

Paper ID: MISS21_70

Prediction of Boston's House Price with Regression

Moutushi Shil

Department of Computer Science and Engineering, Techno College of Engineering Agartala, India

Corresponding author

Abstract

Background: Data science is a process by which some data are fed to a machine and it will give us output on the basis of data's that is fed on it, particularly called Machine Learning. We know that houses with more rooms have higher prices, though a lot of factors are being act there at same time such as no. of windows, transport facilities, etc. House besides with lower class workers (higher 'LSTAT'¹ value) will worth less [1]. Similarly, neighborhoods with more students to teacher's ratio (higher 'PTRATIO'² value) will be worth less [2]. If people earn less money it is, that their houses are worth less. They are inversely proportional variables.

Objectives: To build a ML model which can perform and predict prices of houses from the trained and tested data from houses in suburbs of Boston, Massachusetts. This model can act as someone like a real estate agent whose work is to analyses these on a daily basis.

Methodology: It is necessary to split the total data into training and testing part for better performance of the model. Usually at a ratio of 80:20 we are dividing the data's in two mentioned parts. Here we will calculate the coefficient of determination [3] i.e. R^2 , to analyses the model's performance. . Usually R^2 range from 0 to 1, which means the percentage of squared correlation between the predicted and actual values of the target variable. A model with an R^2 of 1 exactly predicts the target variable. Any value between 0 and 1 indicates what percentage of the target variable.

Results and discussion: After we split the data into k partitions of equal size. For each partition i, we train the model on the remaining k-1 parameters and evaluate it on partition i.e., The final score is the average of the K scores obtained.

Conclusions and future work: Now that we use these results to discuss whether the built model can or cannot be used in a real- world. It is fair to say that generally to estimate or predict the price of an individual home based on the properties of the its neighborhood. In the same neighborhood there can be huge differences in prices.

References

- [1] LSTAT: This is the percentage lower status of the population, towards data science.com
- [2] PTRATIO This is the pupil-teacher ratio by town.
- [3] Coefficient of determination. It is the amount of the variation in the out, GFG

