Explainable AI

First level of explicabity: statistics

In this class we will use data from the same dataset we used in the previous TD. We will build upon the work done previously.

- We will be using python for the data analysis (R is accepted as well if needed)
- The main libraries we are going to use are pandas,SHAP, matplotlib and sklearn
- This TD will not be graded.

Question 1.

Create a new notebook and from TD2 load the variable selection and the different models that you created

Question 2.
Install the shap package (documentation here https://shap.readthedocs.io/en/latest/index.html)

Question 3.
Let's begin with an enrichment on the random forest interpretation.

- Get the variable importance of your features
- Recall the explanation you gave on them
- using the Treeexplainer function of the shap package, find the shapley values
- Visualise a couple of the explanations that are given with shap.forceplot. Can you explain them?
- Using the "summary plot" with the "plot type" equal to bar feature can you compare the importance that shapley gives to the importance that the native "variable importance" has given you?
- Do you understand the plot if you remove the plot type?

Question 4.
Now lets work on a different model. Let's train an Xgboost
How does Xgboost work ? what is boosting? you can search the internet for these answers

- Install the Xgboost library if it's not already installed
- Fit an Xgboost to your data and fine tune it
- Is it better than your Random Forest? why?

Question 5.

Let's compare the feature importance of the Random Forest and the Xgboost

- Using shap's tree explainer get the shapley values for this new model.
- Select the variables that are the most important and plot a dependency plot. Does the result confirm your intuition?
- Compare the summary plot to the summary plot of the random forest. What are the changes?

Question 6.

We can use shapley values from the Xgboost to learn more about the properties. Let's cluster the properties based on the shapley values.

- Let's reduce the dimension of our data for visualisation purposes. Do a PCA with the shapley values and visualise the 2 principal axis.
- Given the visualisation choose a clustering algorithm (K-Means, DBSCAN, gausian mixture...) and try to cluster the shapley values of the property.
- Visualise your results
- What characteristics can you give to each cluster?
- What conclusion can you reach?