

1) Zoo Datensatz:

Correctly Classified Instances 100 99.0099 %

Incorrectly Classified Instances 1 0.9901 %

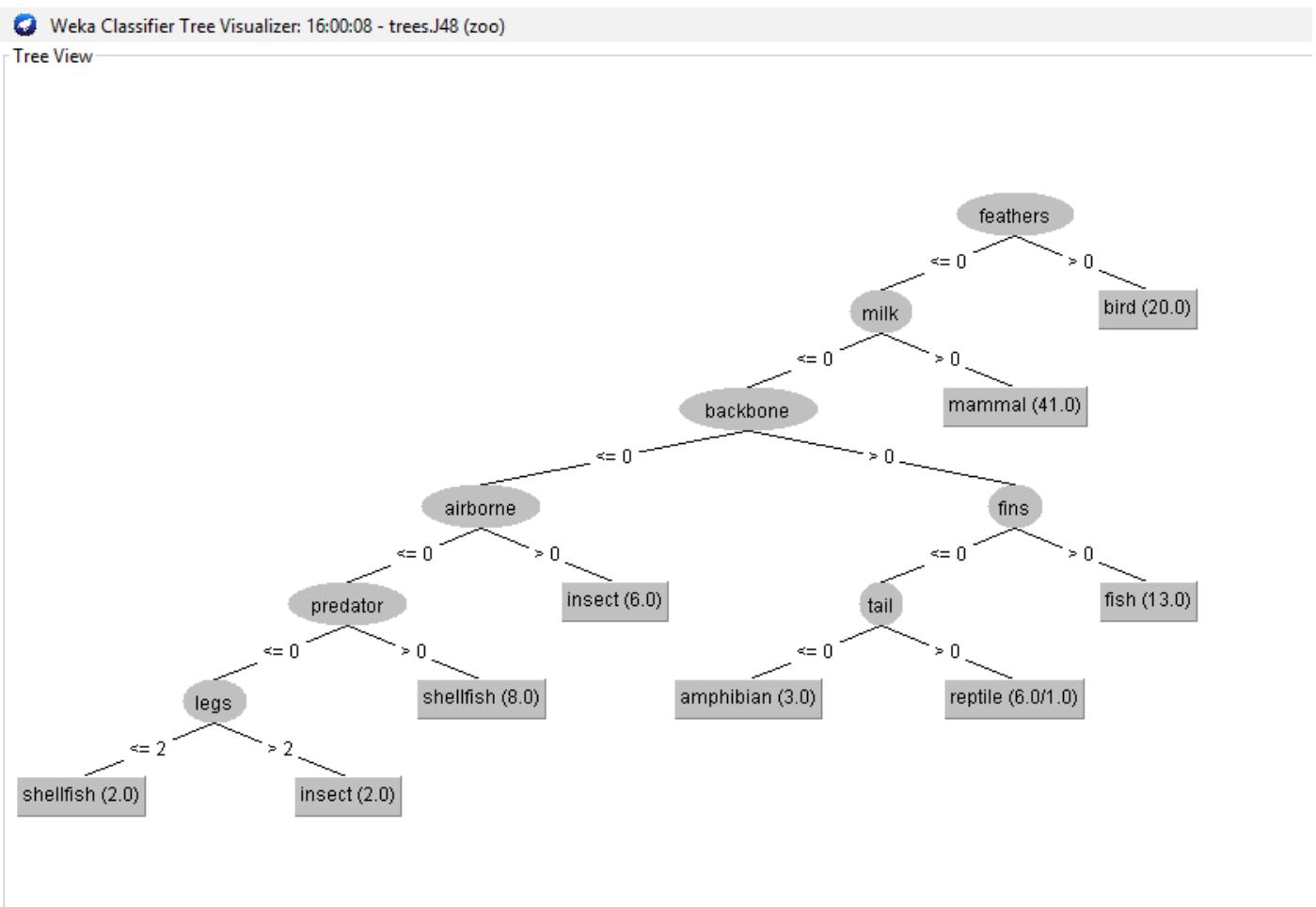
== Confusion Matrix ==

a	b	c	d	e	f	g	<-- classified as
41	0	0	0	0	0	0	a = mammal
0	13	0	0	0	0	0	b = fish
0	0	20	0	0	0	0	c = bird
0	0	0	10	0	0	0	d = shellfish
0	0	0	0	8	0	0	e = insect
0	0	0	0	0	3	1	f = amphibian
0	0	0	0	0	0	5	g = reptile

Die Matrix zeigt wie viele Einträge der CSV Datei richtig bzw. falsch klassifiziert wurden.

Hier wurden alle Klassen korrekt erkannt, bis auf eine Amphibie, die als Reptil erkannt wurde.

Der Baum der erzeugt wurde:



Restaurant-Datensatz:

Correctly Classified Instances 8 66.6667 %

Incorrectly Classified Instances 4 33.3333 %

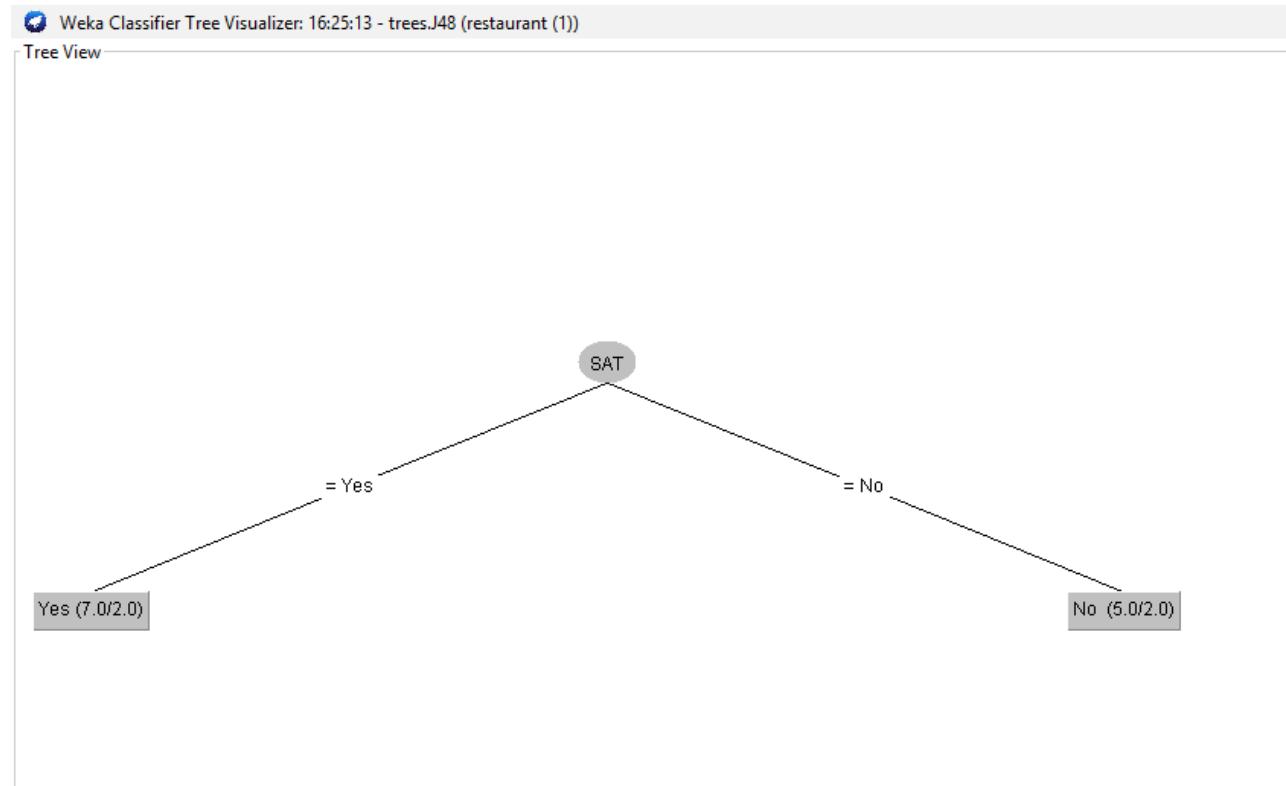
== Confusion Matrix ==

a b c <- classified as

5 0 1 | a = Yes

1 0 1 | b = No

1 0 3 | c = No



Die Klassifizierung finde ich merkwürdig. Ich würde ein Restaurant nach dem Typ von Essen klassifizieren, das angeboten wird, wie Italian oder Burger. Aber hier wird anscheinend auf die Entscheidung hingearbeitet, ob man wartet.

ARRFF-Format

String ist ein Text, Buchstaben-Datentyp. Man kann Attribute mit beliebigen Textwerten erstellen.

Numeric sind ganze oder reelle Zahlen.

Nominal benennt die Klasse, es ist eine Liste mit allen möglichen Klassen.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **Id3**

Test options

- Use training set
- Supplied test set
- Cross-validation Folds
- Percentage split %
-

(Nom) type

Result list (right-click for options)

18:45:13 - trees.Id3

Classifier output

	101	100	%
Correctly Classified Instances	101	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Relative absolute error	0		
Root relative squared error	0		
Total Number of Instances	101		

==== Detailed Accuracy By Class ====

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	mammal
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	bird
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	reptile
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	fish
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	amphibian
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	insect
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	shellfish
Weighted Avg.	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	

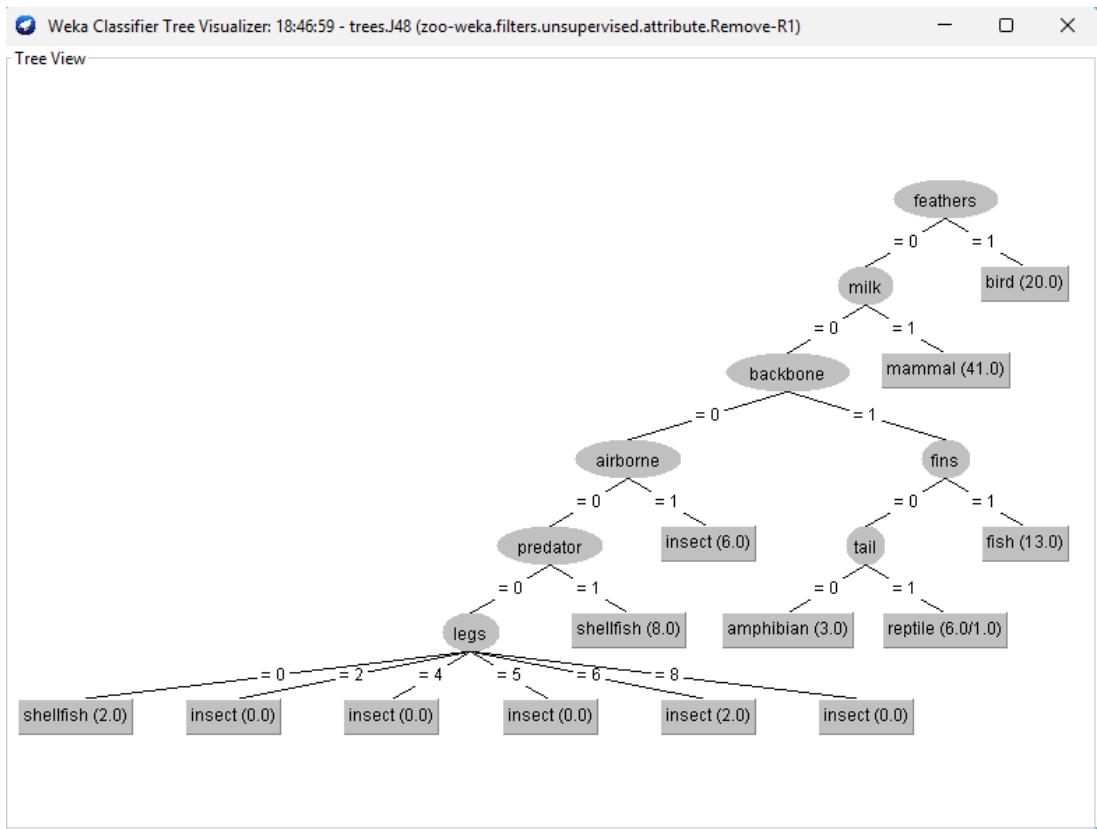
==== Confusion Matrix ====

a	b	c	d	e	f	g	<-- classified as
41	0	0	0	0	0	0	a = mammal
0	20	0	0	0	0	0	b = bird
0	0	5	0	0	0	0	c = reptile
0	0	0	13	0	0	0	d = fish
0	0	0	0	4	0	0	e = amphibian
0	0	0	0	0	8	0	f = insect
0	0	0	0	0	0	10	g = shellfish

Status OK

Ich habe versucht, den Baum mit ID3 zu generieren. Ich habe dazu nur nominale Datentypen genutzt, dann hat er funktioniert. Den Baum konnte ich aber nicht generieren.

Mit J48 konnte ich den Baum generieren, er zeigt jetzt alle Ausprägungen an und ist viel detaillierter. Ich



Ich musste den Namen entfernen, da in diesem Format oder mit diesen Bäumen anscheinend nicht in String gearbeitet werden kann. Nur Numeric oder Nominal. Dasselbe gilt für den Restaurant-Datensatz. Ich hab versucht, ihn mit ID3 darzustellen, konnte den Baum aber nicht generieren.

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **Id3**

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 10

Percentage split % 66

More options...

(Nom) Wait

Start Stop

Result list (right-click for options)
18:53:13 - trees.Id3

Classifier output

```

| Res = Yes: Yes
| Res = No: No

Time taken to build model: 0 seconds

==== Evaluation on training set ====

Time taken to test model on training data: 0 seconds

==== Summary ====

```

Correctly Classified Instances	12	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Relative absolute error	0	%	
Root relative squared error	0	%	
Total Number of Instances	12		

```

==== Detailed Accuracy By Class ====

```

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	Yes
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	No
Weighted Avg.	1,000	0,000	1,000	1,000	1,000	1,000	1,000	

```

==== Confusion Matrix ====

```

```

a b  <- classified as
6 0 | a = Yes
0 6 | b = No

```

Status OK Log x0

J48 hat diesmal mehr erkannt, als bei dem csv Format:

Correctly Classified Instances 11 91.6667 %

Incorrectly Classified Instances 1 8.3333 %

Die Confusion Matrix hat diesmal zwei Werte, wie es sein soll. Nicht drei (a,b,c) wie bei der csv.

a b <-- classified as

5 1 | a = Yes

0 6 | b = No

Der Baum sieht aber fast gleich aus

Weka Classifier Tree Visualizer: 18:54:23 - trees.J48 (restaurant)

Tree View

