

Аверина Е.С.

*Научный руководитель: ст. преподаватель К.О. Мацкевич
Муромский институт (филиал) федерального государственного образовательного
учреждения высшего образования «Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»
602264, г. Муром, Владимирская обл., ул. Орловская, 23
email: katerina_140900@mail.ru*

Machine learning method - decision trees

I'm going to tell you about one of the machine learning methods which are called decision trees method. The last decade has seen considerable growth in interest in Artificial Intelligence and Machine Learning. First, let's decipher at what is machine learning. Machine Learning (ML) is a branch of the artificial intelligence theory, the subject of which is the search for solving tasks methods by learning in the process of solving similar tasks.

The decision trees method is one of the most popular methods for solving classification and forecasting tasks. A decision tree is a set of rules ("If – then") that can be combined into some hierarchy.

The structure of the tree consists of "leaves" and "branches":

- The edges ("branches") of the solution tree contain the attributes which the target function depends on;
- The "leaves" contain the values of the target function;
- The other nodes contain the attributes by which the cases differ.

Classifying a new case is going down the tree to the desired leaf and determining the appropriate value. To understand better what a decision tree represents, we can consider a funny example. Let's say we need to answer the question: "Go to a lecture?".

To solve a task, that is, to make a decision whether to go to a lecture, one should classify the current situation as one of the well-known classes - in this case, "Go" or "Don't go". This requires answering a number of questions that are located in the nodes of this tree, starting from its root (Figure 1).

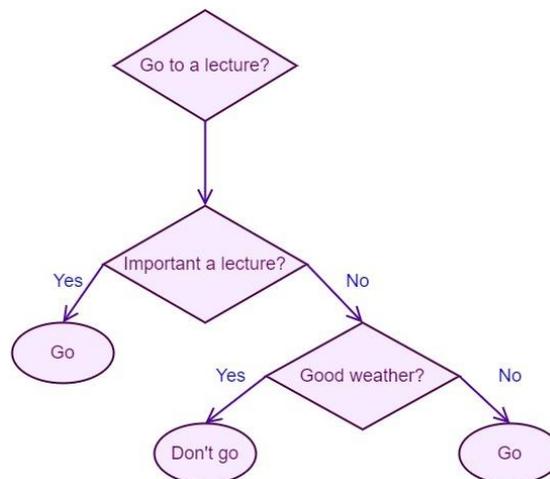


Figure 1: Example of a decision tree.

Well, it becomes clear that the top question is the root node, the answer options "Yes" or "No" are branches and our classes "Go" or "Don't go" these are leaves.

Now let's consider at a more complex real-world example. So, we will have a look what the decision tree looks like using the famous Iris dataset. By the way, this dataset is almost 85 years old. It has already become classical and is often used in the literature to illustrate the work of various statistical algorithms.

If we talk more detail about this set, it is worth noting that the Iris Fisher dataset, or simply the iris flower dataset, is a multidimensional dataset that biologist Ronald Fisher introduced in 1936 his article as an example of linear discriminate analysis. Just for the record the dataset consists of 150

samples whose characteristics are the length and width of sepals and petals in centimeters. With the help of these samples, you can determine to which class a particular Iris belongs to. Dataset includes Iris species such as Iris setosa, Iris virginica and Iris versicolor (Figure2). Using these four characteristics, Fischer was able to develop a linear discriminate model to be able to classify and distinguish species from each other (Figure3).

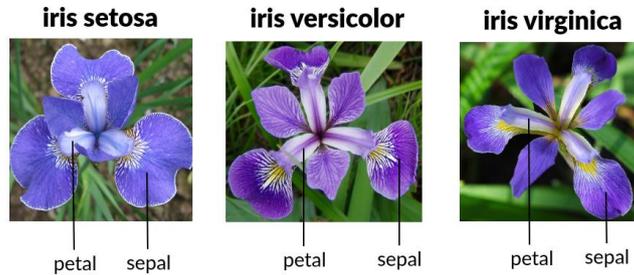


Figure2: Types of iris.

sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3.0	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
5.0	3.6	1.4	0.2	Iris-setosa
7.0	3.2	4.7	1.4	Iris-versicolor
6.4	3.2	4.5	1.5	Iris-versicolor
6.9	3.1	4.9	1.5	Iris-versicolor
5.5	2.3	4.0	1.3	Iris-versicolor
6.5	2.8	4.6	1.5	Iris-versicolor
6.3	3.3	6.0	2.5	Iris-virginica
5.8	2.7	5.1	1.9	Iris-virginica
7.1	3.0	5.9	2.1	Iris-virginica
6.3	2.9	5.6	1.8	Iris-virginica
6.5	3.0	5.8	2.2	Iris-virginica

Figure3: Few part of the Iris dataset.

Well, based on the Iris dataset, a decision tree is built, the purpose of which is to divide the dataset, where there is only one type of Iris, in accordance with the specified features: here, the width of the petal (Figure4).

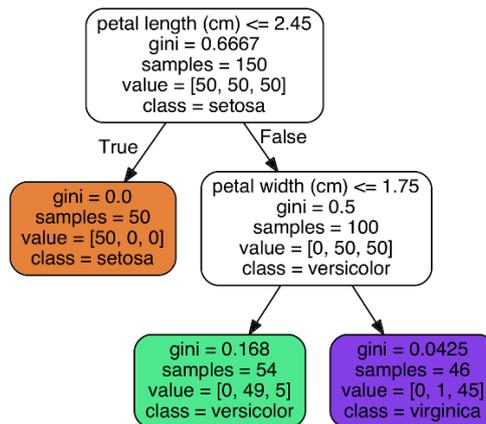


Figure4: Example of a decision tree.

In conclusion, it is necessary to underline that in data mining, a decision tree describes data and as we said before decision trees are the most popular machine learning algorithms because they are intelligible and simple.