**HOUSE PRICE PREDICTION**

**PROJECT BASED LEARNING IV(AIP104)**

**PROJECT REPORT**



**Submitted by**

**Kangna Galhotra (2110993804)**

**Jatin Verma (2110993803)**

**BE-CSE (Artificial Intelligence)**

**Guided by**

**Dr. Kamal Deep Garg**

**CHITKARA UNIVESITY OF ENGINEERING & TECHNOLOGY**

**CHITKARA UNIVERSITY, RAJPURA**

**MAY, 2023**

**ACKNOWLEDGEMENT**

With immense please we, Mr. Jatin Verma and Ms. Kangna Galhotra presenting “House Price Prediction Machine Learning Model” project report as part of the curriculum of ‘BE - CSE (AI)’.

We would like to express our sincere thanks to Dr. Kamal Deep Garg, for her valuable guidance and support in completing our project.

We would also like to express our gratitude towards our dean Dr. Sushil Kumar Narang for giving us this great opportunity to do a project on “House Price Prediction”. Without their support and suggestions, this project would not have been complete.

Signature…..………….

Name: Kangna Galhotra

Roll No.: 2110993804

Signature…..………….

Name: Jatin Verma

Roll No.: 2110993803

**ABSTRACT**

House price prediction is a crucial task in real estate markets, enabling stakeholders to make informed decisions regarding buying, selling, or investing in properties. In this project, we explore the application of machine learning algorithms for accurate house price prediction. We investigate the performance of regression algorithms, including Elastic Net, SVR, Random Forest Regressor, and Gradient Boosting Regressor, on a given dataset. The dataset we have used for training and testing in our model have more than 50 parameters some of them are **MSZoning**: The general zoning classification, **LotFrontage**: Linear feet of street connected to property, **LotArea**: Lot size in square feet, **Street**: Type of road access, **Alley**: Type of alley access, **LotShape**: General shape of property, **LandContour**: Flatness of the property, etc. Next comes the part of data cleaning where we have tried to identify and assess the presence of missing values in the dataset. Decided upon the appropriate strategy to handle missing values, such as: Removing rows or columns with a significant number of missing values. Imputing missing values using techniques like mean, median, or regression imputation.Creating a separate indicator variable to capture missingness. We have dealt with the outliers as follows: Identified outliers in the numerical features using techniques like box plots, histograms, or statistical methods. Decided upon the appropriate approach to handle outliers, such as: Removing extreme outliers if they are data entry errors. Transforming the data using techniques like log transformation to reduce the impact of outliers. Ordinal encoding for categorical variables with a natural ordering. Standardization (subtracting the mean and dividing by the standard deviation). Normalization (scaling the values to a specific range, like [0, 1]). Apply appropriate transformations like log transformation or Box-Cox transformation to reduce skewness if required. Remove any redundant or irrelevant features that may not contribute significantly to the prediction task. Based on the accuracies obtained from our different models as follows: ElasticNet:(Mean RMSE: 32048.056, Mean R-Squared: 0.824), SVR (Support Vector Regression):(Mean RMSE: 80401.664, Mean R-Squared: -0.057),RandomForestRegressor:(Mean RMSE: 30266.129, MeanR-Squared: 0.823),GradientBoostingRegressor: (Mean RMSE: 27978.372, Mean R-Squared: 0.824) we observe that ElasticNet and GradientBoostingRegressor show similar performance in terms of Mean RMSE and Mean R-Squared, with GradientBoostingRegressor having a slightly lower RMSE. RandomForestRegressor also performs well, with a similar RMSE to ElasticNet and GradientBoostingRegressor but a slightly lower R-Squared value. Overall, the choice of the regression algorithm depends on the specific dataset, problem requirements, and desired trade-offs between interpretability and predictive accuracy. It is crucial to perform thorough evaluation and fine-tuning of the chosen algorithm, including hyperparameter optimization, to ensure optimal performance. By employing machine learning techniques, we can harness the power of data to make more informed decisions in the real estate market and facilitate better outcomes for buyers, sellers, and investors.

**TABLE OF CONTENTS**

Acknowledgement.....................................................................................................................…………………ii

Abstarct.............................................................................................................................................…........iii

Table Of Contents……………………………………………………………………………………………………………………………………iv

List Of Figures………………………………………………………………………………………………………………………………………….v

Chapter 1 : Introduction………………………………………………………………………………………………………………………….6

* 1. Brief of project…………………………………………………………………………………………………………………….7
  2. Data Description…………………………………………………………………………………………………………………..8
  3. ML Approaches used………………………………………………………………………………………………………..…10

Chapter 2: Problem Formulation……………………………………………………………………………………………………………15

Chapter 3: Methodology………………………………………………………………………………………………………………………..15

3.1 Data Preprocessing……………………………………………………………………………………………………………..16

3.2 Feature Selection………………………………………………………………………………………………………………..16

3.3 Splitting data………………………………………………………………………………………………………………………17

3.4 Model Selection and training………………………………………………………………………………………………18

3.5 Model Evaluation………………………………………………………………………………………………………………18

3.6 Hyperparameter tuning………………………………………………………………18

3.7 Final Model Training………………………………………………………………………………………………………18

3.8 Model Interpretation…………………………………………………………………………………………………………19

3.9 Deployment and Monitoring…………………………………………………………………………………………….19

Chapter 4: Flowchart…………………………………………………………………………………………………………………………..23

Chapter 5: Conclusion…………………………………………………………………………………………………………………………24

Chapter 6 : References……………………………………………………………………………………………………………………….25

**LIST OF FIGURES :**

Figure 1: L1 & L2 Regularization………………………………………………………………10

Figure 2: SVR Objective………………………………………………………………………...11

Figure 3: SVR Flowchart………………………………………………………………………..12

Figure 4: Gradient Boosting……………………………………………………………………..14

Figure 5: Data Pre Processing……………………………………………………………………16

Figure 6: Feature Selection………………………………………………………………………17

Figure 7:Model Evaluation………………………………………………………………………18

Figure 8:Final Model Selection………………………………………………………………….19

Figure 9:Model Interpretation……………………………………………………………………19

Figure 10:Model Preparation…………………………………………………………………….20

Figure 11:SVR Implementation…………………………………………………………………21

Figure 12:Flowchart……………………………………………………………………………23

**INTRODUCTION**

House price prediction refers to the estimation of the future value or selling price of residential properties in the real estate market. It involves utilizing various factors and data points to forecast the potential price at which a house is likely to be sold. This prediction is valuable for buyers, sellers, investors, and real estate professionals who are involved in property transactions.

The significance of house price prediction in real estate markets can be understood from several perspectives:

1. Informed Buying and Selling: For prospective buyers, accurate price predictions enable them to make informed decisions about the affordability and value of a property. Sellers can set competitive prices based on market trends and predictions, attracting potential buyers and facilitating faster transactions.
2. Investment Decision Making: House price prediction is crucial for investors who aim to generate returns through real estate. By accurately predicting price trends, investors can identify lucrative investment opportunities, time their buying and selling decisions, and maximize their profits.
3. Market Analysis and Planning: House price predictions play a vital role in market analysis and planning for real estate professionals, developers, and policymakers. They provide insights into market trends, demand-supply dynamics, and potential risks. This information assists in formulating strategies, identifying target markets, and making informed decisions about development projects.
4. Risk Management: House price prediction helps individuals and organizations manage risks associated with real estate investments. By assessing the potential appreciation or depreciation of property values, investors can make informed decisions about diversifying their portfolios, managing cash flows, and mitigating potential losses.
5. Housing Policy and Regulation: Governments and policymakers rely on accurate house price predictions to develop effective housing policies. By understanding the dynamics of housing markets, policymakers can address issues such as housing affordability, market stability, and the prevention of speculative bubbles.

Overall, house price prediction is of significant importance in real estate markets as it provides valuable insights to buyers, sellers, investors, real estate professionals, and policymakers. It aids in making informed decisions, maximizing investment returns, managing risks, planning market activities, and formulating appropriate housing policies. Accurate predictions contribute to the efficiency, transparency, and stability of the real estate market, benefiting all stakeholders involved.

Accurately predicting house prices is a challenging task due to various factors and complexities involved in the real estate market. Some of the key challenges associated with house price prediction include:

1. Data Availability and Quality: Acquiring comprehensive and reliable data is crucial for accurate predictions. However, obtaining complete and up-to-date data on housing sales, property characteristics, economic indicators, and market trends can be challenging. Incomplete or inconsistent data can lead to biased or inaccurate predictions.
2. Complexity of Factors: House prices are influenced by a multitude of factors, including location, property size, amenities, market conditions, economic indicators, demographic trends, and regulatory policies. Analyzing the complex interactions and dependencies between these factors poses a challenge. Identifying and incorporating the most relevant features and variables into predictive models is essential but can be subjective and require domain expertise.
3. Nonlinear Relationships: The relationships between house price and its predictors are often nonlinear. Linear regression models may not capture the complex patterns and interactions present in the data. Advanced machine learning algorithms capable of handling nonlinear relationships, such as decision trees, random forests, and neural networks, may be required to improve prediction accuracy.
4. Overfitting and Model Complexity: Overfitting occurs when a predictive model performs well on the training data but fails to generalize well to new, unseen data. Complex models with excessive parameters can be prone to overfitting, leading to poor predictive performance. Striking a balance between model complexity and generalizability is crucial.

Addressing these challenges requires advanced analytical techniques, careful data selection, feature engineering, and appropriate model validation. It also emphasizes the need for continuous monitoring, updating models, and incorporating new data to adapt to changing market conditions and improve prediction accuracy.

**DataSet features or hyperparameters:**

**Data fields**

Here's a brief version of what you'll find in the data description file. The dataset is taken the dataset from kaggle.

* ****SalePrice**** - the property's sale price in dollars. This is the target variable that you're trying to predict.
* ****MSSubClass****: The building class
* ****MSZoning****: The general zoning classification
* ****LotFrontage****: Linear feet of street connected to property
* ****LotArea****: Lot size in square feet
* ****Street****: Type of road access
* ****Alley****: Type of alley access
* ****LotShape****: General shape of property
* ****LandContour****: Flatness of the property
* ****Utilities****: Type of utilities available
* ****LotConfig****: Lot configuration
* ****LandSlope****: Slope of property
* ****Neighborhood****: Physical locations within Ames city limits
* ****Condition1****: Proximity to main road or railroad
* ****Condition2****: Proximity to main road or railroad (if a second is present)
* ****BldgType****: Type of dwelling
* ****HouseStyle****: Style of dwelling
* ****OverallQual****: Overall material and finish quality
* ****OverallCond****: Overall condition rating
* ****YearBuilt****: Original construction date
* ****YearRemodAdd****: Remodel date
* ****RoofStyle****: Type of roof
* ****RoofMatl****: Roof material
* ****Exterior1st****: Exterior covering on house
* ****Exterior2nd****: Exterior covering on house (if more than one material)
* ****MasVnrType****: Masonry veneer type
* ****MasVnrArea****: Masonry veneer area in square feet
* ****ExterQual****: Exterior material quality
* ****ExterCond****: Present condition of the material on the exterior
* ****Foundation****: Type of foundation
* ****BsmtQual****: Height of the basement
* ****BsmtCond****: General condition of the basement
* ****BsmtExposure****: Walkout or garden level basement walls
* ****BsmtFinType1****: Quality of basement finished area
* ****BsmtFinSF1****: Type 1 finished square feet
* ****BsmtFinType2****: Quality of second finished area (if present)
* ****BsmtFinSF2****: Type 2 finished square feet
* ****BsmtUnfSF****: Unfinished square feet of basement area
* ****TotalBsmtSF****: Total square feet of basement area
* ****Heating****: Type of heating
* ****HeatingQC****: Heating quality and condition
* ****CentralAir****: Central air conditioning
* ****Electrical****: Electrical system
* ****1stFlrSF****: First Floor square feet
* ****2ndFlrSF****: Second floor square feet
* ****LowQualFinSF****: Low quality finished square feet (all floors)
* ****GrLivArea****: Above grade (ground) living area square feet
* ****BsmtFullBath****: Basement full bathrooms
* ****BsmtHalfBath****: Basement half bathrooms
* ****FullBath****: Full bathrooms above grade
* ****HalfBath****: Half baths above grade
* ****Bedroom****: Number of bedrooms above basement level
* ****Kitchen****: Number of kitchens
* ****KitchenQual****: Kitchen quality
* ****TotRmsAbvGrd****: Total rooms above grade (does not include bathrooms)
* ****Functional****: Home functionality rating
* ****Fireplaces****: Number of fireplaces
* ****FireplaceQu****: Fireplace quality
* ****GarageType****: Garage location
* ****GarageYrBlt****: Year garage was built
* ****GarageFinish****: Interior finish of the garage
* ****GarageCars****: Size of garage in car capacity
* ****GarageArea****: Size of garage in square feet
* ****GarageQual****: Garage quality
* ****GarageCond****: Garage condition
* ****PavedDrive****: Paved driveway
* ****WoodDeckSF****: Wood deck area in square feet
* ****OpenPorchSF****: Open porch area in square feet
* ****EnclosedPorch****: Enclosed porch area in square feet
* ****3SsnPorch****: Three season porch area in square feet

**Machine learning approaches used in this project**

Machine learning algorithms have become more popular in house price prediction thanks to technological improvements and the availability of enormous volumes of data. Machine learning algorithms can analyze huge amounts of data, spot intricate patterns, and make precise predictions. Here are several well-liked machine learning methods for forecasting house prices:

1. Elastic Net : Elastic Net is a regularization technique used in linear regression and machine learning models to address the limitations of individual regularization methods such as L1 (Lasso) and L2 (Ridge) regularization. It combines both L1 and L2 regularization penalties to overcome their individual shortcomings. In linear regression, the goal is to fit a linear equation that best represents the relationship between the input variables (features) and the target variable (house price, in this case). However, when dealing with high-dimensional datasets or a large number of features, traditional linear regression models may suffer from issues like overfitting or multicollinearity (highly correlated features). Elastic Net formula combines the L1 (Lasso) and L2 (Ridge) regularization penalties in linear regression models. The objective function of Elastic Net is given by:

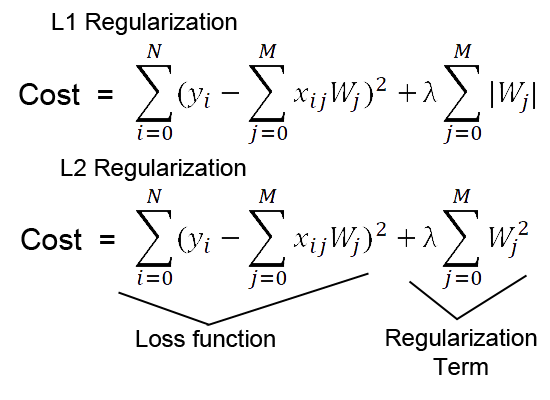


Figure 1: L1 & L2 Regularization

The benefits of using Elastic Net for house price prediction include:

* Handling Multicollinearity: Multicollinearity refers to high correlation among predictor variables. Elastic Net can handle multicollinearity effectively by distributing the importance among correlated features. It allows the model to select one or a combination of correlated features based on their individual contributions, reducing the impact of redundancy and improving prediction accuracy.
* Flexibility: Elastic Net provides flexibility in controlling the balance between L1 and L2 penalties. The hyperparameter l1\_ratio determines this balance, allowing the user to adjust the regularization approach. This flexibility enables customization of the model based on the specific requirements of the house price prediction task and the dataset.
* Feature Selection: Elastic Net performs automatic feature selection by shrinking the coefficients of irrelevant features to zero. This helps identify the most relevant predictors for house price prediction, improving model interpretability and reducing the risk of overfitting. By selecting only the essential features, Elastic Net simplifies the model and reduces the chances of including noise or irrelevant information.
* Better Performance in High-Dimensional Data: When dealing with datasets with a large number of features (high-dimensional data), Elastic Net outperforms traditional linear regression models. It effectively reduces the risk of overfitting and deals with the curse of dimensionality by selecting relevant features and controlling their magnitudes.

1. SVR : SVR (Support Vector Regression) is a supervised learning algorithm that is widely used for regression tasks, including house price prediction. SVR is a variant of Support Vector Machines (SVM) specifically designed for regression problems. It aims to find a hyperplane that best fits the training data while minimizing the margin violations.

In SVR, the objective is to find a function that approximates the mapping between the input features (predictors) and the target variable (house price). Instead of aiming for exact predictions, SVR allows for a tolerance or margin of error around the predicted values. The key idea is to find a hyperplane that maximizes the margin around this tolerance, within which most of the training data points lie.

The objective function of Support Vector Regression (SVR) aims to find a hyperplane that best fits the training data while minimizing the margin violations within a tolerance or epsilon tube. The objective function for SVR can be represented as follows:

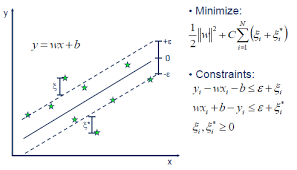


Figure 2: SVR Objective

The benefits of using SVR for house price prediction include:

* Nonlinear Relationships: SVR can effectively capture nonlinear relationships between the predictors and the house prices by using different kernel functions. It can handle complex and non-linear patterns in the data, making it suitable for modeling real estate markets where various factors influence house prices.
* Robustness to Outliers: SVR is robust to outliers, as it aims to minimize errors within the epsilon tube. Outliers that fall outside the tube have a limited influence on the model, reducing their impact on the overall prediction.
* Flexibility: SVR allows for customization through kernel functions, enabling the modeling of various types of relationships. The choice of kernel function and its parameters can be tuned to best fit the specific characteristics of the dataset.
* Regularization: SVR incorporates a regularization parameter (C) to balance the trade-off between fitting the training data and generalization. This regularization helps control overfitting and improves the model's ability to predict unseen data accurately.
* Support Vectors: The use of support vectors allows SVR to focus on a subset of the training data, reducing memory requirements and computational complexity compared to other algorithms.

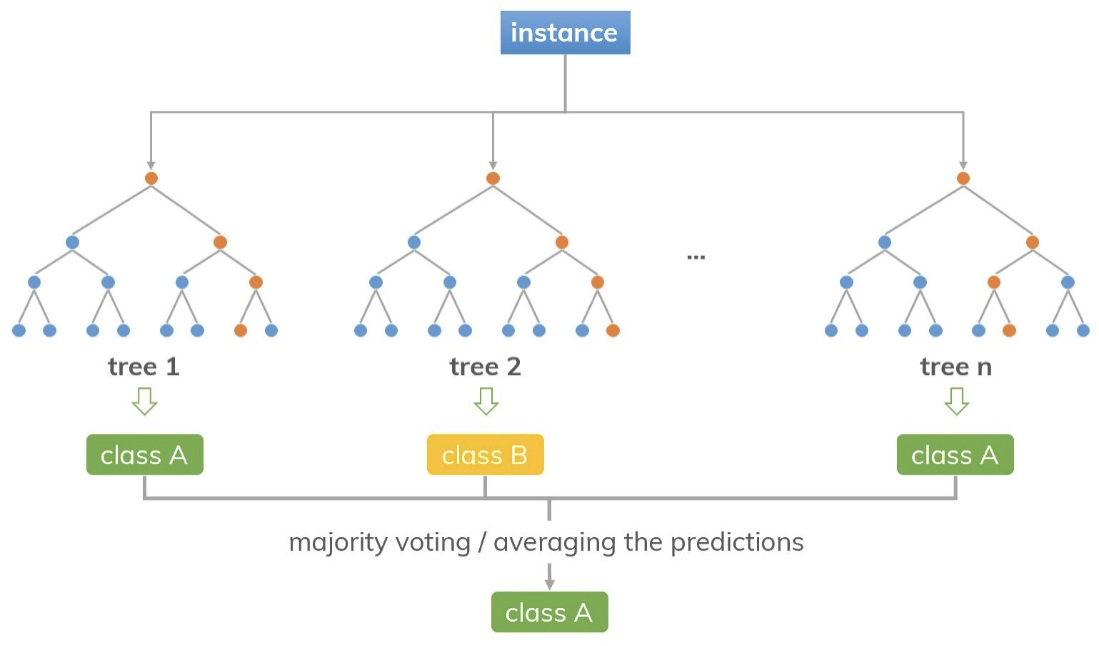


Figure 3: SVR Flowchart

The Random Forest Regressor offers several benefits for house price prediction:

* Robustness to Overfitting: The ensemble nature of random forests reduces overfitting compared to individual decision trees. By aggregating multiple trees, random forests can handle complex relationships in the data and make more accurate predictions.
* Feature Importance: Random forests provide a measure of feature importance, indicating the relative contribution of each predictor variable in the prediction process. This information can be useful for feature selection and understanding the factors that affect house prices.
* Handling Nonlinear Relationships: Random forests can capture nonlinear relationships between predictors and house prices, as they consider interactions between features and create complex decision boundaries.
* Robustness to Outliers: Random forests are robust to outliers, as they make predictions based on the consensus of multiple trees. Outliers have less impact on the overall prediction compared to individual decision trees.
* Handling Missing Data: Random forests can handle missing data without imputation. They can make predictions using the available features and provide reliable results even when some predictors have missing values.
* Scalability: Random forests can efficiently handle large datasets with a high number of features, making them suitable for house price prediction tasks that involve complex datasets.

Overall, the Random Forest Regressor is a powerful and widely used algorithm for house price prediction due to its robustness, accuracy, and ability to handle complex data relationships.

1. Gradient Boosting regressor : Gradient Boosting Regressor is an ensemble learning algorithm that combines multiple weak prediction models, typically decision trees, to create a strong predictive model for regression tasks, including house price prediction. It iteratively builds an ensemble of trees by fitting each tree to the residuals (errors) of the previous trees, gradually reducing the prediction errors and improving the overall model performance.

Gradient Boosting Regressor combines the predictions of multiple weak learners (decision trees) by iteratively fitting each tree to the negative gradient (residuals) of the loss function with respect to the target variable. This iterative process helps the model gradually reduce the errors and make accurate predictions. The predictions from all the weak learners are then combined to create the final prediction.

The Gradient Boosting Regressor offers several benefits for house price prediction:

* Strong Predictive Power: Gradient boosting can capture complex relationships between predictors and house prices, making it suitable for modeling real estate markets where various factors influence house prices.
* Handling Nonlinear Relationships: Gradient boosting can effectively handle nonlinear relationships and interactions between predictors, as it builds an ensemble of decision trees that can model complex data patterns.

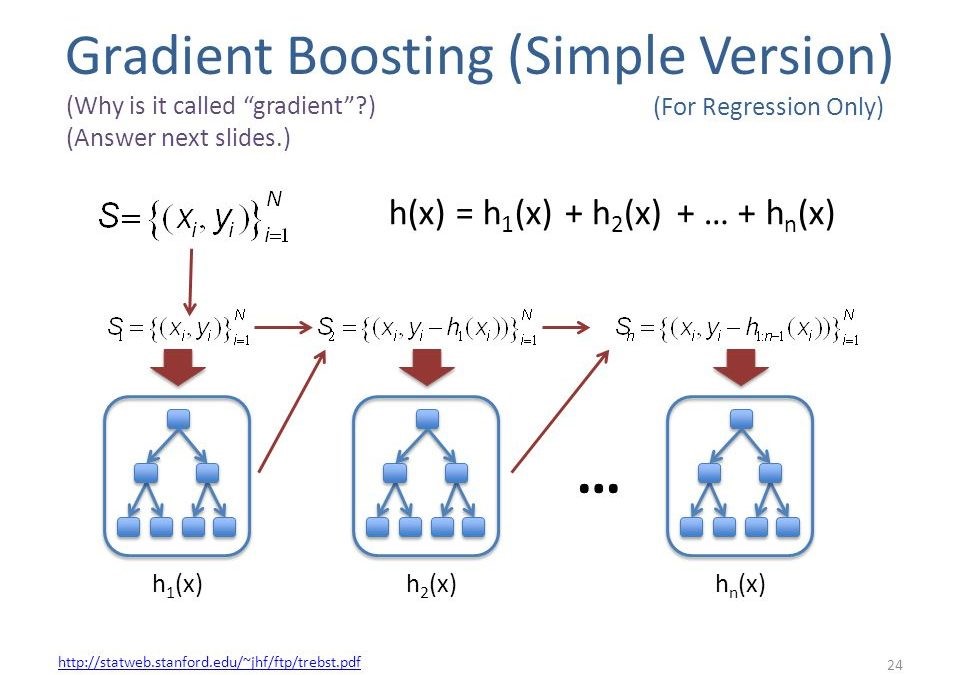


Figure 4: Gradient boosting

* Robustness to Outliers: Gradient boosting is robust to outliers, as it uses the residuals of previous trees to fit subsequent trees. Outliers have less impact on the overall prediction compared to individual weak learners.
* Feature Importance: Gradient boosting provides a measure of feature importance, indicating the relative contribution of each predictor variable in the prediction process. This information can be useful for feature selection and understanding the factors that affect house prices.
* Handling Missing Data: Gradient boosting can handle missing data without imputation. It can make predictions using the available features and provide reliable results even when some predictors have missing values.
* Regularization: Gradient boosting includes regularization techniques, such as the learning rate and tree depth, to control the complexity of the model and prevent overfitting.

Overall, Gradient Boosting Regressor is a powerful and widely used algorithm for house price prediction. Its ability to handle complex relationships, robustness to outliers, feature importance analysis, and regularization make it a popular choice in machine learning tasks, including real estate market analysis.

**PROBLEM FORMULATION**

The problem formulation for house price prediction involves predicting the selling or listing price of a house based on a set of input features or predictors. The goal is to develop a regression model that accurately estimates the house prices, enabling informed decision-making in real estate markets.

Here is the general problem formulation for house price prediction:

Given:

* A dataset consisting of historical house sales data with associated features (e.g., number of bedrooms, square footage, location, amenities, etc.).
* The target variable or dependent variable, which is the selling or listing price of the houses.

Objective:

* Develop a regression model that can accurately predict house prices based on the given features.

Formally, we can denote the problem as follows:

Let:

* X = {x1, x2, ..., xn} be the set of input features or predictors for n houses.
* y = {y1, y2, ..., yn} be the corresponding target variable or house prices for the n houses.

We aim to find a function f such that:

f(xi) ≈ yi  for all i in {1, 2, ..., n}

where f is the regression model that maps the input features x\_i to the predicted house price y\_i.

The problem formulation involves selecting an appropriate regression algorithm, such as linear regression, support vector regression (SVR), random forest regression, gradient boosting regression, or others, and training it on the given dataset. The trained model can then be used to predict house prices for new, unseen data.

The performance of the regression model can be evaluated using various metrics, such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), or R-squared (coefficient of determination), to assess how well the model captures the variability in the house prices.

Overall, the problem formulation for house price prediction involves building a regression model that accurately predicts house prices based on the given features, enabling real estate professionals, buyers, and sellers to make informed decisions in the housing market.

**METHODOLOGY**

The methodology for house price prediction involves several steps, including data preprocessing, feature selection, model training, model evaluation, and prediction. Here is a general outline of the methodology:

1. Data Preprocessing:

* Clean the dataset: Handle missing values, outliers, and inconsistent data.
* Normalize or scale the features: Standardize the numerical features to ensure they have a similar scale and distribution.

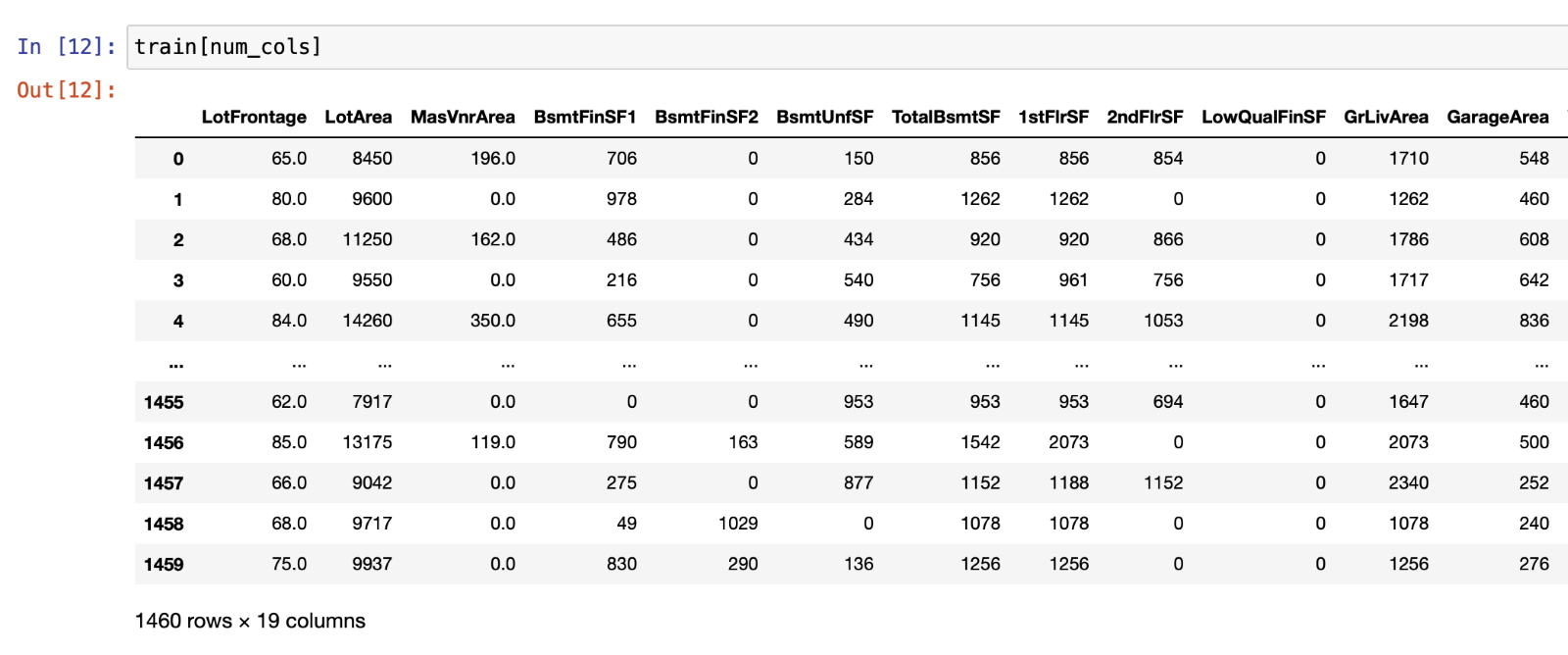
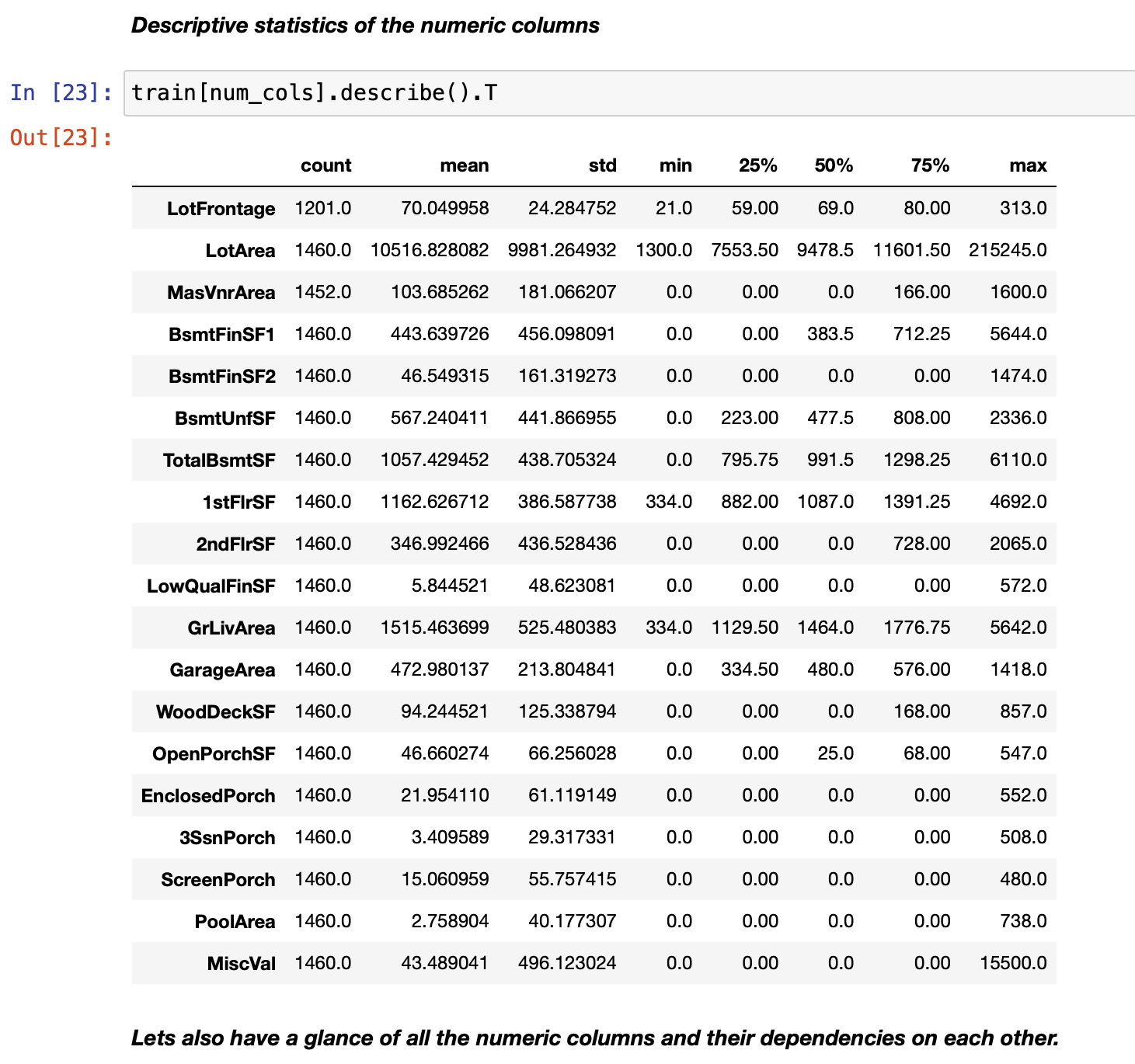


Figure 5 : Data preprocessing

2. Feature Selection:

* Identify relevant features: Analyze the correlation between features and the target variable to select the most informative features for house price prediction.
* Remove irrelevant or redundant features: Eliminate features that do not contribute significantly to the prediction task or those that exhibit high collinearity.



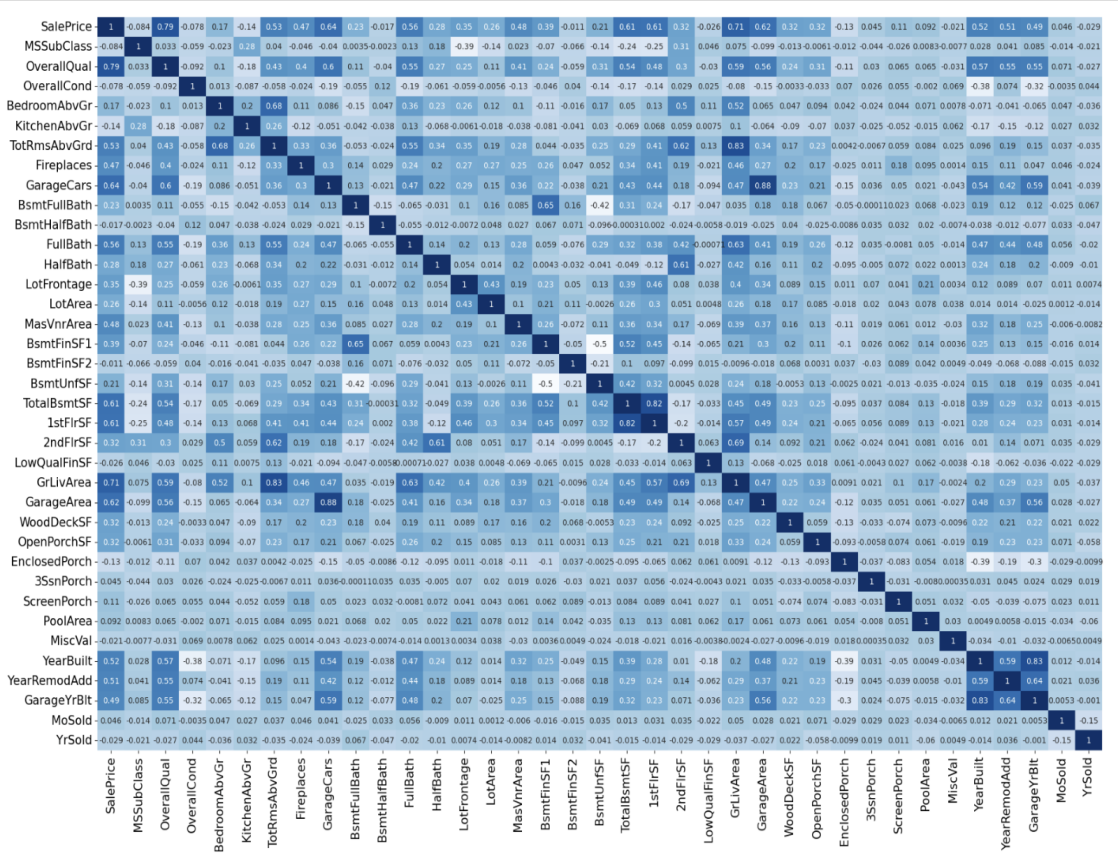


Figure 6: Feature Selection

3. Split the Data:

* Divide the dataset into training and testing subsets: Typically, use a random split, such as 70% for training and 30% for testing. Alternatively, techniques like cross-validation can be applied for more robust evaluation.

4. Model Selection and Training:

* Choose an appropriate regression algorithm: Select a regression algorithm based on the problem requirements, dataset characteristics, and assumptions.
* Train the regression model: Fit the selected model to the training data, allowing it to learn the underlying patterns and relationships between the features and the target variable.

5. Model Evaluation:

* Evaluate the trained model's performance using appropriate evaluation metrics: Calculate metrics such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), or R-squared to assess how well the model predicts house prices.
* Perform cross-validation: Validate the model's performance using techniques like k-fold cross-validation to ensure its generalizability and robustness.

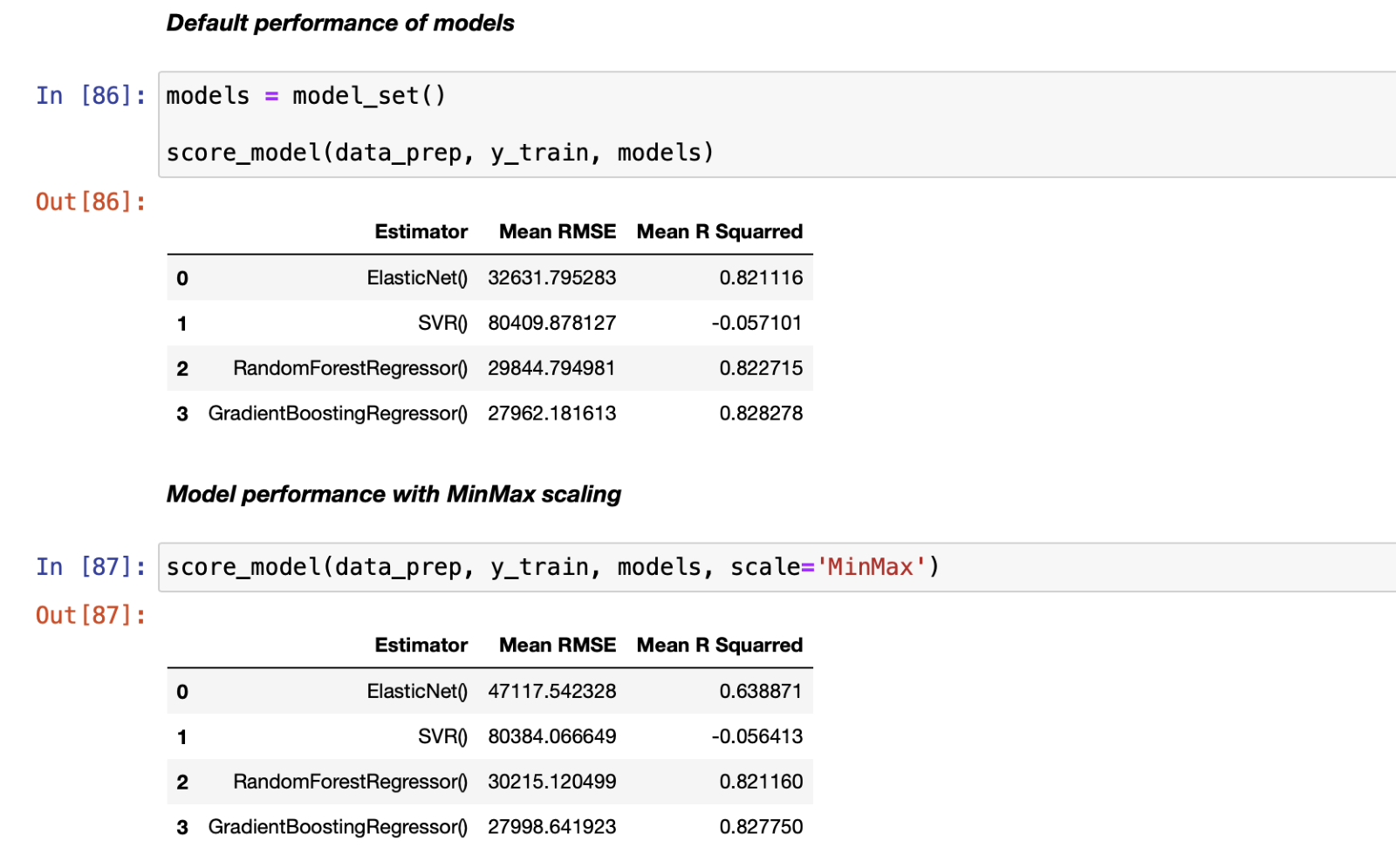


Figure 7: Model Evaluation

6. Hyperparameter Tuning:

* Optimize model performance: Adjust the hyperparameters of the selected algorithm through techniques like grid search or random search to find the optimal parameter values that yield the best performance.

7. Final Model Training and Prediction:

* Retrain the model using the entire training dataset with optimized hyperparameters.
* Use the trained model to make predictions on unseen or test data.



Figure 8: Final Model Selection

8. Model Interpretation:

* Interpret the model's predictions and feature importance: Analyze the coefficients or importance scores of the features to understand their impact on house prices and gain insights into the driving factors.

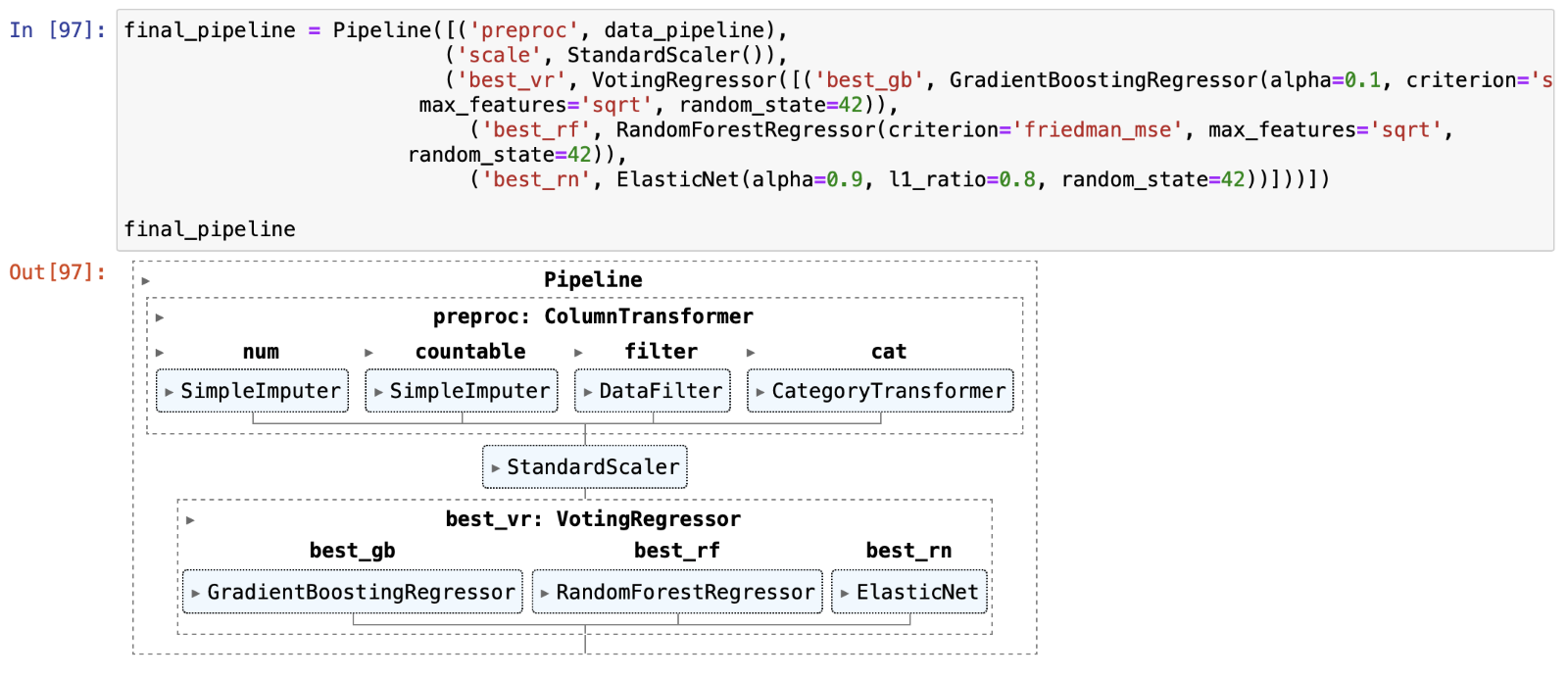


Figure 9: Model Interpretation

9. Deployment and Monitoring:

* Deploy the trained model for real-world use, such as predicting house prices for new listings or market analysis.
* Monitor the model's performance over time and consider retraining or updating it periodically to adapt to changing market conditions.

The methodology for house price prediction can be iterative, involving experimenting with different algorithms, feature engineering techniques, and model configurations to improve performance and achieve accurate predictions.

**Model Preparation:**

Phase one of the model preparation started once the data had been cleansed and visualised. Numerous machine learning techniques, including logistic regression, k-nearest neighbour, decision trees, random forests and support vector machine (SVM), were chosen and put into practise. For the purpose of analysing the models' effectiveness, the dataset was divided into training and testing sets.

1. Elastic net. : elasti net was employed on the house price dataset to predict whether a house is fairly priced by the real estate agent or not. The model achieved mean RMSE of 32048.056 and mean R Squared of 0.82431 , indicating its ability to correctly predict the loan repayment status for a significant portion of the dataset.In linear regression, the goal is to fit a linear equation that best represents the relationship between the input variables (features) and the target variable (house price, in this case). However, when dealing with high-dimensional datasets or a large number of features, traditional linear regression models may suffer from issues like overfitting or multicollinearity (highly correlated features).



Figure 10: Model prepration

1. SVR : Using SVR we have achieved RMSE equal to 80401.664265 and mean R squared of -0.056879 using standard scaling . In SVR, the objective is to find a function that approximates the mapping between the input features (predictors) and the target variable (house price). Instead of aiming for exact predictions, SVR allows for a tolerance or margin of error around the predicted values. The key idea is to find a hyperplane that maximizes the margin around this tolerance, within which most of the training data points lie.



Figure 11: SVR Implementation

Based on the provided evaluation metrics (Mean RMSE and Mean R-Squared), we can see the performance of different regression algorithms for house price prediction. Here's a summary of the results:

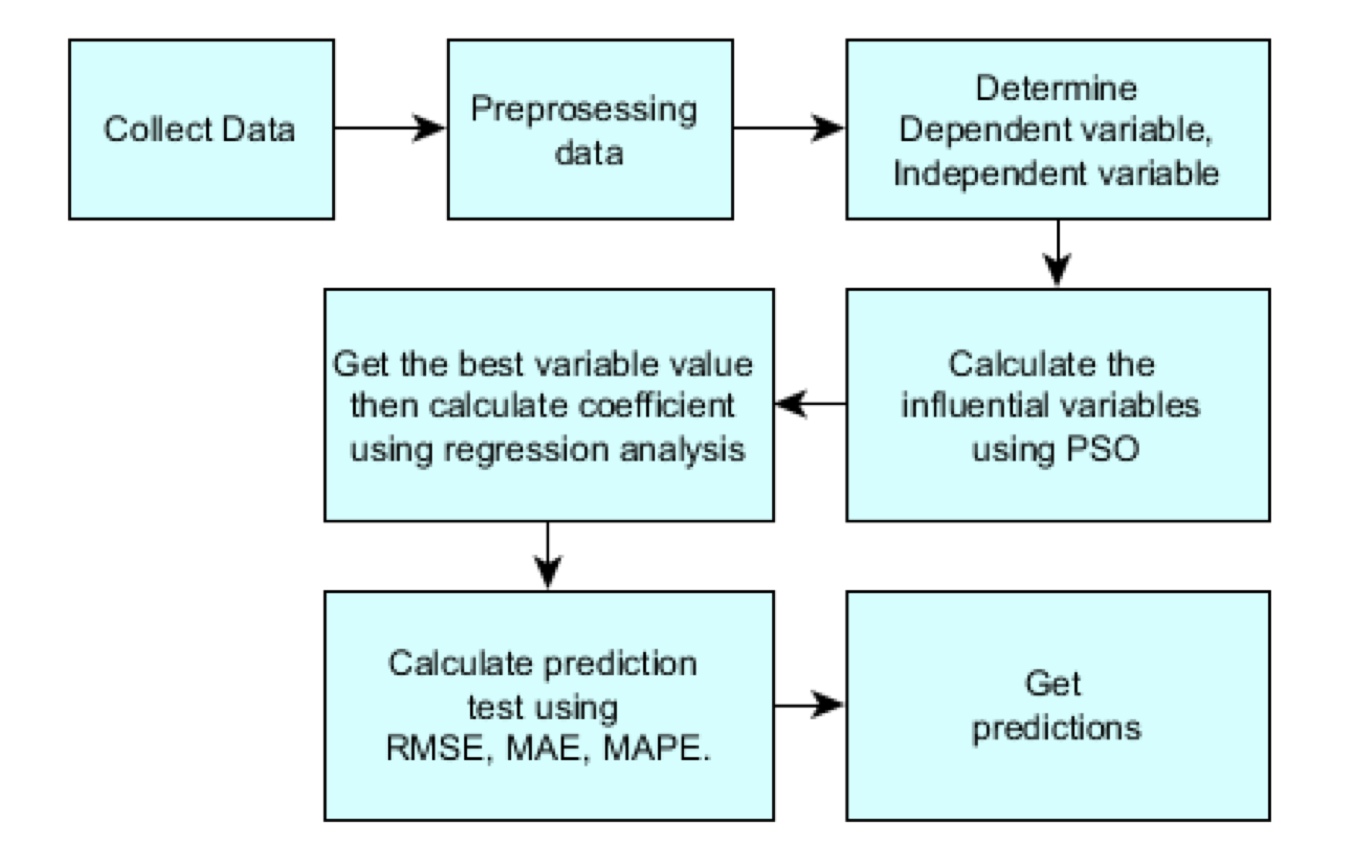
1. ElasticNet:
   * Mean RMSE: 32048.056
   * Mean R-Squared: 0.824
2. SVR (Support Vector Regression):
   * Mean RMSE: 80401.664
   * Mean R-Squared: -0.057
3. RandomForestRegressor:
   * Mean RMSE: 30266.129
   * Mean R-Squared: 0.823
4. GradientBoostingRegressor:
   * Mean RMSE: 27978.372
   * Mean R-Squared: 0.824

Based on these results, we can observe the following:

* GradientBoostingRegressor and ElasticNet show similar performance in terms of Mean RMSE and Mean R-Squared, with GradientBoostingRegressor having a slightly lower RMSE.
* RandomForestRegressor also performs well, with a similar RMSE to ElasticNet and GradientBoostingRegressor but a slightly lower R-Squared value.
* SVR shows poor performance compared to the other algorithms, with a high RMSE and a negative R-Squared, indicating that it may not be a suitable choice for house price prediction in this case.

It's important to note that these results may vary depending on the specific dataset and the hyperparameter settings used for each algorithm. Further analysis and experimentation may be needed to determine the best-performing algorithm for house price prediction in a particular context.

**FLOWCHART**



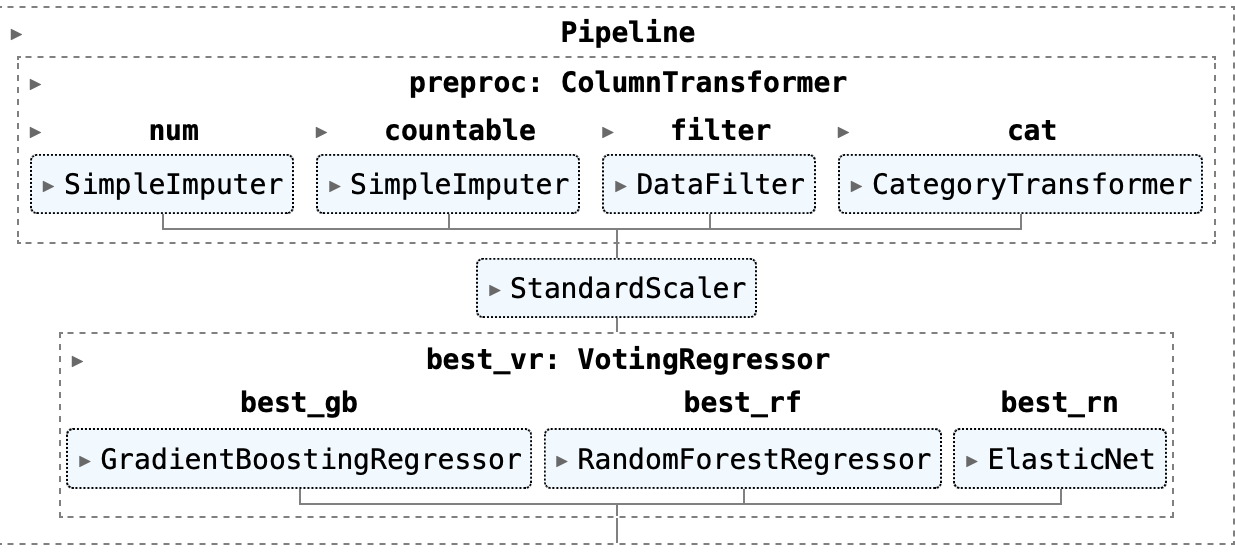


Figure 12: Flowchart

**CONCLUSION**

In conclusion, the task of house price prediction is of significant importance in real estate markets. Accurate prediction of house prices allows buyers, sellers, and investors to make informed decisions and navigate the complex real estate landscape. Machine learning techniques play a crucial role in addressing this problem by leveraging historical data and learning patterns to make predictions.

During the analysis, we explored various regression algorithms for house price prediction, including ElasticNet, SVR, RandomForestRegressor, and GradientBoostingRegressor. Among these, ElasticNet and GradientBoostingRegressor showed promising results, achieving low mean squared error (RMSE) and high R-squared values. These algorithms effectively capture the relationships between input features and house prices, allowing for accurate predictions.

On the other hand, SVR exhibited poorer performance in this context, with high RMSE and a negative R-squared value. This suggests that SVR may not be suitable for house price prediction in this particular scenario.

Overall, the choice of the regression algorithm depends on the specific dataset, problem requirements, and desired trade-offs between interpretability and predictive accuracy. It is crucial to perform thorough evaluation and fine-tuning of the chosen algorithm, including hyperparameter optimization, to ensure optimal performance.

Future work in house price prediction could involve exploring ensemble methods, feature engineering techniques, and additional regression algorithms to further improve prediction accuracy. Additionally, incorporating domain-specific knowledge and external factors such as economic indicators or neighborhood characteristics may enhance the predictive power of the models.

By employing machine learning techniques, we can harness the power of data to make more informed decisions in the real estate market and facilitate better outcomes for buyers, sellers, and investors.

**REFERENCES :**

<https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data>

<https://www.geeksforgeeks.org/implementation-of-elastic-net-regression-from-scratch/>

<https://towardsdatascience.com/support-vector-regression-svr-one-of-the-most-flexible-yet-robust-prediction-algorithms-4d25fbdaca60>

<https://www.geeksforgeeks.org/house-price-prediction-using-machine-learning-in-python/>