What is the problem

A program to predict median house price in Boston is needed.

Task: Predict the median house price in Boston

Experience: A database of houses with their median house price

Performance: Regression accuracy, the RMSE error between the predicted and ground truth values.

Assumptions

The median price is higher where the crime rate is lower

The price is correlated with how much retail area is nearby

The price is not correlated with nearby river parameter

The price is higher when the number of rooms increases

The price is higher when nearby radial highways

Tax doesn’t influence the price

The price is influenced by number of blacks and lower status of population

Similar problems

Predict level of education by residence, predict higher status population by residence.

Why does the problem needs to be solved

It is a learning exercise

Solution Benefits

Prediction of median house prices in an area may tell us if the asset we are interested to buy is given in a reasonable price and is worth investing or not.

Solution Use

The model is expected to live a few years, until a drastic change in the market, which influences the training data we acquired (making it not valid anymore).

It can be used by any individual interested to buy a house in Boston

How would I solve the problem

First read about similar solved problems and the methods used. Then, I would clean the data first – checking Nans and filling them. Checking there are no contradicting duplicate values. I would perform feature engineering – performing feature selection, adding relevant features, scaling. Testing the above assumptions as well. Then I would use the most used machine learning methods – Random Forests, XGBoosting trees, Lightgbm and more.