FEATURE IMPORTANCE

The feature importance is a way to reduce the dimensionality of the dataset that you have with you by removing the unimportant features.

- It is **calculated** by randomly shuffling the values of a particular column of your dataset and then calculating the validation score on that data.
- If the difference in the validation score for your validation score of original data and the validation of your shuffled data is high then THAT FEATURE IS ACTUALLY IMPORTANT.
- Note that that particular **column is not removed** but it is **just shuffled so that you don't have to retrain your forest** on your new dataframe.
- A very important point is to look at **RELATIVE IMPORTANCE AMONG THE FEATURES.**
- **For example-** if feature importance of a particular variable is 0.5 and the second most important is 0.46 and onwards it starts from 0.0006 so don't look at others as the second most important is **LIKE A THOUSAND TIMES MORE IMPORTANT** relatively to third and so on features so drop them.
- Just note that this variable WHICH IS A THOUSAND TIMES IMPORTANT OR SOMETHING is **not** some **DIRECT ENCODING OF YOUR DEPENDENT VARIABLE BUT A FEATURE THAT IS REALLY AVAILABLE FOR ANALYSIS.**