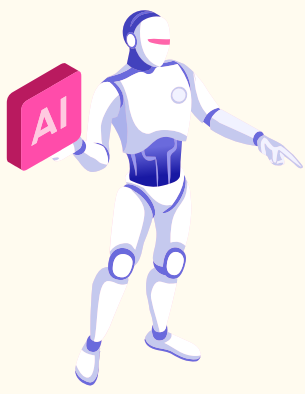
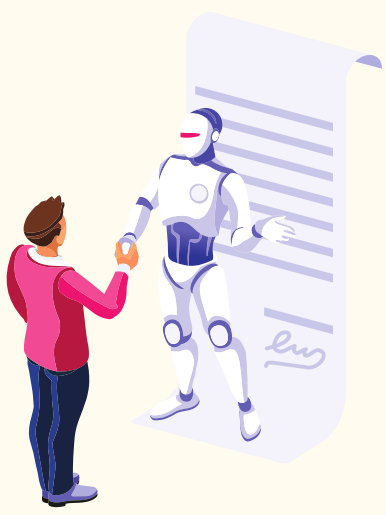
CHALLENGE

XGBoost and SHAP vs House Price Predictions

01 LIBRARIES

SHAP is not part of the pre-installed libraries in Colab. Install it and import other necessary libraries.



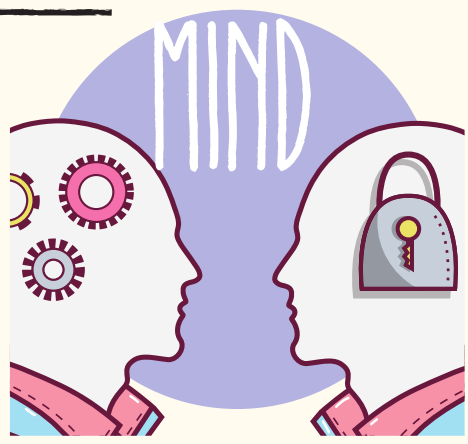


02 STRING VARIABLES

All inputs must be numeric. Transform all string variables into dummy variables or equivalent.

03 DATA PREPARATION

We can swiftly make the preparations for XGBoost. First, isolate X and Y. Then, create XGBoost matrices.

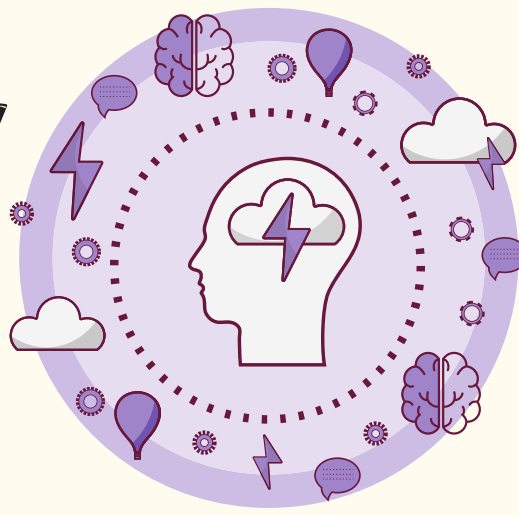


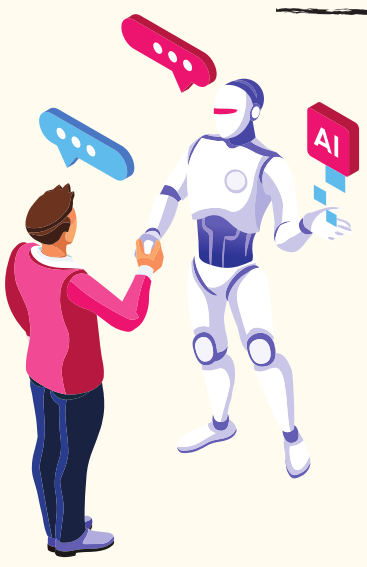
04 XGBOOST

XGBoost has a lot of parameters to play around with. Set them. The most important part is that the *objective* matches your problem: regression, classification, etc... Then, you are ready to create the XGBoost model.

05 LOCAL INTERPRETABILITY

The first part is to explain individual predictions. It is important when you are interested in specific explanations or hedge cases.



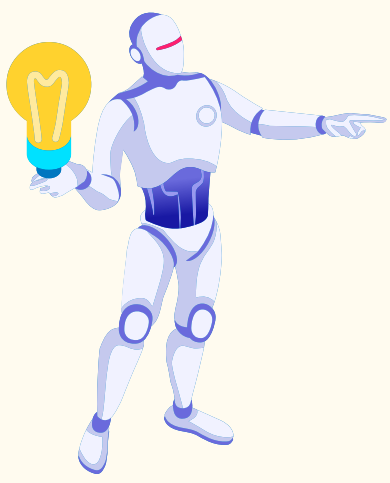


06 DEPENDENCY PLOTS

The dependency plots are particularly interesting to study how variables interact with each other. One thing to have in mind is that Dependency Plots work best with continuous variables.

07 GLOBAL INTERPRETABILITY

Uff, last step. This was a long one, I know. Global Interpretability is the Local Interpretability combined. we see how variables impact the dependent variable.



Diogo Resende