Further information on phase 8.2

Reflection Part 1 - Comparison of the automatic and manual creation process of a decision tree

After the unit, it is reflected how the decision trees were created with the help of the computer. You can only see the final result of the automatic creation process, but something similar has happened in the background to the manual process with the data cards. A discussion can be structured as follows. First, start with the thesis that a decision tree is created with the help of the computer in a similar way to the manual process with the data cards. The following aspects should be compared:

* A feature was selected manually and cards were sorted according to this. The difficulty is to use all features manually because it takes a lot of time.
* Various threshold values were tested manually. However, threshold values were not systematically tested in all gaps.
* The computer is used to test **all** features and try out **all** sensible threshold values, i.e. a threshold value is tried out for **each** "gap" between two data cards.
* The number of incorrect/correct classifications was evaluated manually depending on the threshold value in order to compare threshold values/decision rules. This is implemented by the computer in exactly the same way, only for all features and all possible "gaps".
* However, the computer performs the necessary steps (sorting, identifying gaps, selecting threshold values, calculating misclassifications and comparing decision rules) much faster than is possible manually.
* The computer can thus systematically try out all features and gaps.
* The computer therefore finds the best decision rule at each level.

Reflection Part 2 - Possible applications and limitations of the self-created decision trees for classifying foodstuffs

A resulting decision tree is not an absolutely valid system of rules for recommended lifestyles, but it can still give us good pointers. We have seen that decision trees can always make mistakes

The following aspects should be discussed with pupils in plenary.

* It is not enough to look at one characteristic to decide whether a food is recommendable.
* The characteristics of fat, sugar and energy are particularly good indicators (but should not be used alone as a decision criterion).
* It is possible to gain and visualize insights that are hidden in data by creating decision trees.
* However, the decision trees created also have weaknesses with regard to their general validity as a recommendation system. Reasons for weaknesses that can be traced back to data:
* The data was labeled rather intuitively. The class was not always in agreement or certain when it came to assigning rather recommendable/rather not recommendable labels. The quality of the decision tree stands and falls with the modeling of the data. (If necessary, refer to phase 8 if a run was made in the Jupyter Notebook with data that was labeled randomly or incorrectly) --> Data is the most important building block as the basis of a decision tree
* In nutritional science, there are other important characteristics of foods that have not even been surveyed (e.g. fiber, vitamins, etc.).
* Reasons for weaknesses due to the context:
* A recommended diet is first and foremost versatile and cannot be covered one-to-one by a decision tree.