

Prediction of House Prices using Machine Learning Algorithms

Guided by Mr.Suresh Kumar Choudhary

Pradyumna Kumar Sahoo, Bimal Kumar

2018MSBDA016@curaj.ac.in,2017MSBDA002@curaj.ac.in

Dept. of Data sciences and Analytics

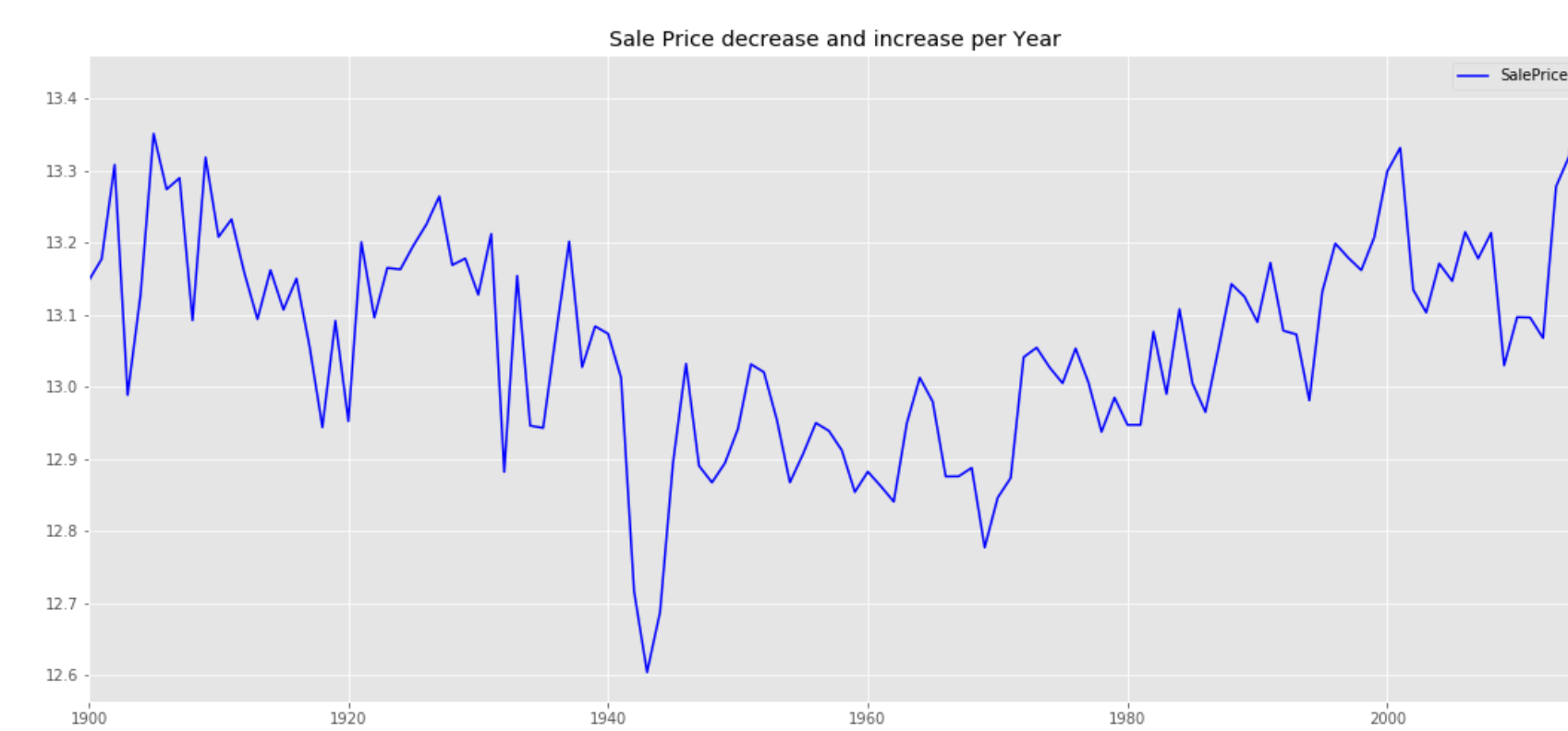


Abstract

Housing sales price are determined by numerous factors such as area of the property,floors,age of the property,number of bedrooms and bathrooms so on.We have used machine learning algorithms like Lasso,Ridge,Elastic-net,Random Forest and Gradient boosting regression to build the prediction model for houses. Here, machine learning algorithms have been employed to the data of 21646 properties.Further, we have compared these algorithms based on error metrics such as MSE,RMSE and R2 score.



1 Why ??



Investment in real estate generally seems to be profitable because their property values do not decline rapidly thus an attractive choice for the investments.Yet we lack proper standardized ways to measure the real estate property values.We require an automated prediction model that can help to predict the property values without any bias and help the first-time buyers and less experienced customers to understand whether the property rates are overrated or underrated.We have considered various intrinsic and external parameters.Then we have applied these parameter values to different machine learning algorithms comparing them over various error metrics like R-Squared value and Root Mean Squared Error(RMSE).

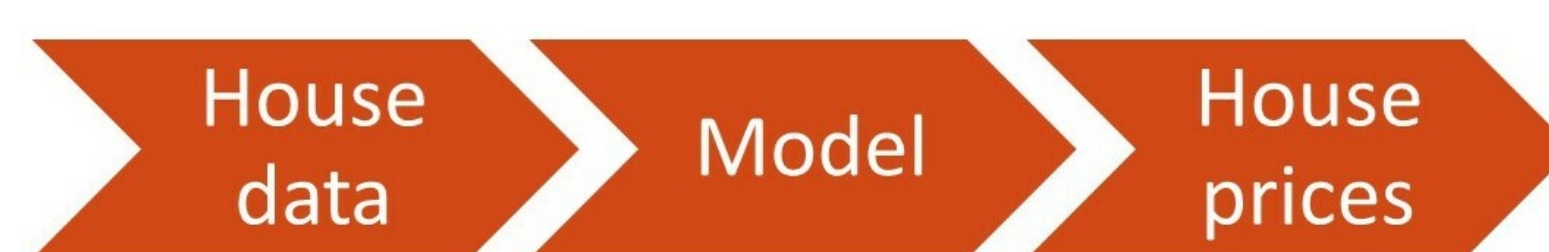


2 What ??

We hereby collect data,clean it appropriately,then train our Machine learning Model and then use it to provide estimates based on the de-

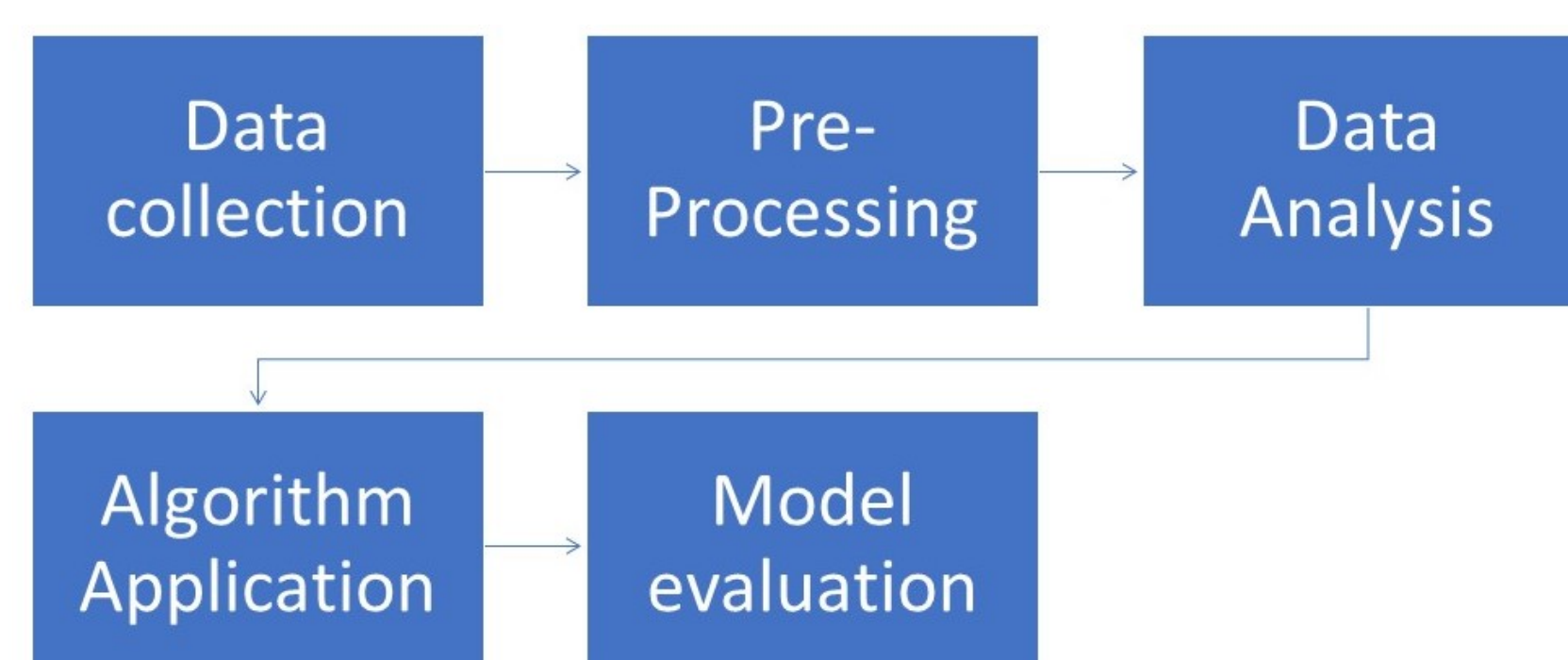
mand of the users.

WE ARE PREDICTING HOUSE PRICES BASED ON THEIR FEATURES.



3 How ??

The following diagram (figure 1) represents step-wise tasks that need to be completed:

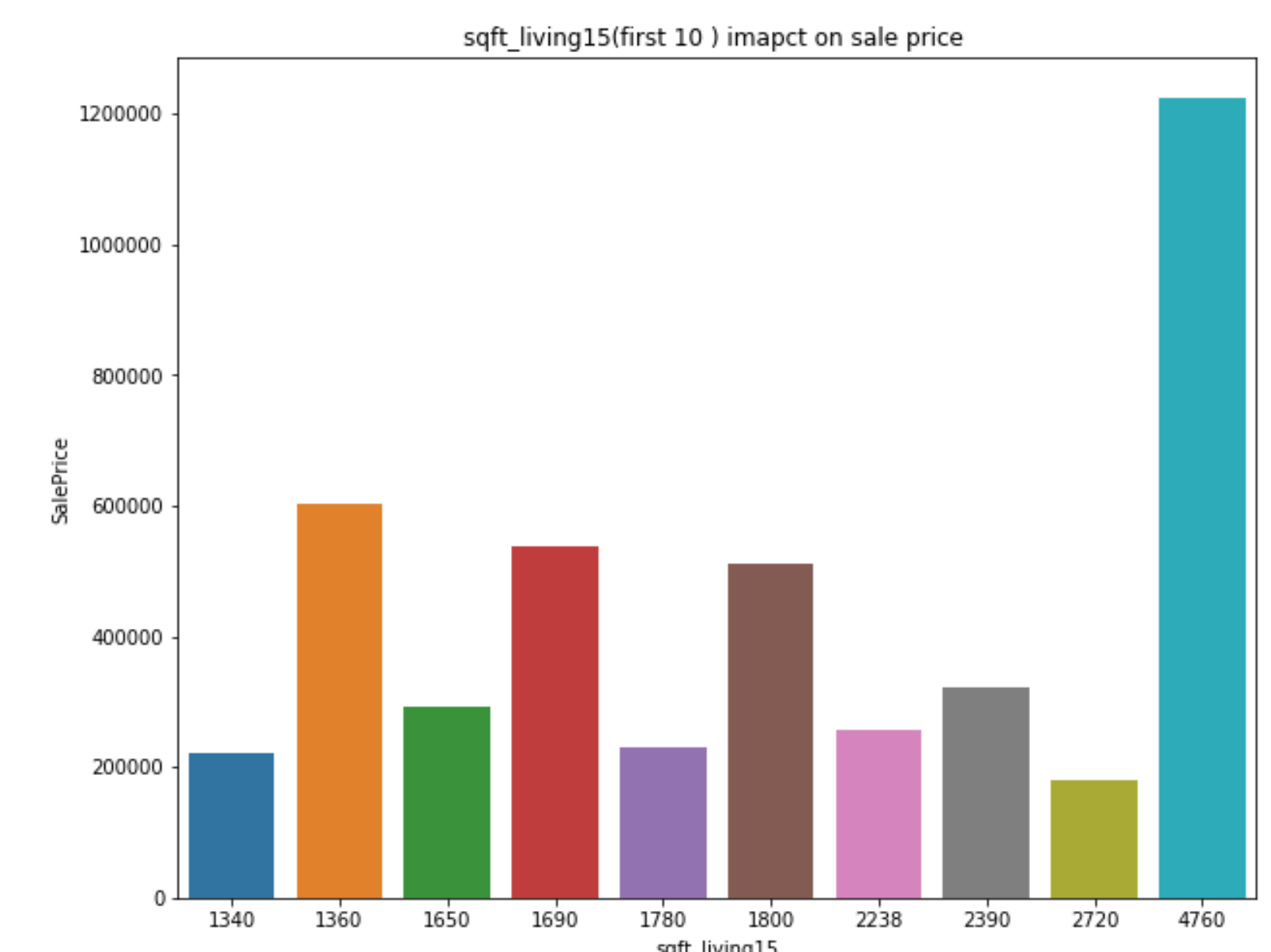
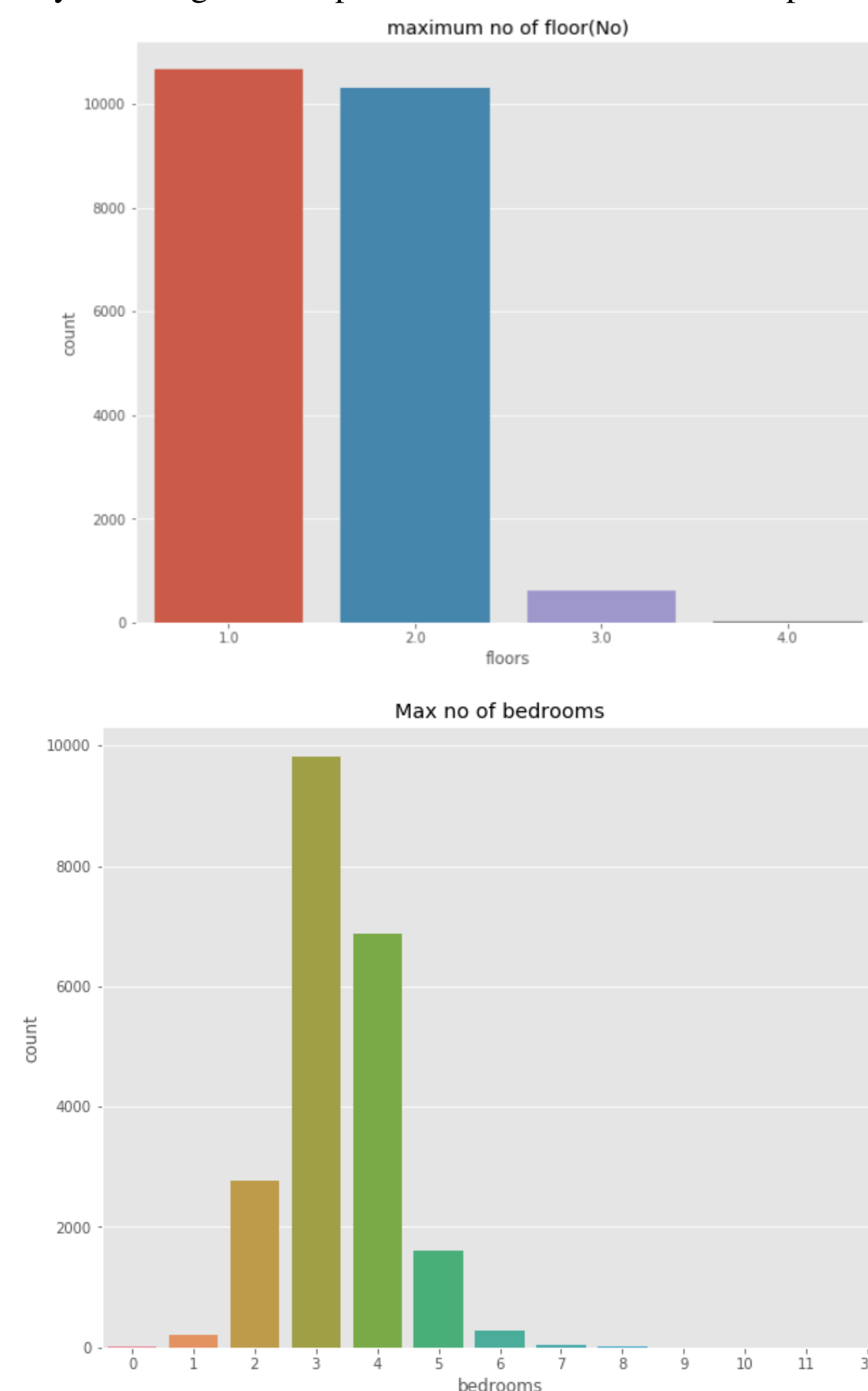


3.1 Data Collection and Pre-processing

The data set used in this project was an open source data set from Kaggle. It consists of 21614 records with 19 parameters that have the possibility of affecting the property prices.Entire data set is checked for NaN and whichever observation consists of NaN will be deleted or imputed.However,in our data set,there was no missing values found,meaning that every record was constituted its corresponding feature values.

3.2 Data Analysis

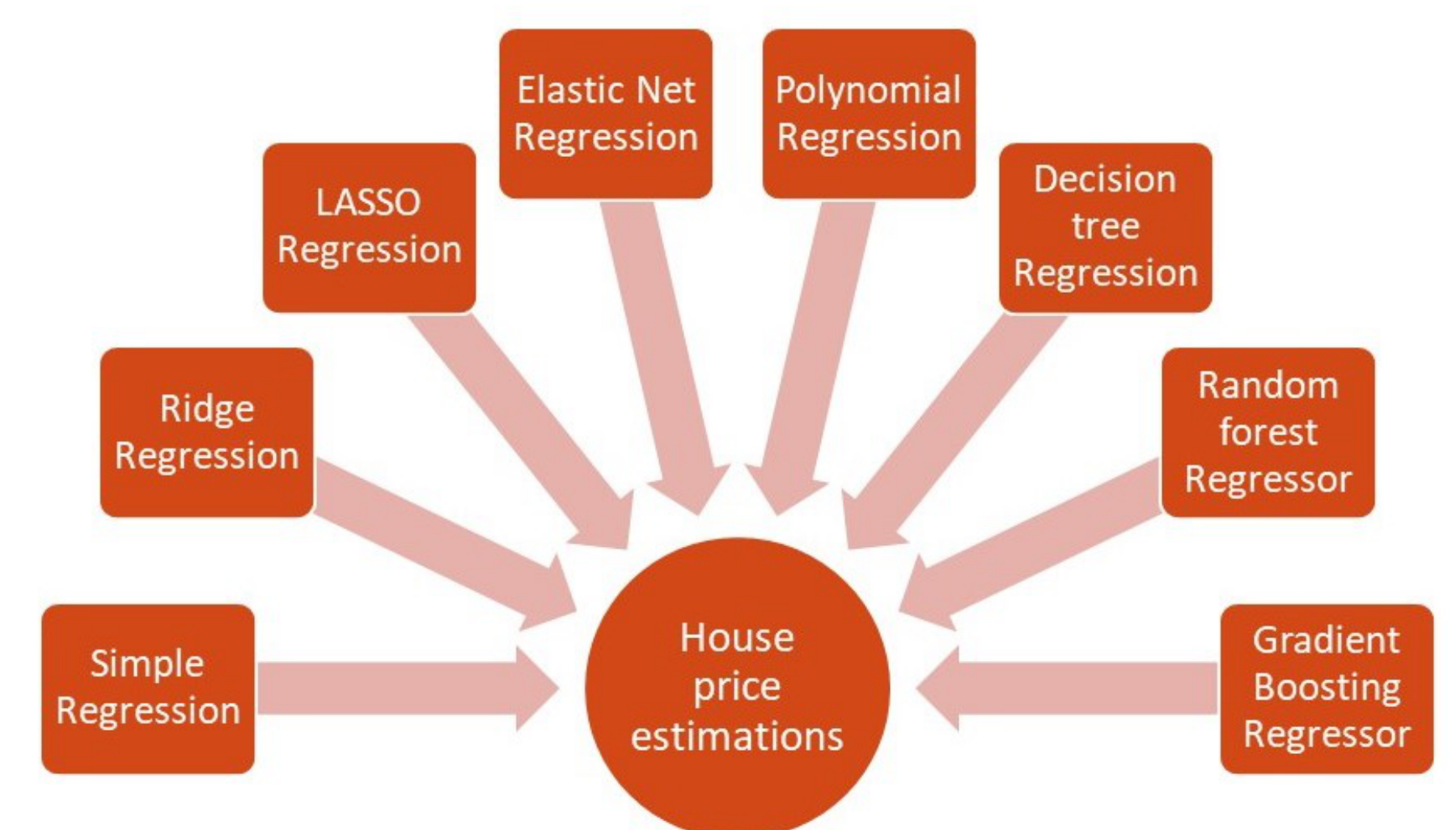
we need to analyse our data set and study the different parameters and relationship between these parameters.We weeded out the outliers present in our data set.We did some univariate and bivariate analysis taking most important features.Here are some plots.



4 Application of Machine Learning Algorithms

Once the data is clean and we have gained insights about the data set, we can apply an appropriate machine learning model that fits our dataset.The following algorithms were implemented with the help of python's SciKit-learn Library.

Machine algorithms applied



4.1 Results

| Methods | R-Square | Explained variance of model | Mean squared error | RMSE | Cross validation score | Model Score |
|----------------------------|----------|-----------------------------|--------------------|------|------------------------|-------------|
| Simple Regression | 0.73 | 73% | 0.06 | 0.24 | 78% | 78% |
| Ridge Regression | 0.73 | 73% | 0.06 | 0.24 | 78% | 78% |
| Lasso Regression | 0.72 | 72% | 0.06 | 0.24 | 78% | 78% |
| Elastic Net Regression | 0.72 | 72% | 0.06 | 0.24 | 78% | 78% |
| Polynomial Regression | 0.77 | 77% | 0.05 | 0.23 | 78% | 78% |
| Decision Tree Regressor | 0.72 | 72% | 0.08 | 0.28 | 71% | 71% |
| Random Forest Regressor | 0.79 | 79% | 0.05 | 0.21 | 83% | 83% |
| GradientBoosting Regressor | 0.87 | 87% | 0.03 | 0.17 | 88% | 88% |

Conclusion

The Gradient boosted regression model shows highest accuracy in predicting house prices with 88 percent.There are various other methods out there like KNN regression,Support Vector regression,Multi-Layer perceptrons.Meanwhile,We will be working on those. However,there always exists scope for improvements.So awaiting your feed backs.

THANK YOU