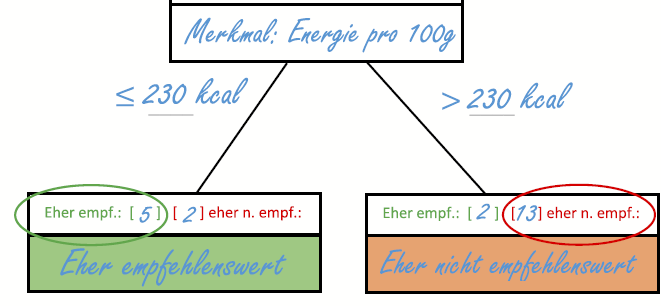
Further information on lesson 3

Notes on the implementation of *vital statistics*

Due to the limited number of pupils, not all cards can usually be used for the living statistics. Care should be taken not to use unfavorable samples of the cards. Some samples simplify the problem too much, making it too easy to find perfect rules. We suggest a truncated dataset with which the phase can be carried out well. Suggestions for hull data sets of different sizes can be found in the files "Rumpfdatensatz\_22.csv" and "Rumpfdatensatz\_28.csv". Slight deviations from the body data sets are not a problem.

Living Statistics is performed with the entire class to perform data splits based on thresholds and derive data-based decision rules.

1. Each pupil takes a data card (including a colored bracket) and represents this food.
2. The teacher names a threshold value (e.g. energy and 350 kcal). Each pupil whose food has an energy value≤ 350 goes to the left side of the classroom, each pupil with a value > 350 goes to the right side.
3. In the resulting partial data sets, the number of red and green brackets ("rather recommendable" or "rather not recommendable") is determined, e.g. by a show of hands. The teacher records the majority ratios on the board.
4. A joint decision is made based on the majority ratios for both sets of partial data ("rather recommendable" or "rather not recommendable")
5. The teacher verbalizes the decision rule ("If a food has 350 kcal or less, we classify it as ...") and notes it on the board as a one-step decision tree.
6. The number of foods that are incorrectly classified in both data sets is counted. As this number is still quite high, the teacher formulates the goal of finding a new decision rule with which fewer foods are classified incorrectly. For this purpose, a new threshold value is proposed (by the teacher or pupils) for the energy feature, e.g. 300. Points 2-5 are repeated for the new threshold value.
7. A judgment is made as to which data split provides the better decision rule. The criterion for this is the respective number of misclassified foods. The technical term for this is the number of misclassifications. It is up to the teacher whether this technical term is used or only paraphrased.
8. Points 6-7 are carried out again for a third threshold value.

Possible table image for the 230 kcal threshold value:  


Background: Establish decision rules

The aim of the lesson series is to create a multi-level rule system for classifying food. In this lesson, the pupils first learn to derive decision rules (single-level decision trees) from the data. This is implemented with the concept of data splitting, which is introduced in this lesson. The data cards are divided into two subgroups based on a characteristic and a so-called threshold value (e.g. foods with up to 10 g fat and over 10 g fat). In both sub-groups, it is then determined whether the majority is recommended or not recommended. If there are different labels in the subgroups (which is the case in the vast majority of cases), there are foods in both subgroups that deviate from the majority decision. These are referred to as errors or misclassifications. The aim is therefore to find the threshold value that produces as few errors as possible (number of misclassifications). In this lesson, the decision rules are created using the energy feature as an example.

Development of the objective for the creation of a decision rule:

[Presentation 2](https://unterrichtsmaterial-ddi.cs.upb.de/images/1/14/Pr%C3%A4sentation2_Datensplit_Einf%C3%BChrung.pptx) can be used to work out the objective when formulating a decision rule. The optimal case (objective) is when a threshold value is found so that on one side of the threshold value there are only foods that are not recommended and on the other side only foods that are recommended. Such a "perfect" rule, with which no food is misclassified from the available training data, is usually not found, but one tries to come as close as possible so that as few foods as possible are misclassified.