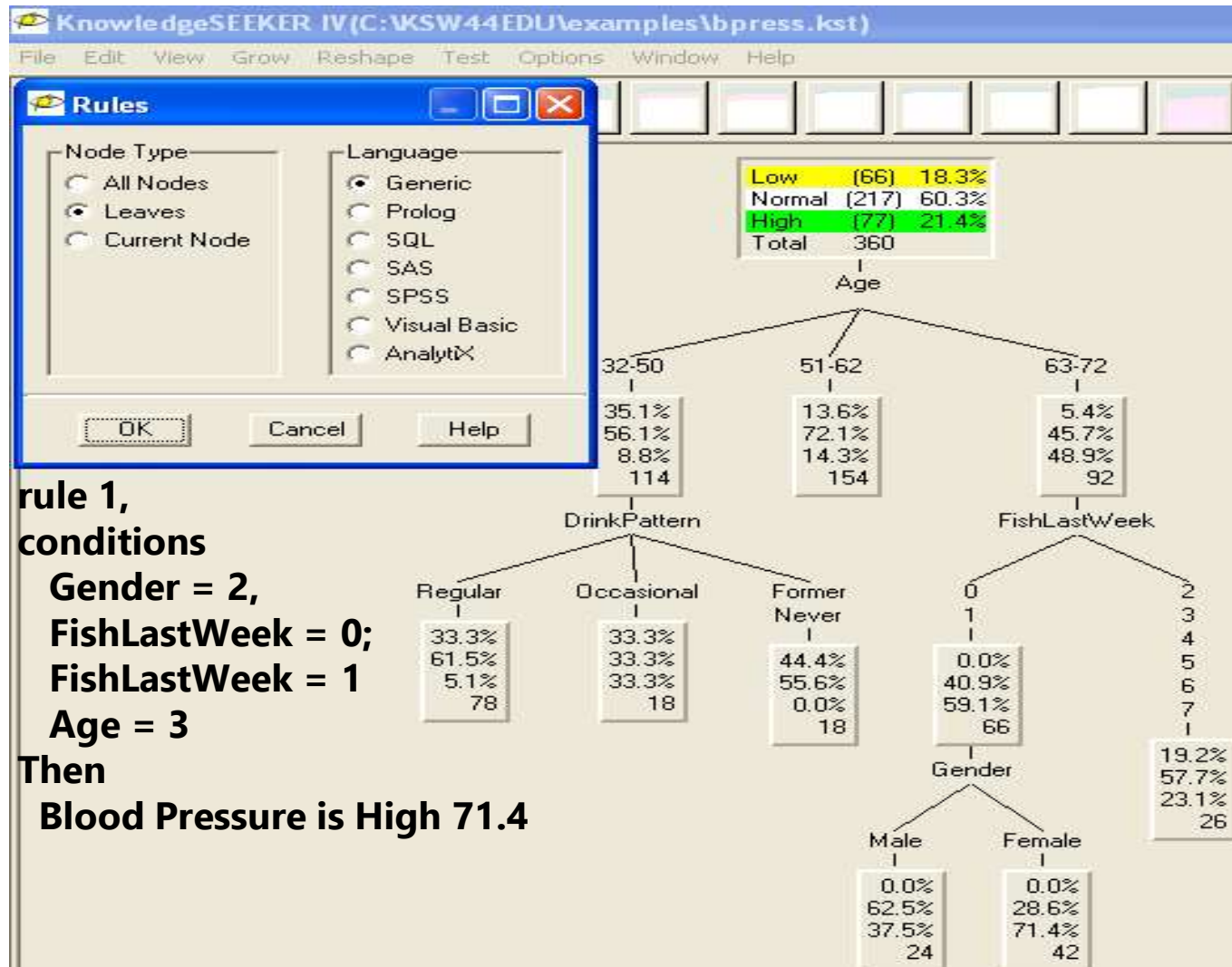


Decision Tree Demo and Example 14.5.21

The screenshot shows the KnowledgeSEEKER IV software interface. The main window is titled 'View Data' and contains a table with 15 rows and 10 columns. The columns are labeled: Pattern, DrinksEveryDay, Age, YearsEducation, Income, Gender, Weight, Height, zscore, and Hypertension. The 'Hypertension' column is highlighted in red. The data is as follows:

Pattern	DrinksEveryDay	Age	YearsEducation	Income	Gender	Weight	Height	zscore	Hypertension
1	2	3	2	16	1	141	626	4.530000	2
2	2	2	2	9	1	187	635	5.030000	3
3	1	2	2	11	1	191	707	3.840000	2
4	9	2	2	3	1	160	657	4.430000	2
5	2	2	2	15	1	216	709	4.430000	2
6	2	1	2	12	1	206	720	2.760000	1
7	9	3	2	9	2	215	632	4.900000	3
8	9	3	5	10	2	89	641	4.510000	2
9	2	1	2	11	1	188	679	3.750000	2
10	9	1	2	16	2	284	677	3.680000	2
11	9	3	5	10	2	150	626	4.370000	2
12	2	2	1	13	1	180	649	4.150000	2
13	2	1	2	7	2	167	619	4.320000	2
14	2	1	2	16	1	206	714	3.480000	2
15	1	1	2	9	1	182	746	4.190000	2

Decision Tree Tool



Decision Tree Example

Obj. Nbr	Class	Var ₁	Var ₂	Var ₃	Var ₄	Var ₅
2	1	1	2	0	2	51
4	1	0	5	0	0	63
6	1	0	3	0	0	32
7	2	6	1	0	0	51
7	2	6	1	0	0	51
7	2	6	1	0	0	51
8	2	4	1	0	1	63
8	2	4	1	0	1	63
8	2	4	1	0	1	63
9	2	1	1	2	0	51
Ent		3.25			6.11	

Im Node 1 wird der Best Split bei Var₁, Var₂, Var₃, Var₄ und Var₅ gesucht. Um einfach zu tun nehmen wir Var₁ und Var₄.

Var₁ hat den kleineren „Entropie“-Wert

Var₁ ist der best Split.

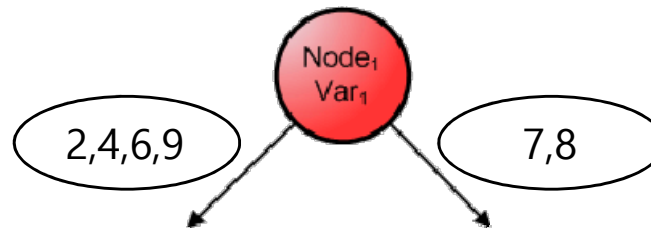
$$\text{"Entropie"} = -p_+ \ln\left(\frac{p_+}{n}\right) - p_- \ln\left(\frac{p_-}{n}\right)$$

p₊ = Anz. richtig klassifizierte

p₋ = Anz. falsch klassifizierte

n = Anzahl Objekte

Node 1



Rule 1: Wenn Var 1 \leq 2.5 dann Class = 1 , Gehe zu Node 2
Die Daten werden mit der Rule 1 aufgeteilt.

Node 2

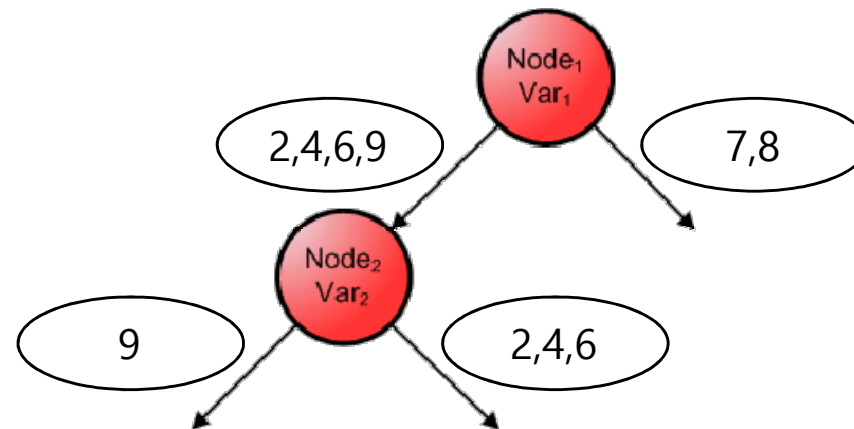
Obj. Nbr	Class	Var ₁	Var ₂	Var ₃	Var ₄	Var ₅
2	1	1	2	0	2	51
4	1	0	5	0	0	63
6	1	0	3	0	0	32
9	2	1	1	2	0	51
Ent		2.25	0			

Im Node 2 wird der Best Split bei Var₁ und Var₂ gesucht.

Var₂ hat den kleineren „Entropie“-Wert

Var₂ ist der best Split.

Node 2



Rule 2: Wenn $Var_2 \leq 1.5$ dann Class = 2, Gehe zu Node 4
Die Daten werden mit der Rule 2 aufgeteilt.
Usw....bis ganze Baum ist erstellt.