KNIME - Decision tree

Let us have data, generated by logical implication $Imp\,(A, B) \equiv A \Rightarrow B$, i.e.

A	B	Imp
fA	fB	t
fA	tB	t
tA	fB	f
tA	tB	t

where, for lucidity, we distinguish logical values of A, B and the result Imp.

Then, for the sequence of column selection A, B we get the sub-tables:

A =fA

A	B	Imp
fA	fB	t
fA	tB	t

A =tA

A	B	Imp
tA	fB	f
tA	tB	t

From the first table we can see, that the result is unambiguous: For A =fA we have Imp =t (independently of B)

The second table results in both f and t in dependence on the variable B. So here we have to continue

B =fB

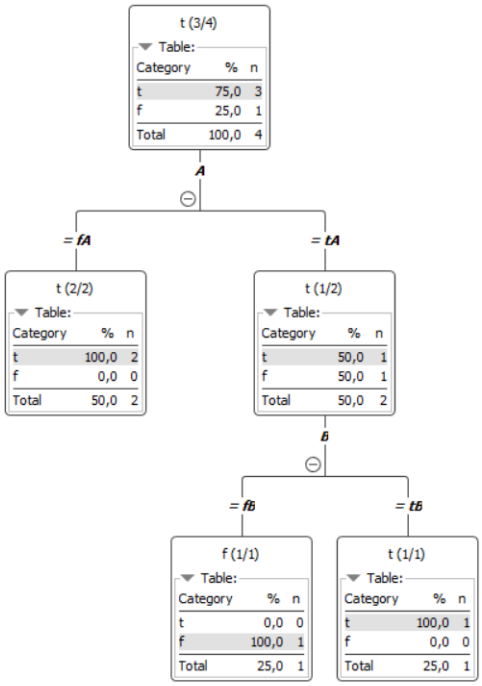
B	Imp
fB	f

B =tB

B	Imp
tB	t

and here we see, that for B =fB we have Imp =f and for B =tB we have Imp =t; both unambiguous.

This is also expressed in the tree, from KNIME (Decision three learner/Decision tree view)



Here we can follow the subsequent division of the table:

The first rectangle shows the situation, when no information from A and B is taken into account. Here, we can see the number of t and f in the column Imp. From it follows, that without any other information we would guess the value of Imp to be t - this is indicated in the top of the upper rectangle.

Remark: This means $\begin{array}{c|cc} Imp & \text{t} & \text{f} \\ \hline f\left(Imp\right) & \frac{3}{4} & \frac{1}{4} \end{array}$ and $\hat{Imp}$ =t.

Now we take into account the variable A - i.e. we have $f\left(Imp|\text{A}\right)$. This situation is reflected by the second layer of the tree.

Here: the left rectangle is finished (in the table we have 100%) but the right one is still not ready, no 100% occurs. So, here we must still continue.

Taking into account also the variable B we treat $f\left(Imp|A,B\right)$, specifically $f\left(Imp|A=tA,B\right)$. This is in the third layer of the tree.

Result

The result for the measured values of A and B is obtained in the top of the rectangle we get to using the measured value - they are the rectangles in the bottom of the tree.

This result can also be obtained from the Decision tree to ruleset as follows

☐	Row1	A = "fA" AND TRUE => "t"
☐	Row2	B = "fB" AND A = "tA" => "f"
☐	Row3	B = "tB" AND A = "tA" => "t"

The whole KNIME program is here