

Info sheet for the article "Classifying food with decision trees" (published in: mathematik lehren 244 (2024))

Susanne Podworny, Yannik Fleischer & Rolf Biehler

The Nutri-Score

The Nutri-Score can be used to compare foods with a label (A to E). This is intended provide consumers with simple guidance when shopping including the mandatory nutritional information and other ingredients. The Nutri-Score is intended to help consumers to identify the product with the most favorable nutritional composition from several products in the same product group (e.g. different types of muesli).

The labeling

The letters A to E and additional traffic light colors indicate the quality of food in five levels. The letter A on a green background (illustration) stands for a comparatively favorable the letter E on a red background stands for a comparatively unfavorable composition of nutritional values and other ingredients.



Example of a Nutri-Score label with the letter A (courtesy of RAL LOGO LIZENZ)

Determination of the Nutri-Score

Based on a scientifically developed calculation model of the Scientific Committee of the Nutri-Score (ScC), the ingredients of foods are given points. The fewer points a food receives, the better its Nutri-Score. Here, "favorable ingredients" such as nuts, fiber, proteins, fruits/vegetables, legumes and selected oils are negative points (from -5 to 0), while "unfavorable ingredients" such as sugar, salt, saturated fatty acids and energy content are positive points (from 0 to 10). Once the total number of points has been calculated, a food is classified according to the total number of points as shown in the following table:

Calculated total score (Food)	Classification in Nutri-Score
Min. to -1	A
0 to 2	B
3 to 10	C
11 to 18	D
19 and more	E

Applying the algorithm to, for example, pretzel sticks with the nutritional values shown in the picture (and assuming a fiber content of 2.9 g/100 g) results in the following scoring according to the document "Conditions of use of the trademark 'Nutri-Score'"¹.

Feature	Amount of nutrients in pretzel sticks (per 100 g)	value is in the Range	Received Points
Energy	392 kcal (1640 KJ)	> 1340; ≤ 1675	4
Saturated fatty acids	0,8 g	≤ 1	0
Sugar	3,4 g	≤ 4,5	0
Salt	4.5 g (1800 g sodium)	> 900	10
Protein	12 g	> 8	-5
Dietary fiber	2,9 g	> 2.8; ≤ 3.7	-3
Fruit/vegetable content	0%	≤ 40	0

The calculation of the total number of points then depends on whether the sum of the positive points (i.e. the points for energy, saturated fatty acids, sugar and salt) is 11 or more.

- If the positive point total is less than 11, then all negative points (protein, fiber and fruit/vegetable content) are deducted.
- If the positive point total is 11 or more, then only the negative points for fiber and fruit/vegetable content are included; protein is not included in the calculation.

In our case, we have 14 positive points, so the total number of points is calculated from the sum of the positive points and the sum of the points for fiber and fruit/vegetable content. The result for the pretzel sticks is therefore

Total points: 4+ 10 -3= 11 points

According to the table above, 11 points are to be classified with the letter D. This Nutri-Score classification can actually be found on many pretzel stick packages.

¹ The exact allocation of points depending on the respective nutritional content can be found in the document "Conditions of use of the trademark 'Nutri-Score'" from p. 23:

https://www.santepubliquefrance.fr/content/download/150258/file/Nutriscore_reglement_usage_EN_240423.pdf



Salzstangen	
	
Nährwerte pro 100g	
Energie	392 kcal
Fett	5,9 g
davon gesättigte Fettsäuren	0,8 g
Kohlenhydrate	71,0 g
davon Zucker	3,4 g
Eiweiß	12,0 g
Salz	4,5 g

On the website of the Federal Ministry of Food and Agriculture (BMEL) you can find an Excel table² that makes it easier to calculate the Nutri-Score of foods. The exact score can either be calculated in detail using the corresponding document or with the help of the Excel table, in which it is sufficient to enter the values of the food in order to obtain the Nutri-Score.

All information on the Nutri-Score states that only foods in the same product group (e.g. different types of pizza or chocolate) should be compared³. The official documentation distinguishes between four groups (general foods, cheese, oils & fats & nuts, and drinks), so the same formula is used for the vast majority of "general foods". However, comparing products from different product groups seems to lead to results that could be confusing, so the restriction comparison within product groups was added, although the same formula applies to all.

² <https://www.bmel.de/DE/themen/ernaehrung/lebensmittel-kennzeichnung/freiwillige-angaben-und-label/nutri-score/nutri-score-labeling-help-positions.html#doc73490bodyText8>

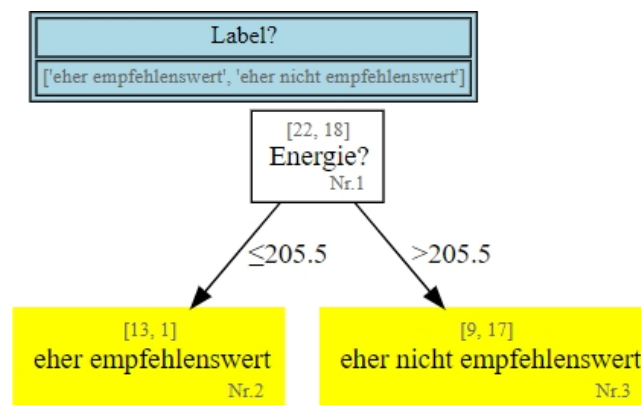
³ E.g. here: <https://www.bmel.de/DE/themen/ernaehrung/lebensmittel-kennzeichnung/freiwillige-angaben-und-label/nutri-score/nutri-score-erklart-verbraucherinfo.html> under "What the Nutri-Score actually "

The recommendation by a decision tree of the ProDaBi lesson series "Apple or popcorn?"

In the following, the series of lessons and the decision trees created there for recommending foods related to the Nutri-Score. This helps teachers establish a reference to the Nutri-Score, which may be familiar from everyday life. We do not intend to put these two approaches on the same level, as the decision trees created in class are of course not nearly as well-founded. Nevertheless, we would like to emphasize the value that the creation of decision trees can have for learning about the context of food.

Similar to the Nutri-Score, the data-based creation of a decision tree during the series of lessons classifies foods based on nutritional information (per 100g). The recommendation issued is only two-stage (rather recommended or rather not recommended). An important difference in the approach is that the Nutri-Score uses scientific knowledge about the various ingredients to assess and evaluate their influence, while the series of lessons follows a data-based approach in the sense of supervised machine learning. In other words, an assessment is made of a set of foods that we "obviously" consider to be rather recommendable (e.g. cucumber) or rather not recommendable (e.g. potato chips). An attempt is then made to create a recommendation system with the help of these so-called training examples with labels (assessment) and their nutritional information. Only the mandatory nutritional information for foods, which can be found on all packaging, is used as the data basis. Other data that cannot be read directly from the nutritional value table, such as dietary fiber, are not included in the series of lessons.

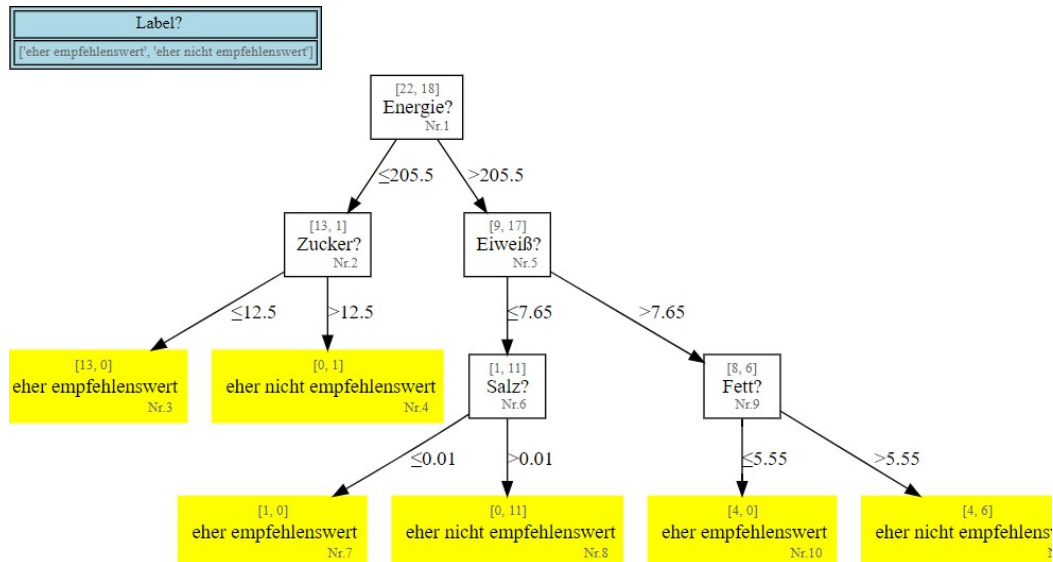
The data is then used to systematically search for decision rules that correctly classify as many training examples as possible (procedure: see article). First, single-level decision trees are sought, as in the following figure, which recommendations based on one feature. The best rule for a sample data set of foods is based on the feature energy and classifies 30 out of 40 training examples correctly.



The rule is as follows: if energy is less than or equal to 205.5 kcal per 100 g, then it is recommended, if energy is greater than 205.5 kcal per 100 g, then it is not recommended. This is of course very simplistic and therefore 10 out of 40 foods are classified incorrectly. However, it does give an impression of how the energy characteristic should be viewed.

In the course of the series of lessons, the decision trees become more complex by adding further levels in order to classify as few training examples as possible incorrectly.

Good students arrive at three-stage decision trees which, as shown in the following image, already provide relatively detailed insights and include several particularly influential nutritional values. For example, sugar is still considered for foods with a low energy value and if the sugar content high, the food is rated as not recommended despite its low energy content.



This means that the content of the trees can be interpreted after they have been created in order to derive information about the individual nutritional values. Our pretzel stick from above is classified here in the far right-hand path of the tree (392 kcal, 12 g protein, 5.9 g fat) and therefore not recommended.

For this method to work, it is crucial that the training examples are given a meaningful label (recommended, not recommended) at the beginning. This is not obvious for all possible foods, so it may make sense to put food aside if you are unsure about the label. In the series of lessons, students initially assign labels to foods subjectively based on their assessments. In order to reach a consensus, the labels are discussed in class with the teacher and there is an opportunity to compare the labels with a suggestion included in the lesson series. Food can also be set aside if there is disagreement. In this way, awareness of the topic of nutrition is raised the beginning of the series of lessons and discussed in class.

Focus of the ProDaBi lesson series "Apple or popcorn?"

The focus of the series of lessons is data-based decision-making with so-called decision trees, as a first approximation to machine learning. To this end, both procedures for creating decision trees and aspects of the context of nutrition are simplified for teaching purposes.

The context of *food* is not typical for the field of machine learning, but is suitable for connecting to the world of experience of all students (regardless of age, gender, etc.) and

is therefore easily accessible. The recommendation system created can be representative of many recommendation systems that are familiar from everyday life (e.g. for products on online platforms). There are possible links to biology lessons, for example, and the treatment of the context can contribute to general education lessons.

Classification of the lesson series in relation to the Nutri-Score

As with the Nutri-Score, the resulting decision trees do not provide a rule system for a recommended diet as a whole; it is only ever about individual foods.

This means that just because a food is classified as recommended by the decision tree does not mean that you should only eat it. Nor does a classification not recommended mean that you should never eat this food. You can only very carefully deduce which foods you should include in your diet more often and which less often. Neither Nutri-Score nor the decision trees provide any information about allergies or other intolerances or suitable portion sizes. With the Nutri-Score, an A can be awarded for a food that you should eat a small amount of, and the same applies to the classification as recommended by a decision tree. Careful reflection on the classifications is necessary for both the Nutri-Score and the decision trees in the lesson series. However, both the decision trees and the Nutri-Score make it possible to obtain indications of foods that tend to be "more recommendable". The Nutri-Score is officially restricted to comparisons within product groups.

Classifications can be easily understood with the help of the decision trees, as the individual criteria are directly visualized in the decision tree (e.g. from what amount of fat and what amount of salt is classified as rather recommendable or not). When discussing various decision trees at the end of the series of lessons, characteristics can also be identified (e.g. energy, fat or sugar) that often have a negative or positive effect on the recommendation of a food.

In the Nutri-Score, the positive and negative components are weighted differently with up to -5 and up to 10 points respectively. This weighting remains invisible, however, as long as the background to the Nutri-Score is not very closely. No such weighting is carried out with the decision trees.

Nevertheless, it should be noted that a scientific committee works behind the Nutri-Score and the categorization by the Nutri-Score is based on scientific findings and a somewhat broader database than the decision trees in the series of lessons. Nevertheless, the decision trees give tendencies that often go in similar directions to the Nutri-Score. However, due to the classification by the decision trees of the lesson series, one should not be tempted to be tempted to of "healthy" or "unhealthy" foods, because they cannot make this recommendation.

Assigning labels to food has a major influence on whether the decision tree based on it can classify food in a meaningful way. This can be explicitly addressed in the series of lessons, for example by deliberately labeling the data in a less meaningful way when automatically creating the decision trees in the web application and discussing the resulting decision trees. In this way, weaknesses of the data-based approach can also be uncovered.