TD2 : optimisation de modele

For this TD we are using a classic real estate price base. The objective is to predict the "SalePrice".

The goal is to select the best model and optimize it.

This is a regression our performance function could be the R2. As a reminder, R2 measures the part of explained variance. R2 = 1 at most (perfect model .. doubtful ..), R2 = 0: the model is as frustrating as an average (we reduce the same price for all houses, and R2 <0: it is possible .. you are doing worse than predicting the average for all houses.

The price distribution to be predicted is spread over several orders of magnitude: a good reflex is to work on the log of the price rather than the price: this avoids too much sensitivity to outliers, "Gaussianizes" the price distribution (do this in observation) as well as the summaries (very useful for linear regressions

The feature engineering is already done (already proposed rather .. we can still continue it but this is not our goal today

What we need to do

1) **Select a good model (the best one ... would be pretentious)**

At least 3 differet algorithms, with justification of hyperparamter tuning

2) **Exploring this model : important features, at least**
3) **Optimize the model**

The target is to use just the minimum useful features, keeping the performance.

What is the use of this operation ?