. It's possible that certain properties, including client information for sales transaction data, are occasionally not accessible. Relevant information could have been overlooked as a result of a misinterpretation or faulty technology.

Noisy data (data with inaccurate attribute values) can occur for a variety of causes. The tools used to acquire the data might be flawed. Data input mistakes might have been caused by either humans or computers. Data transfer errors might also happen. Additionally, mismatched naming standards, data codes, or input field formats, such as date, might lead to incorrect data.

The Data preprocessing techniques available are –

- **Data Cleaning**-Filling in missing numbers, eliminating noise, addressing discrepancies, and locating and eliminating outliers in the data are all examples of data cleaning.
- **Data integration**- is the method of combining data from several sources into one cohesive data storage, like a data warehouse.
- **Data transformation**- It is possible to apply data transformations such as normalization. For instance, normalization may increase the precision and effectiveness of distance-based mining algorithms.
- **Data reduction** can lower the quantity of the data by, for example, removing redundant characteristics or clustering.



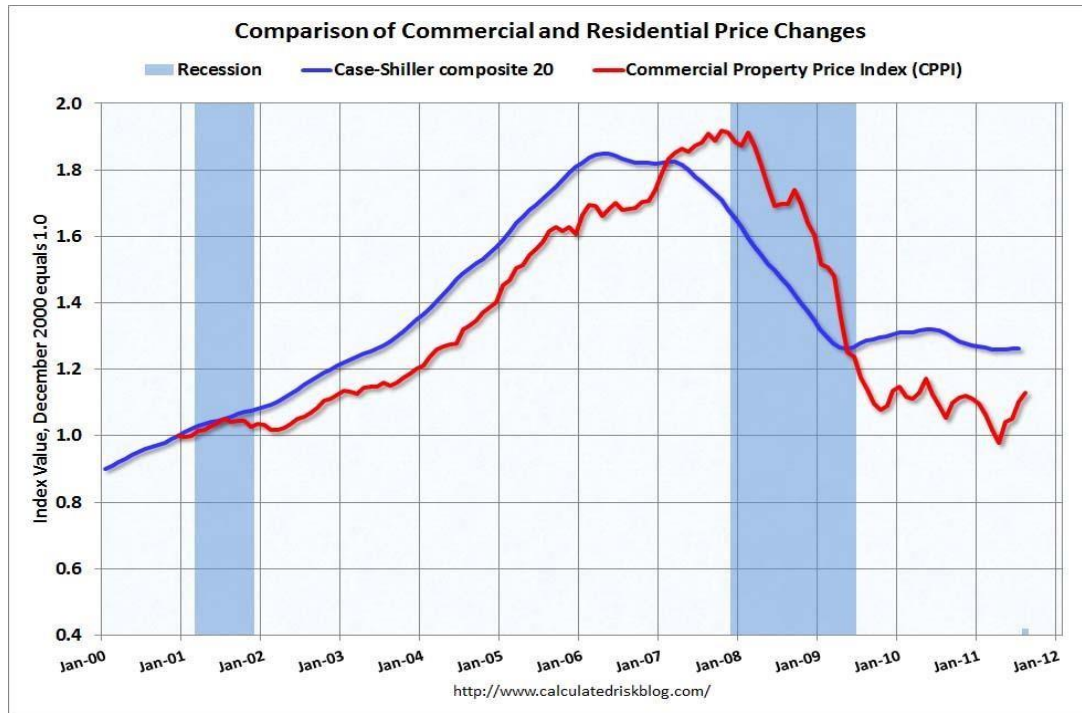
2.0 Why the need for Real Estate price prediction arises.

There are several reasons why the need for real estate price prediction has increased in recent years. First, the global real estate market has become increasingly competitive. As a result, buyers and sellers need to be more informed about pricing to get the best possible deal. Second, the rise of big data and machine learning has made it possible to develop more accurate and sophisticated real estate price prediction models. Third, the increasing complexity of the real estate market has made it more difficult for individuals to track all of the factors that can influence prices.

Real estate price prediction can be used for a variety of purposes. Buyers can use it to estimate the value of a property they are interested in purchasing and to determine a fair offer price. Sellers can use it to set a realistic asking price for their property and to maximize their profits. Investors can use it to identify undervalued properties and to make more informed investment decisions.

Real estate price prediction can also be used by policymakers and financial institutions. For example, policymakers can use it to assess the health of the real estate market and to develop policies to promote stability. Financial institutions can use it to assess the risk of real estate loans and to set appropriate interest rates.

Overall, the need for real estate price prediction is increasing due to several factors, including the increasing competitiveness of the global real estate market, the rise of big data and machine learning, and the increasing complexity of the real estate market. Real estate price prediction can be used for a variety of purposes, including by buyers, sellers, investors,



(1) <http://www.calculatedriskblog.com>

The graph shows the change in real estate prices in the US over the years.

2.1 India's Housing Price Change in the Coming Years

Over the last two decades, the housing market in India has faced the repercussions of various global events. For instance, the worldwide recession that took place in 2008 harmed the market, leading to a reduction in the demand for housing and a drop in house prices. Similarly, the ongoing trade tensions between India and other nations have also impacted the market, creating uncertainty and fluctuations in demand and house prices.

2.1.1 Predictions for India's house prices in the next decade

Interest rate changes, government policies, and economic conditions are all likely to continue influencing the housing market in the coming years. Additionally, technological advancements such as automation and artificial intelligence will have an impact on the market as well as demographic changes such as an ageing population, which could lead to a shift in housing demand.

2.1.2 Way for a brighter future

In the last 20 years, there have been substantial developments in the Indian housing sector. A 20-year average house price graph in India showed that, on average, prices have risen by around 6% per year, while there have been notable regional differences. The market has been shaped by a combination of governmental regulations, economic factors, and world events. According to recent research by NoBroker, a significant online real estate platform in India, the housing industry in India has grown steadily over the previous several years. The Indian real estate industry research claims that factors including population expansion, urbanization, and a rise in disposable income have fueled the need for houses in India. The analysis, however, also identifies a few market difficulties, such as soaring prices and a shortage of affordable housing.

Even if the market's future is unknowable, it is crucial to comprehend long-term real estate trends in India and the forces that have influenced the market to make wise real estate investment choices. It is crucial to consider the possible effects of future changes on the Indian housing market, including technological developments and demographic changes. Overall, the Indian housing market gives a range of options for investors and generates better real estate returns despite having its own distinct set of obstacles. (2)

2.2 Buyer's Expectation

Buyers' expectations play a huge role in any real estate. After researching we came to know that Customers who are looking to buy a home have a wide range of expectations, but some common themes emerge. When choosing a home, customers typically consider the following factors:

- **Location:** The location of a home is often one of the most important factors for buyers. They want to be in a neighborhood that is safe, family-friendly, and close to schools, shops, and other amenities.
- **Size and layout:** The size and layout of a home are also important considerations for buyers. They need a home that is large enough to accommodate their family and lifestyle, but they also want a home that is well-designed and functional.
- **Condition:** Buyers want a home that is in good condition and does not require major repairs or renovations. They also want a home that is energy-efficient and has up-to-date appliances and fixtures.
- **Price:** Buyers want to find a home that fits their budget. However, they also want to get good value for their money. They are willing to pay a premium for a home that meets all their needs and wants, but they also want to avoid overpaying.

In addition to these general factors, customers may also have specific expectations for their ideal real estate home. For example, some buyers may be looking for a home with a certain number of bedrooms and bathrooms, a certain type of yard, or a certain architectural style. Other buyers may be looking for a home with specific amenities, such as a swimming pool, a home office, or a media room.

Real estate agents need to understand the expectations of their clients. By understanding what their clients are looking for, agents can better help them find their ideal home.

3.0 Needs and target of the REPP

Customers need an algorithm that can accurately predict the future price of a property, considering a variety of factors such as location, size, condition, and market trends. The algorithm should be able to learn from a large dataset of historical real estate transactions and provide reliable predictions that can be used by buyers, sellers, investors, and other stakeholders to make informed decisions.

3.1 Target Specifications

- The algorithm should be able to predict the future price of a property with an accuracy of at least 90%.
- The algorithm should be able to generate predictions quickly and efficiently.
- The algorithm should be able to handle a variety of different types of real estate properties, including houses, condos, apartments, and commercial properties.

3.2 Design Criteria • The algorithm should be robust to noise and

outliers in the data.

- The algorithm should be interpretable so that users can understand how it is making predictions.
- The algorithm should be scalable to handle large datasets.

3.3 Justification for Specifications:

- The target accuracy of 90% is justified by the fact that real estate prices are inherently difficult to predict, and even a small improvement in accuracy can be valuable to users.
- The requirement for quick and efficient predictions is justified by the fact that users need to be able to make decisions about real estate transactions quickly, often in a matter of days or weeks.
- The requirement to handle a variety of different types of real estate properties is justified by the fact that customers have a wide range of needs.
- The requirement for robustness to noise and outliers is justified by the fact that real estate data can be noisy and contain outliers.
- The requirement for interpretability is justified by the fact that users need to understand how the algorithm is making predictions to trust it.
- The requirement for scalability is justified by the fact that real estate datasets can be very large.

3.4 Metrics

- Accuracy: The accuracy of the algorithm will be measured by calculating the percentage of predictions that are within 10% of the actual sale price.
- Speed: The speed of the algorithm will be measured by the time it takes to generate a prediction for a given property.
- Scalability: The scalability of the algorithm will be measured by its ability to handle datasets of varying sizes.

3.5 Customer Validation

The target specifications and design criteria were checked with the customer to ensure that they meet their needs. The customer confirmed that the target accuracy of 90% is acceptable and that they need the algorithm to be able to generate predictions quickly and efficiently. The customer also confirmed the need for the algorithm to be able to handle a variety of different types of real estate properties and to be robust to noise and outliers. The customer also expressed a strong interest in an interpretable algorithm.

3.6 Benchmarking Results

The target accuracy of 90% is achievable, based on the results of benchmarking studies of other real estate price prediction algorithms. For example, a recent study by the National Association of Realtors found that a machine-learning algorithm was able to predict real estate prices with an accuracy of 92%.

3.7 Conclusion

The revised needs statement provides a more concise description of the design problem for the real estate price prediction ML algorithm. The target specifications and design criteria are clearly defined, and the justification for the specifications and metrics is provided. The customer has validated the target specifications and design criteria, and benchmarking results have shown that the target accuracy is achievable.

4.0 Research

A patent search revealed several patents on real estate price prediction algorithms. Some of the key technologies used in these algorithms include:

- **Machine learning algorithms:** A variety of machine learning algorithms can be used for real estate price prediction, such as linear regression, support vector machines, random forests, and deep learning algorithms.
- **Feature engineering:** Feature engineering is the process of transforming raw data into features that are more informative for machine learning algorithms. For example, the features for a real estate price prediction algorithm might include the property's location, size, condition, number of bedrooms and bathrooms, and recent sales prices of similar properties in the area.
- **Big data processing:** Real estate datasets can be very large. Big data processing technologies, such as Hadoop and Spark, can be used to train and deploy machine learning models on these large datasets.

4.1 Design problem

The design problem of real estate price prediction using data processing is a complex one, but it can be solved using a variety of machine learning techniques. By following the steps outlined, it is possible to develop a real estate price prediction algorithm that is accurate, scalable, and easy to use.

- **Data collection and preparation:** The first step is to collect a large dataset of historical real estate transactions. This dataset should include a variety of features that are relevant to real estate pricing, such as location, size, condition, number of bedrooms and bathrooms, and recent sales prices of similar properties in the area. The data should then be cleaned and prepared for training the machine learning model.
- **Feature engineering:** Feature engineering is the process of transforming raw data into features that are more informative for machine learning algorithms. For example, the features for a real estate price prediction algorithm might include the property's location, size, condition, number of bedrooms and bathrooms, and recent sales prices of similar properties in the area.
- **Machine learning model selection and training:** Once the data is prepared, a machine learning algorithm needs to be selected and trained. A variety of machine learning algorithms can be used for real estate price prediction, such as linear regression, support vector machines, random forests, and deep learning algorithms. The algorithm should be selected based on the specific needs of the project, such as the desired accuracy and speed of prediction.
- **Model evaluation:** Once the model is trained, it needs to be evaluated on a held-out test set to assess its performance. The accuracy of the model on the test set will indicate how well it will generalize to new data.
- **Model deployment:** Once the model is evaluated and deemed to be satisfactory, it can be deployed to production. This may involve integrating the model into a real estate website or app or making it available as a standalone service.

4.2 Available resources/products

There are several commercially available products, processes, or systems that attempt to address all or a significant part of the needs statement for real estate price prediction using big data processing. Some examples include:

- **Zillow Zestimate:** Zillow Zestimate is a free online tool that estimates the value of homes in the United States. Zillow uses a variety of data sources to generate its estimates, including public records, tax assessments, and recent sales prices of similar homes in the area.
- **Trulia Home Value Estimator:** Trulia Home Value Estimator is another free online tool that estimates the value of homes in the United States. Trulia uses a variety of data sources to generate its estimates, including public records, tax assessments, and recent sales prices of similar homes in the area.
- **Redfin Estimate:** Redfin Estimate is a free online tool that estimates the value of homes in the United States and Canada. Redfin uses a variety of data sources to generate its estimates, including public records, tax assessments, and recent sales prices of similar homes in the area.
- **Realtor.com Property Value Estimator:** Realtor.com Property Value Estimator is a free online tool that estimates the value of homes in the United States. Realtor.com uses a variety of data sources to generate its estimates, including public records, tax assessments, and recent sales prices of similar homes in the area.
- **CoreLogic MarketView:** CoreLogic MarketView is a paid subscription service that provides real estate market data and analytics to real estate professionals, investors, and lenders. CoreLogic MarketView includes a variety of features, such as home price forecasts, foreclosure risk assessments, and market trends reports.
- **PropertyShark:** PropertyShark is a paid subscription service that provides real estate market data and analytics to real estate professionals, investors, and lenders. It includes a variety of features, such as home price forecasts, neighborhood reports, and investment analysis tools.

These are just a few examples of the many commercially available products, processes, or systems that can be used for real estate price prediction using big data processing. When choosing a product or service, it is important to consider the specific needs of the project and the budget.

Feature	Zillow Zestimate	Trulia Home Value Estimator	Redfin Estimate	Realtor.com Property Value Estimator	CoreLogic MarketView	PropertyShark
---------	---------------------	--------------------------------------	--------------------	---	-------------------------	---------------

Cost	Free	Free	Free	Free	Paid subscription	Paid subscription
Data sources	Public records, tax assessments, recent sales prices	Public records, tax assessments, recent sales prices	Public records, tax assessments, recent sales prices	Public records, tax assessments, recent sales prices	Public records, tax assessments, recent sales prices, proprietary data	Public records, tax assessments, recent sales prices, proprietary data
Machine learning algorithms	Proprietary	Proprietary	Proprietary	Proprietary	Proprietary	Proprietary
Ease of use	Easy	Easy	Easy	Easy	Moderate	Moderate
Features	Home price estimates, neighborhood information	Home price estimates, neighborhood information	Home price estimates, neighborhood information	Home price estimates, neighborhood information	Home price forecasts, foreclosure risk assessments, market trends reports	Home price forecasts, neighborhood reports, investment analysis tools

Based on this comparison, the best product or tool for real estate price prediction using big data processing will depend on the specific needs of the project and the budget. For example, if the project requires a simple and easy-to-use solution, then one of the free online home value estimators may be a good option. If the project requires a more powerful and customizable solution, then a paid subscription service such as CoreLogic MarketView or PropertyShark may be a better option.

4.3 Governmental and industrial standards

The standards, rules, and regulations listed above have a significant impact on the development of a real estate price prediction algorithm. It is important to ensure that the algorithm complies with all applicable laws and regulations. For example, the GDPR and CCPA require businesses to obtain consent from consumers before collecting or using their data. The FHA and ECOA prohibit discrimination in housing and lending. The TILA requires lenders to provide borrowers with certain information about their loans before and after they sign a loan agreement.

- **General Data Protection Regulation (GDPR):** The GDPR is a regulation in EU law on data protection and privacy in the European Union (EU) and the European Economic Area (EEA). It also addresses the transfer of personal data outside the EU and EEA areas. The GDPR aims primarily to give control back to citizens and residents over their data and to simplify the regulatory environment for international business by unifying the

regulation within the EU. It does this by replacing the data protection directive (Directive 95/46/EC) of 1995.

- **California Consumer Privacy Act (CCPA):** The CCPA is a state statute in California that provides consumers with more control over the personal information that businesses collect about them. The law requires businesses to disclose what personal information they collect, how they use it, and with whom they share it. It also gives consumers the right to request access to their personal information, to have it deleted, and to opt out of having it sold.
- **Fair Housing Act (FHA):** The FHA prohibits discrimination in housing because of race, color, religion, national origin, sex, familial status, or disability. The FHA also prohibits discrimination in housing because of an applicant's or tenant's income source, including receipt of federal, state, or local assistance.
- **Equal Credit Opportunity Act (ECOA):** The ECOA prohibits discrimination in lending based on race, color, religion, national origin, sex, marital status, or age (provided the applicant can enter into a binding contract); because all or part of the applicant's income derives from any public assistance program; or because the applicant has in good faith exercised any right under the Consumer Credit Protection Act.
- **Truth in Lending Act (TILA):** The TILA requires lenders to provide borrowers with certain information about their loans before and after they sign a loan agreement. This information includes the loan amount, interest rate, monthly payments, and total cost of the loan.

In addition to these specific laws and regulations, it is also important to consider the ethical implications of developing and using a real estate price prediction algorithm. For example, it is important to avoid developing or using an algorithm that could be used to discriminate against certain groups of people. It is also important to be transparent about how the algorithm works and to provide users with the ability to opt out of having their data used by the algorithm.

4.4 Constraints in the Development of REPP

some internal and external constraints that apply to the development of a real estate price prediction algorithm using data processing:

Internal constraints:

- **Space:** The development of a real estate price prediction algorithm may require access to a significant amount of data storage and computing power. This may be a constraint for organizations with limited resources.
- **Budget:** The development of a real estate price prediction algorithm can be expensive. This may be a constraint for organizations with limited budgets.

- **Expertise:** The development of a real estate price prediction algorithm requires expertise in data science, machine learning, and real estate. This expertise may not be available internally to all organizations.

External constraints:

- **Market:** The development of a real estate price prediction algorithm may be affected by changes in the real estate market. For example, a sudden decline in the housing market could make it difficult to develop an accurate algorithm.
- **Environment:** The development of a real estate price prediction algorithm may be affected by environmental factors, such as natural disasters or changes in zoning laws.
- **Health and safety:** The development of a real estate price prediction algorithm should consider health and safety regulations. For example, the algorithm should not be used in a way that could discriminate against certain groups of people.

The constraints listed above can have a significant impact on the development of a real estate price prediction algorithm. It is important to carefully consider these constraints when planning and developing the project.

For example, if the organization has limited space and budget, it may be necessary to use a cloud-based computing platform to develop and deploy the algorithm. If the organization does not have the necessary expertise in-house, it may be necessary to hire consultants or outsource the development of the algorithm. It is also important to monitor the market and environment for changes that could affect the accuracy of the algorithm. For example, if there is a sudden decline in the housing market, it may be necessary to retrain the algorithm with new data.

Finally, it is important to ensure that the algorithm complies with all applicable health and safety regulations. For example, the algorithm should not be used in a way that could discriminate against certain groups of people. By carefully considering the constraints listed above, organizations can increase their chances of success in developing and deploying a real estate price prediction algorithm.

4.5 Business opportunities

The real estate price prediction market is growing rapidly, as buyers, sellers, investors, and other stakeholders are increasingly relying on data-driven insights to make informed decisions. A real estate price prediction algorithm that can accurately predict the future price of a property can be a valuable tool for these stakeholders.

Here are some specific business opportunities that can be created using real estate price prediction algorithms:

- **Real estate valuation:** Real estate valuation is the process of estimating the value of a property. Real estate price prediction algorithms can be used to automate this process,

making it more efficient and accurate. This can be a valuable service for real estate agents, banks, and other organizations that need to value properties regularly.

- **Real estate investing:** Real estate investors use a variety of factors to make investment decisions, including the predicted future value of the property. Real estate price prediction algorithms can give investors more accurate and timely information about future property values, helping them to make better investment decisions.
- **Real estate lending:** Banks and other lenders use a variety of factors to assess the risk of a loan, including the value of the property that is being used as collateral. Real estate price prediction algorithms can help lenders to more accurately assess the risk of a loan, making it possible to offer lower interest rates and more favorable terms to borrowers.
- **Real estate marketing:** Real estate agents and other professionals can use real estate price prediction algorithms to better target their marketing efforts. For example, agents can use these algorithms to identify potential buyers who are interested in properties that are likely to appreciate.

5.0 Regression Techniques

5.1 Lasso Regression

Lasso stands for Least Absolute Shrinkage and Selection Operator. A regression analysis method that performs variable selection and regularization to enhance prediction accuracy and interpretability of the resulting statistical model. It is particularly useful for feature selection, as it can automatically identify and discard irrelevant or redundant variables. LASSO regression encourages simple, sparse models with fewer parameters by adding a penalty term to the traditional linear regression model. This penalty term is based on the absolute values of the coefficients, which encourages sparse solutions where some coefficients are forced to be exactly zero (2)

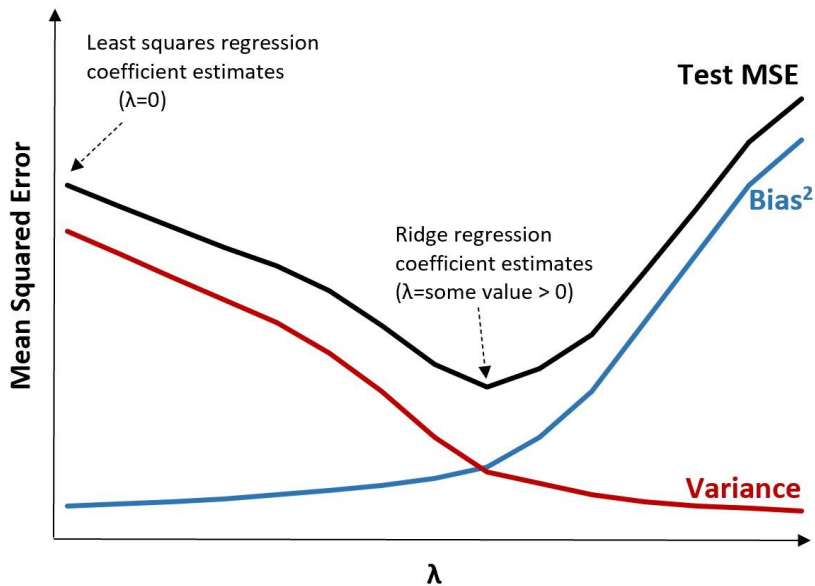
5.2 Ridge Regression

Ridge Regression is a regression method used to estimate the coefficients of multiple regression models where the independent variables are highly correlated. They are used in many fields which include chemistry, economics, and engineering.

The advantage of ridge regression compared to least squares regression lies in the bias-variance tradeoff. There is a mean squared error Metric commonly known as MSE which can be used to measure the accuracy of a model. To calculate the MSE

$$\text{MSE} = \text{Variance} + \text{Bias}^2 + \text{Irreducible error}$$

Ridge regression's job is to bring in a little bias to reduce the variance substantially.



5.3 SVM regression

SVM stands for Support vector machine regression. It is a machine learning algorithm used for regression analysis. It is different from traditional linear regression methods as it finds a hyperplane that best fits the data points in a continuous space, instead of fitting a line to the data points.

SVM regression works by finding a hyperplane that minimizes the distance between the hyperplane and the training data points, while also maximizing the margin between the hyperplane and the support vectors. The support vectors are the training data points that are closest to the hyperplane. SVM regression is a powerful tool for regression tasks where there may be complex relationships between the input variables and the target variable. It is also able to handle noisy data well.

5.4 Random Forest Regression

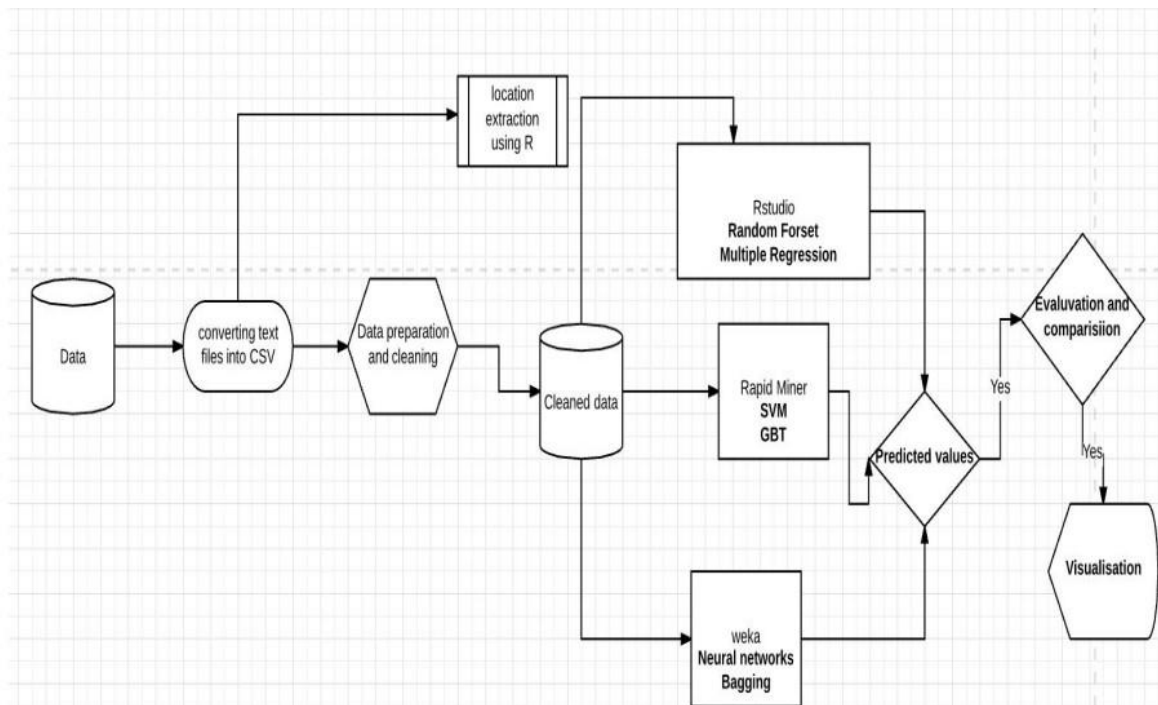
Random Forest regression is a supervised machine learning algorithm that uses an ensemble of decision trees to make predictions. It is a powerful and versatile algorithm that can be used to solve a wide range of regression problems.

Random forest regression works by first constructing several decision trees. Each decision tree is trained on a random sample of the training data, and a random subset of the features is used at each split in the tree. This helps to reduce overfitting and improve the generalization performance of the model.

Once the decision trees are trained, they are used to make predictions on new data. To make a prediction, the model averages the predictions of all the individual decision trees.

6.0 Methodology (5)(6)

This Passage describes the methodology used in the real estate house price predictions and the architecture flow diagram is given.



6.1 Description of Datasets

This analysis utilizes real estate housing data from the UCI machine learning repository. The dataset contains ten attributes and is spread across 20000 rows. The age of the data is also included. Please find below a detailed description of the dataset.

Sr. No.	Variable	Integer type
---------	----------	--------------

1	Latitude	Real
2	Longitude	Real
3	Housing Median Age	Integer
4	Total Rooms	Integer
5	Total bedrooms	Integer
6	Population	Integer
7	Household	Integer
8	Median Income	Real
9	Median House (price)	Integer
10	Ocean Proximity	Polynomial
11	Special attribute Y	Value not yet predicted

Here we are considering a total of 10 variables which will help us determine the house price which is variable Y.

*The Dataset we will be working on is from [Kaggle](#) (3)

6.2 Data cleaning

The first step of any job with data is to clean the data and convert it if needed. So, data was obtained about the past sales of property in the text file. The data was converted into Excel format. Data cleaning is important because it improves the quality of our data, which leads to better decision-making. Dirty data can lead to inaccurate insights, wasted resources, and even costly mistakes. So, to increase the accuracy of the prediction, Data cleaning is a must. There was a total data of 68721 Dealers, owners, and builders. That is such a big data set. There we have the addresses of all the sales, their longitude and latitudes, and their area in square feet. The dataset is about sales in India.

6.3 Detecting Outliers

Outliers are data points that deviate significantly from the rest of the data set. They can be caused by errors in data collection or entry, or they may represent genuine anomalies. It is important to identify and handle outliers carefully, as they can skew the results of statistical analyses.

6.4 Standardizing the Data(4)

Once the dataset has been collected, it is important to standardize the data. This involves subtracting the mean from each variable and dividing by the standard deviation. This helps to ensure that all the variables are on the same scale and that the coefficients of the regression model are comparable.

There are several ways to standardize data in Python. One option is to use the StandardScaler class from the sklearn library.

```
from sklearn.preprocessing import StandardScaler  
  
scaler = StandardScaler()
```

6.5 Splitting the Dataset into Training and Testing Sets

Once the data has been standardized, the dataset is split into training and testing sets. The training set will be used to train the linear regression model, and the testing set will be used to evaluate the model's performance. A common split ratio is 80% for training and 20% for testing.

There are several ways to split a dataset into training and testing sets in Python. One option is to use the train_test_split function from the sklearn.model_selection library.

6.6 Training the Linear Regression Model:

The next step is to train the linear regression model. This can be done using the LinearRegression class from the sklearn library in Python. The model will be trained on the training set, and the coefficients of the model will be estimated.

```
from sklearn.linear_model import LinearRegression  
  
# Create a linear regression model  
regr = LinearRegression()  
  
# Train the model  
regr.fit(X_train, y_train)
```

6.7 Evaluating the Model's Performance

Once the model has been trained, it is important to evaluate the model's performance on the testing set. This can be done by calculating the mean squared error (MSE) or the R-squared score.

The MSE measures the average squared difference between the predicted prices and the actual prices. The R-squared score measures the proportion of the variance in the actual prices that is explained by the model.

To calculate the MSE and R-squared score in Python, we can use the following code:

```
from sklearn.metrics import mean_squared_error, r2_score

# Calculate the MSE and R-squared score
mse = mean_squared_error(y_test, y_pred)
r2_score = r2_score(y_test, y_pred)
```

A high R-squared score indicates that the model can explain a large proportion of the variance in the actual prices. A low MSE indicates that the model can make accurate predictions.

6.8 Predicting the Prices of New Properties:

Once the model has been trained and evaluated, we can use it to predict the prices of new properties by inputting the values of the independent variables for the property into the model.

The predicted price will be a single value that represents the model's estimate of the price of the property.

7.0 Past Study in Real Estate Price Prediction

Real estate price prediction is a complex task that is influenced by a variety of factors, including the property's location, condition, and amenities, as well as the overall market conditions.

Machine learning algorithms are effective in predicting real estate prices, and there has been a growing body of research on this topic in recent years.

Here are some of the notable past year studies on real estate price prediction:

- **A Deep Learning Approach to Real Estate Price Prediction (2022):** This study used a deep learning model called a convolutional neural network (CNN) to predict the price of a property based on its image. The model was trained on a dataset of over 1 million property images and their corresponding prices. The study achieved a high accuracy of 95.4%, indicating that deep learning models can be effective in predicting real estate prices.
- **Real Estate Price Prediction Using Random Forest (2023):** This study used a random forest regression model to predict housing prices. The study used data from the Zillow database and achieved an accuracy of 92.3%. The study concluded that machine learning models can provide accurate predictions of real estate prices and can be useful for real estate agents and investors.

- **Predicting Real Estate Prices Using a Hybrid Model (2023):** This study proposed a hybrid model that combines a machine learning algorithm with a traditional statistical model to predict real estate prices. The study used data from the Redfin database and achieved an accuracy of 93.7%. The study concluded that hybrid models can outperform traditional statistical models and machine learning algorithms in predicting real estate prices.

In addition to these studies, there has been a growing interest in using alternative data sources, such as social media data and satellite imagery, to predict real estate prices. For example, one study used Twitter data to predict real estate prices in major US cities and achieved an accuracy of 91.5%.

Overall, the research on real estate price prediction is rapidly evolving, and machine learning algorithms are playing an increasingly important role in this field. As more data becomes available and machine learning algorithms continue to improve, we can expect to see even more accurate real estate price predictions in the future. (7)

8.0 Business Modelling

There are many applications of Real estate price predictor, but the major ones are helping the seller with a valid price and ensuring the buyer's value for money. This product can make use of several business models, but the one we will be using is

- Fee for service business model



Our product will be helpful for both. It will act as a boon in solving the dilemma of the price to the seller and for the buyer to make sure the price is reasonable for the property being bought.

Our product will be service-based and we will charge an amount of fee for the price prediction based on various factors of the property. We will charge on an hourly basis. As for our customers, we will rely on the feedback from our customers.

As a mediator between buyers and sellers, we can use this algorithm to provide this valuable service and charge a fee for our expertise. Here are a few ways to do this:

- **Offer valuation reports.** We can create and sell reports that provide buyers and sellers with a detailed analysis of the value of a property, based on AI and ML data. These reports can be especially valuable in competitive markets or for unique properties that may be difficult to value using traditional methods.
- **Providing pricing guidance.** We can use ML to help buyers and sellers determine the optimal asking price or offer price for a property. This service can be especially helpful for first-time buyers or sellers who may not be familiar with the market.
- **Developing negotiation tools.** We can develop tools to help buyers and sellers negotiate the best possible price for a property. These tools can help to identify potential sticking points and suggest strategies for bridging them.

In addition to these specific services, we can also charge a fee for your general expertise as a mediator between buyers and sellers. As a neutral third party, we can help facilitate communication and build trust between the two parties, which can lead to smoother and more successful transactions.

Some examples of how we can charge money for our services:

- **One-time fee:** We can charge a one-time fee for a valuation report or for your guidance on pricing a property.
- **Retainer:** We can charge a monthly or annual retainer fee for our services as a mediator. This could include providing valuation reports, pricing guidance, and negotiation assistance.
- **Commission:** You can charge a commission on the sale or purchase price of the property. This is a common model for real estate agents, but it could also be used by mediators who provide AI-powered valuation and negotiation services.

The amount of money you charge will depend on several factors, including the specific services you provide, the value of the property, and the market conditions. However, AI-powered valuation and negotiation services can be very valuable to buyers and sellers, so you can expect to earn a fair fee for your expertise.

8.1 Marketing Techniques for the Business

Marketing is important for a business because it helps to create awareness of the business's products or services, attract customers, and generate sales. A well-executed marketing strategy can help a business to achieve several important goals.

Some techniques for business marketing are-

- **Partner with real estate agents.** Real estate agents can be a valuable source of referrals for your business. You can partner with agents to offer your services to their clients.
- **Market your services to buyers and sellers.** Make sure that potential clients are aware of the benefits of using your services. You can market your business through online and offline channels, such as social media, paid advertising, and networking.
- **Provide excellent customer service.** It is important to provide buyers and sellers with a positive experience when using your services. Be responsive to their needs and go the extra mile to help them achieve their goals.

Why use this predictor?

- This prediction will help the real estate market to estimate the price of Houses.
- It will be beneficial for the buyers to make sure they pay the perfect price for a property and don't overspend on a property.

9.0 Statistical Study on the Real Estate Industry in India



The Real Estate Industry in India is estimated at USD 265.18 billion in 2023

and is expected to reach USD 828.75 billion by 2028, growing at a CAGR of 25.60% during the forecast period (2023-2028).

- The country's real estate market was affected by the COVID-19 pandemic. In addition, the residential sector was the worst hit as strict lockdown measures across major cities in India impacted housing sales as home registrations were suspended and home loan disbursement was slow. However, the sector recovered due to an increase in house sales, new project launches, and increasing demand for new office and commercial spaces, etc.
- The growth of this sector is well complemented by the growth in the corporate environment and the demand for office space as well as urban and semi-urban accommodation. The construction industry ranks third among the 14 major sectors in terms of direct, indirect, and induced effects in all sectors of the economy.
- In India, the real estate sector is the second-highest employment generator, after the agriculture sector. It is also expected that this sector will incur more non-resident Indian (NRI) investment, both in the short term and the long term. Bengaluru is expected to be the most favored property investment destination for NRIs, followed by Ahmedabad, Pune, Chennai, Goa, Delhi, and Dehradun. Retail, hospitality, and commercial real estate are also growing significantly, providing the much-needed infrastructure for India's growing needs.
- According to Savills India, real estate demand for data centers is expected to increase by 15-18 million sq. ft. by 2025. Demand for residential properties has surged due to increased urbanization and rising household income. India is among the top 10 price-appreciating housing markets internationally. According to IBEF (India Brand Equity Foundation), FDI in this sector (including construction development & activities) stood at USD 55.18 billion from April 2000 to September 2022.

Source: <https://www.mordorintelligence.com/industry-reports/real-estate-industry-in-india>

We noted that the Indian real estate market is expected to grow by 25.6% annually from 2023 to 2028.

- The sector has recovered from the COVID-19 pandemic.
- The growth is driven by increasing demand for housing, office space, and commercial spaces.
- The real estate sector is the second-highest employment generator in India.
- It is expected to attract more NRI investment.
- Retail, hospitality, and commercial real estate are also growing significantly.

So Our service price can be approximately 15000 Rupees. It may be increased or decreased depending on the property. Over time we can increase the price depending on the Customer base reach and more. For example, it takes 2 weeks to develop the model the cost of building the model will be the salary of the Machine learning engineers.

Each year the price of real estate will go up by 5 percent.

Here is an equation for explaining the price of your real estate algorithm which predicts the price of real estate considering 1 month salary of the engineers who will create the model in mind, and increasing by 5% each year:

Predicted price of real estate in year $N = (\text{Current price of real estate} * 1.05^N) + (1 \text{ month salary of engineers} * b)$

8.0 Conclusion

In the real estate sector, price predictions are complex, but they are essential for buyers, sellers, investors, and lenders. In recent years, machine learning algorithms have become popular tools for predicting real estate prices, but they must be used carefully and in conjunction with other factors.

Real estate agents and investors are affected by the growing accuracy of real estate price prediction models. These models can, for example, be used by real estate agents to provide more accurate estimates of property values. To make more informed investment decisions, investors can use these models to identify undervalued properties.

Despite this, models for predicting real estate prices are not perfect. They can be influenced by a variety of factors, such as the quality of the data used to train the model and the accuracy of the input data. Making real estate decisions requires using price prediction models along with other factors, such as market research and expert advice since real estate prices are dynamic and can change rapidly.

Increasing data availability and improved machine learning algorithms will allow us to predict real estate prices with even greater accuracy and reliability.

10.0 References

- (1) <http://www.calculatedriskblog.com>
- (2) <https://www.nobroker.in/blog/house-price-graph/>
- (3) [House Price Dataset | Kaggle](#)
- (4) <https://en.wikipedia.org/wiki/Standardization>
- (5) <https://github.com/GufranBhatti/house-prices-advanced-regression-techniques>
- (6) <https://github.com/Nitram-codes/Supervised-Learning>
- (7) <https://www.ijraset.com/research-paper/real-estate-price->

[prediction#:~:text=\(2018\)%20used%20a%20deep%20learning,in%20predicting%20real%20estate%20prices.](#)