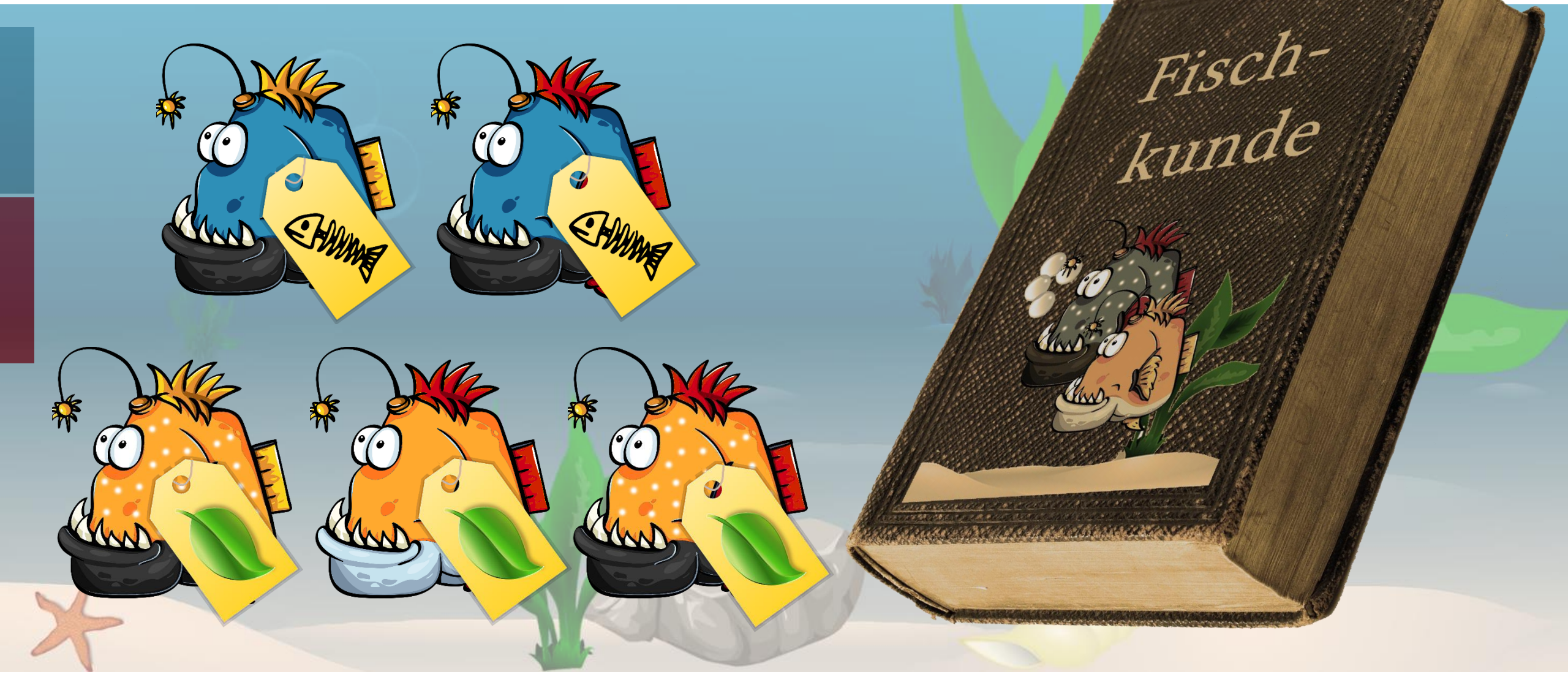
EMIL-VON-BEHRING GYMNASIUM

Künstliche Intelligenz – Entscheidungsbaum

Fische – friedlich oder feindselig?

Fische – friedlich oder feindselig?



Wie würdet ihr euch entscheiden?

Ist dieser Fisch bissig oder nicht?
Versuchen Sie dies aus den anderen Fischen abzuleiten?

Entscheidungsbaum aus Trainingsdaten

Trainingsdaten

- Aus den Trainingsdaten soll nun automatisiert ein Entscheidungsbaum aufgebaut werden
- Aber wie? Nach welchem Attribut unterscheidet man zuerst?

Entscheidungsbaumalgorithmus im Detail

Trainingsdaten

Muster

Flossenfarbe

Bauchfarbe

Schuppenfarbe

Tabellarische Notation: Alles auf einen Blick

Trainingsdaten

Muster

Muster			
Fehler vor Aufteilen der Datenmenge:			4 Fehler
Attributwert / Label	*Friedlich*	*Feindselig*	Fehler
Ohne	2	2	2 Fehler
Punkte	2	3	2 Fehler
		Gesamt:	4 Fehler
		Informationsgewinn:	4 Fehler – 4 Fehler = 0 Fehler

Aufteilen nach Attribut „Flossenfarbe“

Trainingsdaten

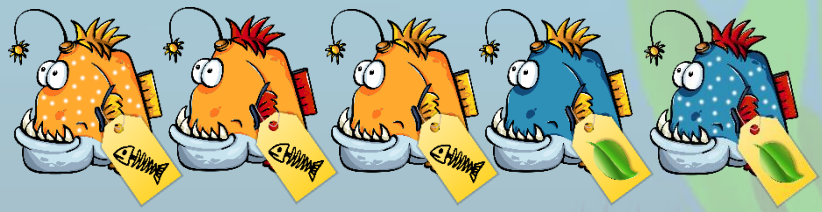
Flossenfarbe

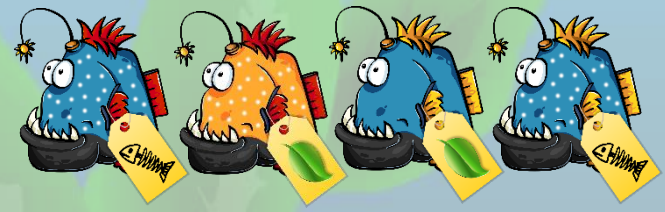
Flossenfarbe			
Fehler vor Aufteilen der Datenmenge:			4 Fehler
Attributwert / Label	*Friedlich*	*Feindselig*	Fehler
Gelb	2	3	2 Fehler
Rot	2	2	2 Fehler
		Gesamt:	4 Fehler
		Informationsgewinn:	4 Fehler – 4 Fehler = 0 Fehler

Aufteilen nach Attribut „Bauchfarbe“

Trainingsdaten

Bauchfarbe



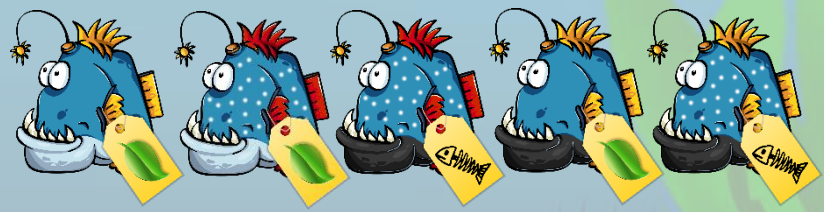


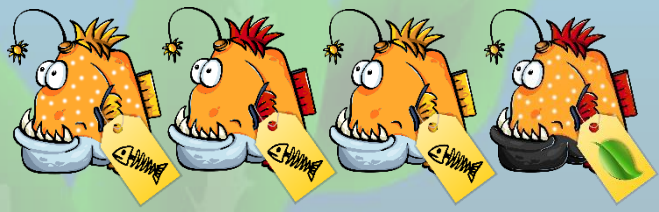
Bauchfarbe			
Fehler vor Aufteilen der Datenmenge:			4 Fehler
Attributwert / Label	*Friedlich*	*Feindselig*	Fehler
Weiß	2	3	2 Fehler
Schwarz	2	2	2 Fehler
		Gesamt:	4 Fehler
		Informationsgewinn:	4 Fehler – 4 Fehler = 0 Fehler

Aufteilen nach Attribut „Schuppenfarbe“

Trainingsdaten

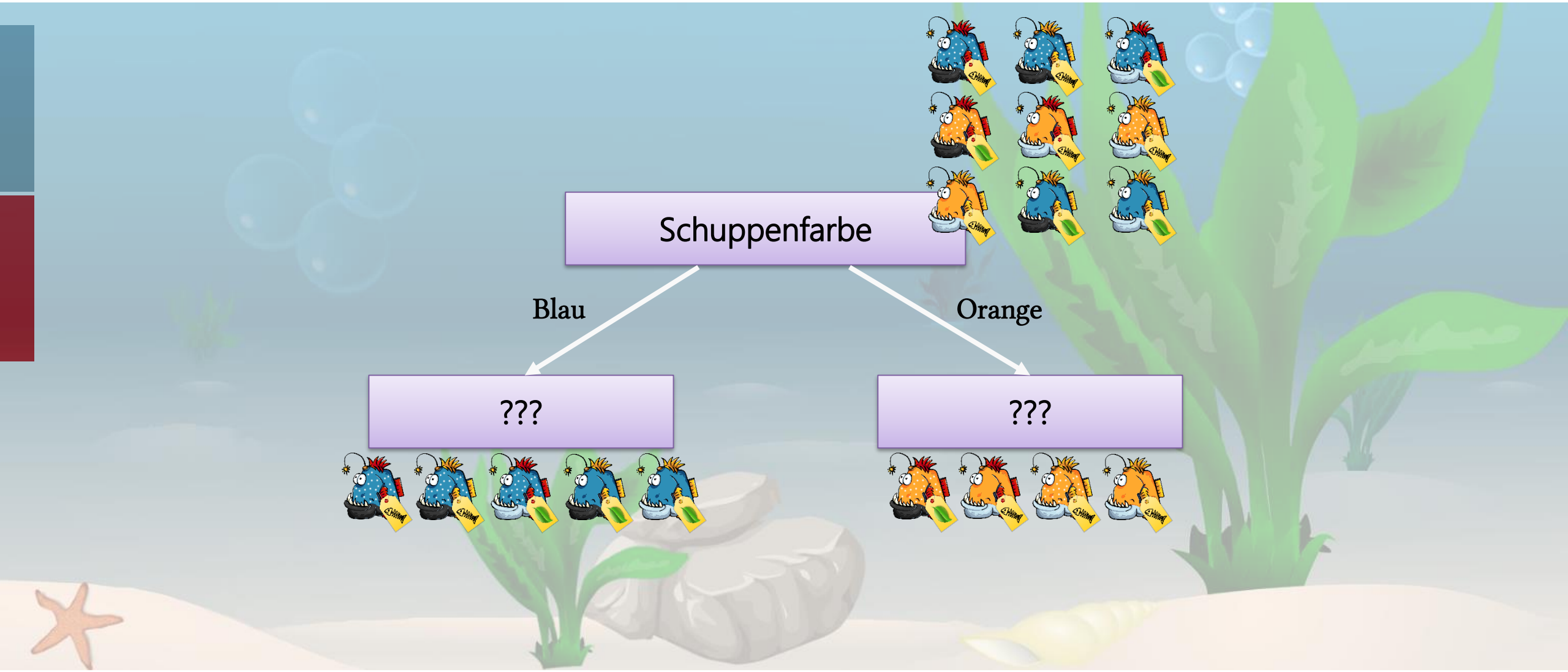
Schuppenfarbe



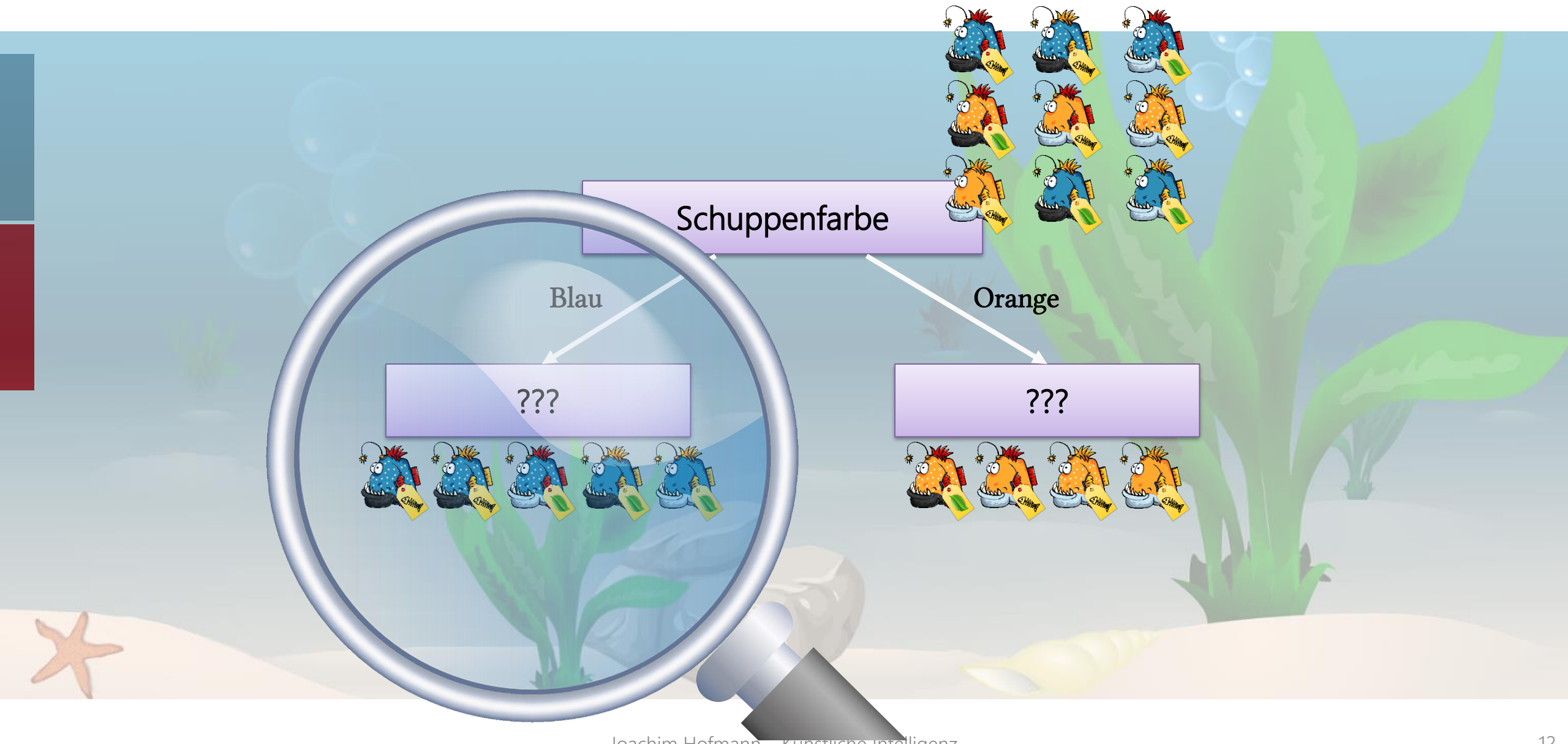


Schuppenfarbe			
Fehler vor Aufteilen der Datenmenge:			4 Fehler
Attributwert / Label	*Friedlich*	*Feindselig*	Fehler
Blau	3	2	2 Fehler
Orange	1	3	1 Fehler
		Gesamt:	3 Fehler
		Informationsgewinn:	4 Fehler – 3 Fehler = 1 Fehler

„Wichtigstes Attribut“: Schuppenfarbe



„Wichtigstes Attribut“: Schuppenfarbe



Suche nach dem nächsten „wichtigsten“ Attribut

Teilmenge der Trainingsdaten

Muster

Muster			
Fehler vor Aufteilen der Datenmenge:			2 Fehler
Attributwert / Label	*Friedlich*	*Feindselig*	Fehler
Ohne	2	0	0 Fehler
Punkte	1	2	Fehler
		Gesamt:	1 Fehler
		Informationsgewinn:	2 Fehler – 1 Fehler = 1 Fehler

Suche nach dem nächsten „wichtigsten“ Attribut

Teilmenge der Trainingsdaten

Flossenfarbe

Flossenfarbe			
Fehler vor Aufteilen der Datenmenge:			2 Fehler
Attributwert / Label	*Friedlich*	*Feindselig*	Fehler
Gelb	2	1	1 Fehler
Rot	1	1	1 Fehler
		Gesamt:	2 Fehler
		Informationsgewinn:	2 Fehler – 2 Fehler = 0 Fehler

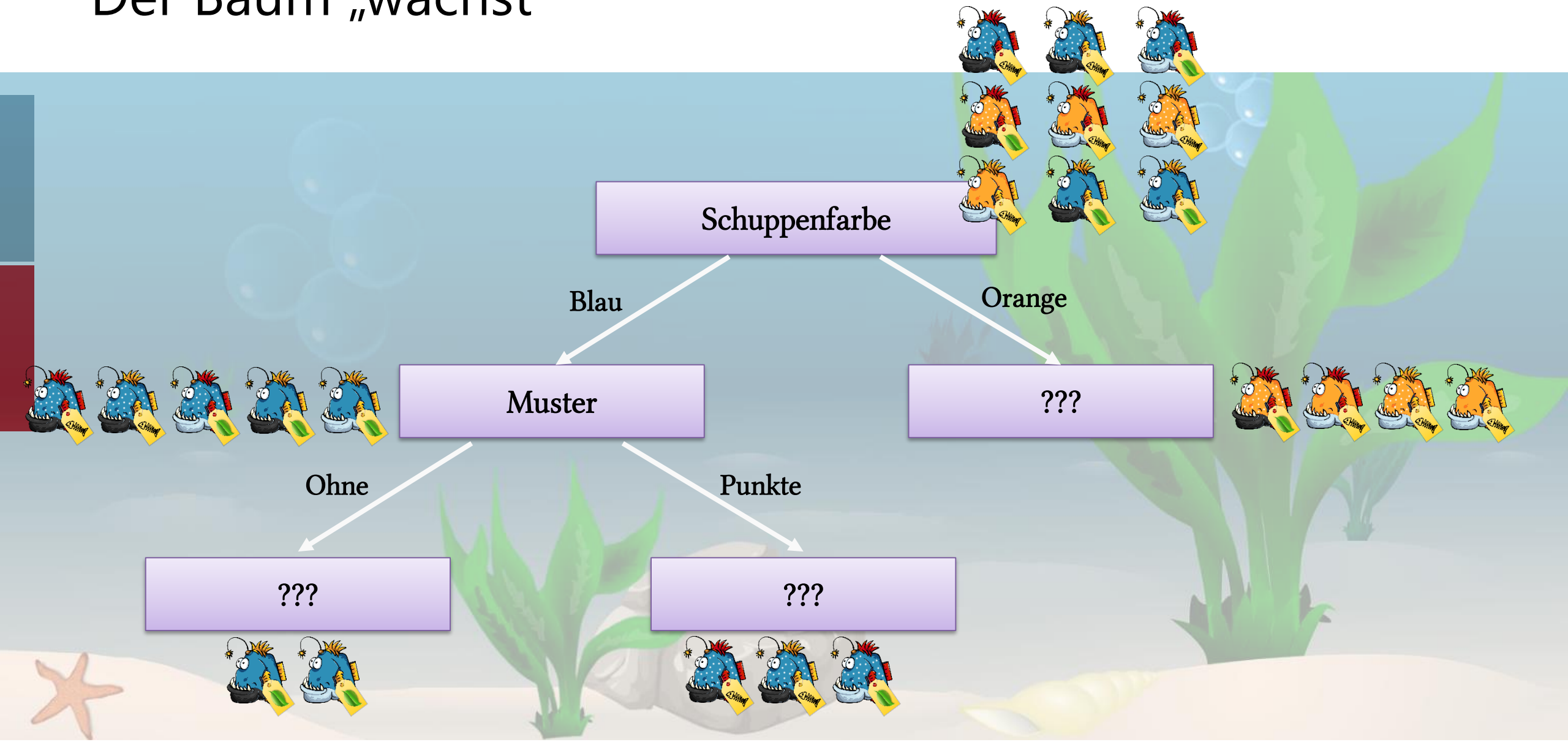
Suche nach dem nächsten „wichtigsten“ Attribut

Teilmenge der Trainingsdaten

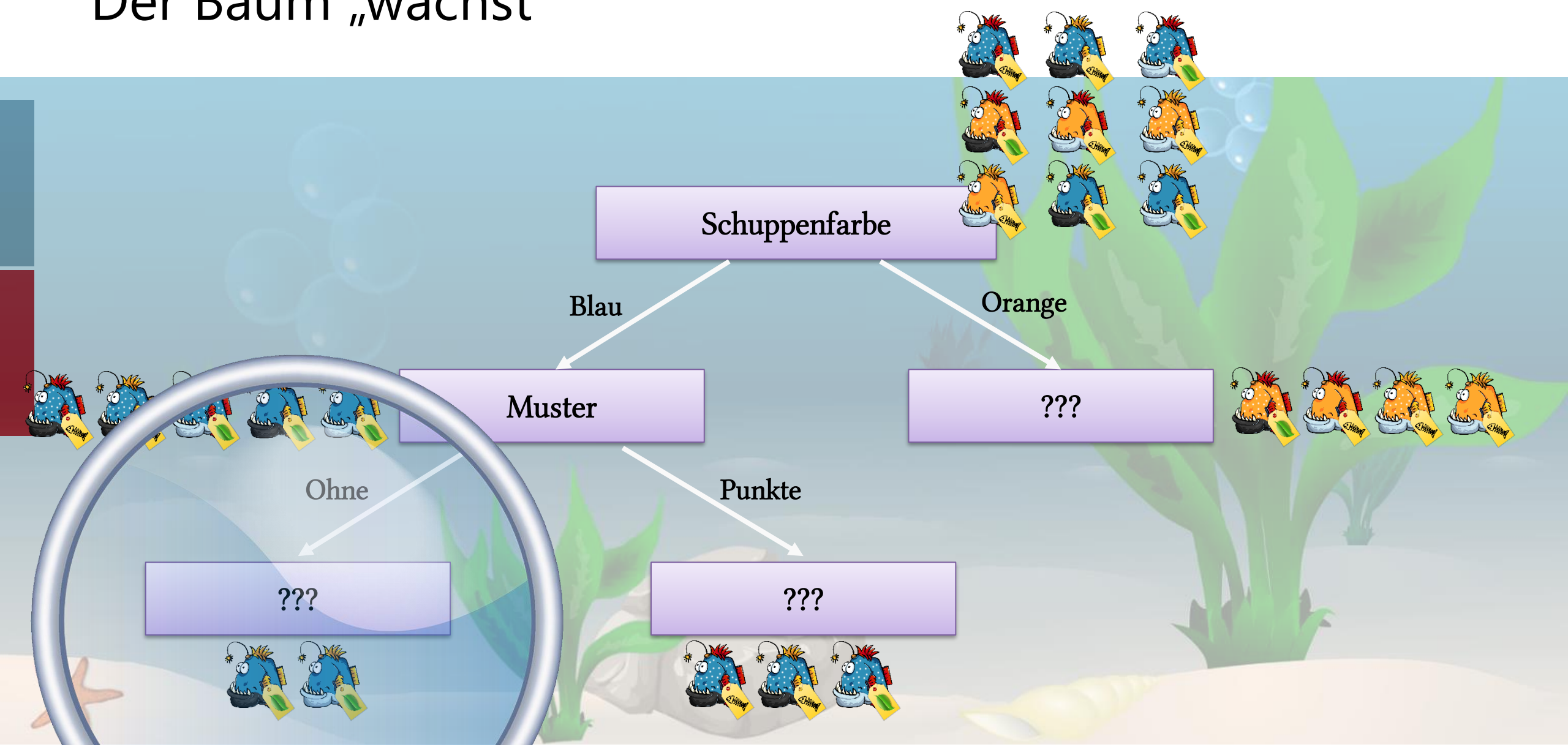
Bauchfarbe

Bauchfarbe			
Fehler vor Aufteilen der Datenmenge:			2 Fehler
Attributwert / Label	*Friedlich*	*Feindselig*	Fehler
Weiß	2	0	0 Fehler
Schwarz	1	2	1 Fehler
		Gesamt:	1 Fehler
		Informationsgewinn:	2 Fehler – 1 Fehler = 1 Fehler

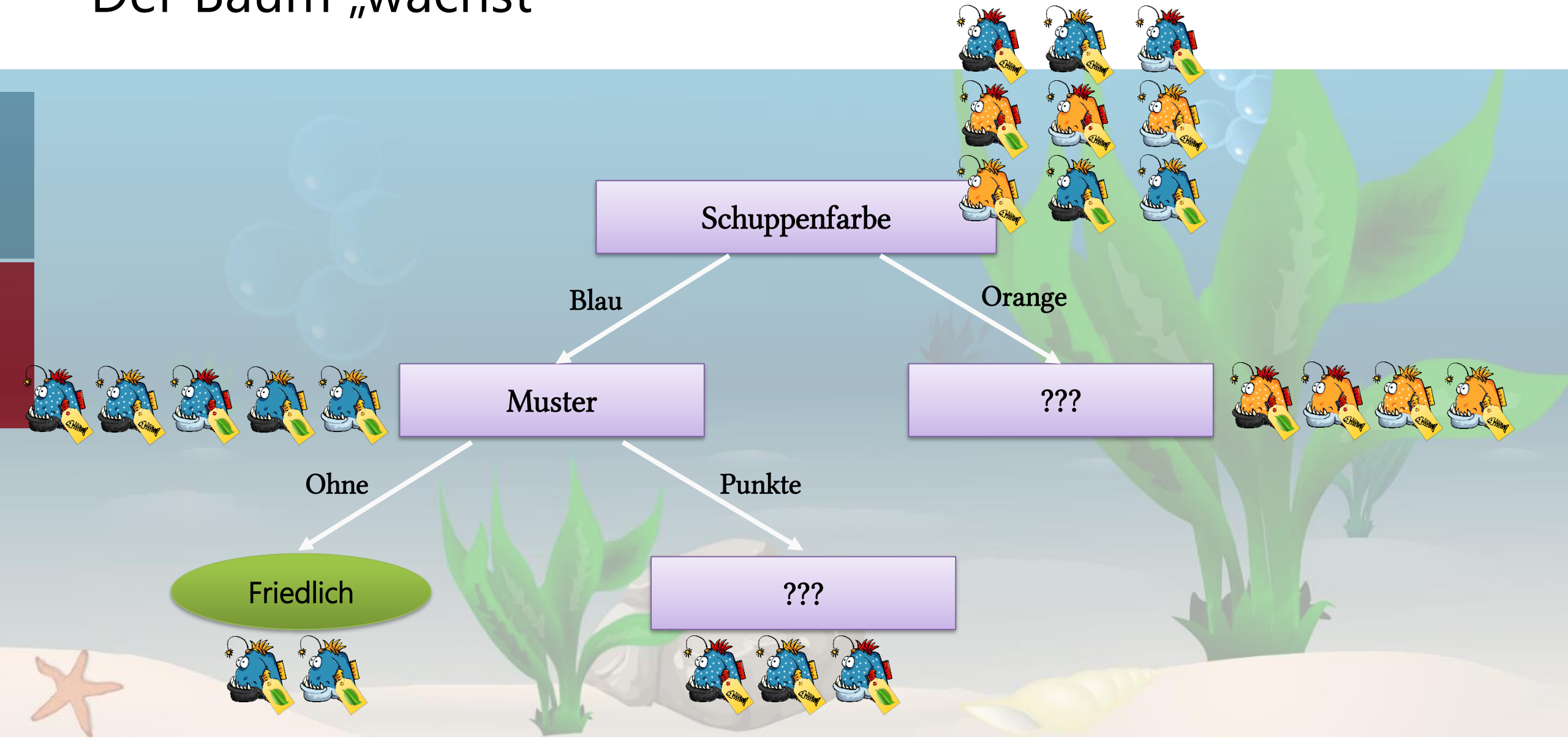
Der Baum „wächst“



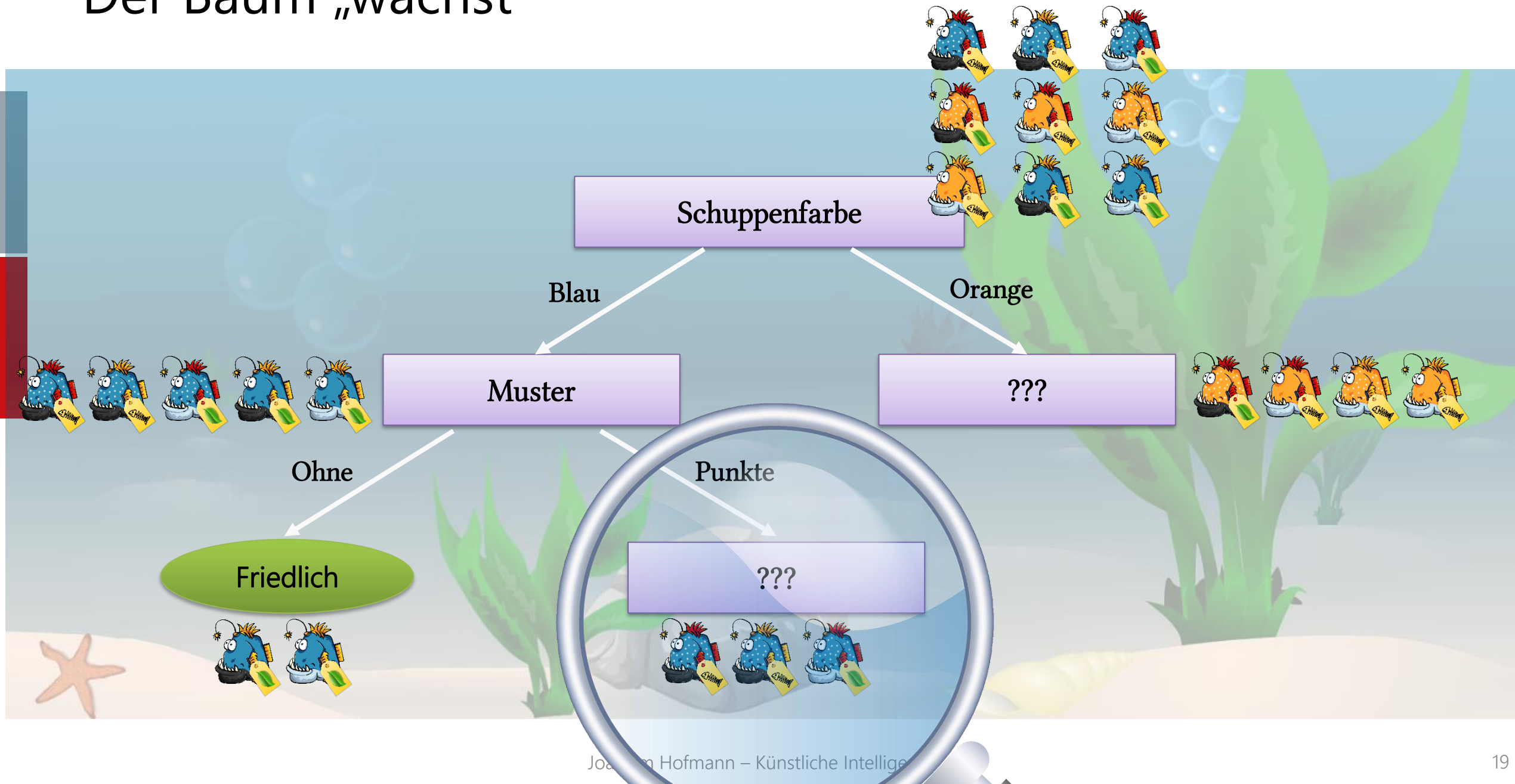
Der Baum „wächst“



Der Baum „wächst“



Der Baum „wächst“



Suche nach dem nächsten „wichtigsten“ Attribut

Teilmenge der Trainingsdaten

Bauchfarbe

Bauchfarbe			
Fehler vor Aufteilen der Datenmenge:			1 Fehler
Attributwert / Label	*Friedlich*	*Feindselig*	Fehler
Weiß	1	0	0 Fehler
Schwarz	0	2	0 Fehler
		Gesamt:	0 Fehler
		Informationsgewinn:	1 Fehler – 0 Fehler = 1 Fehler

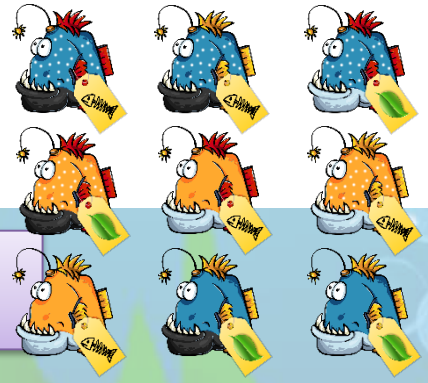
Suche nach dem nächsten „wichtigsten“ Attribut

Teilmenge der Trainingsdaten

Flossenfarbe

Flossenfarbe			
Fehler vor Aufteilen der Datenmenge:			1 Fehler
Attributwert / Label	*Friedlich*	*Feindselig*	Fehler
Gelb	0	1	0 Fehler
Rot	1	1	1Fehler
		Gesamt:	1 Fehler
		Informationsgewinn:	1 Fehler – 1 Fehler = 0 Fehler

Der Baum „wächst“



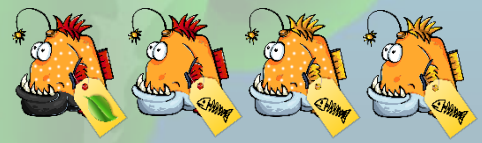
Schuppenfarbe

Blau

Orange

Muster

???



Ohne

Punkte

Friedlich

Bauchfarbe

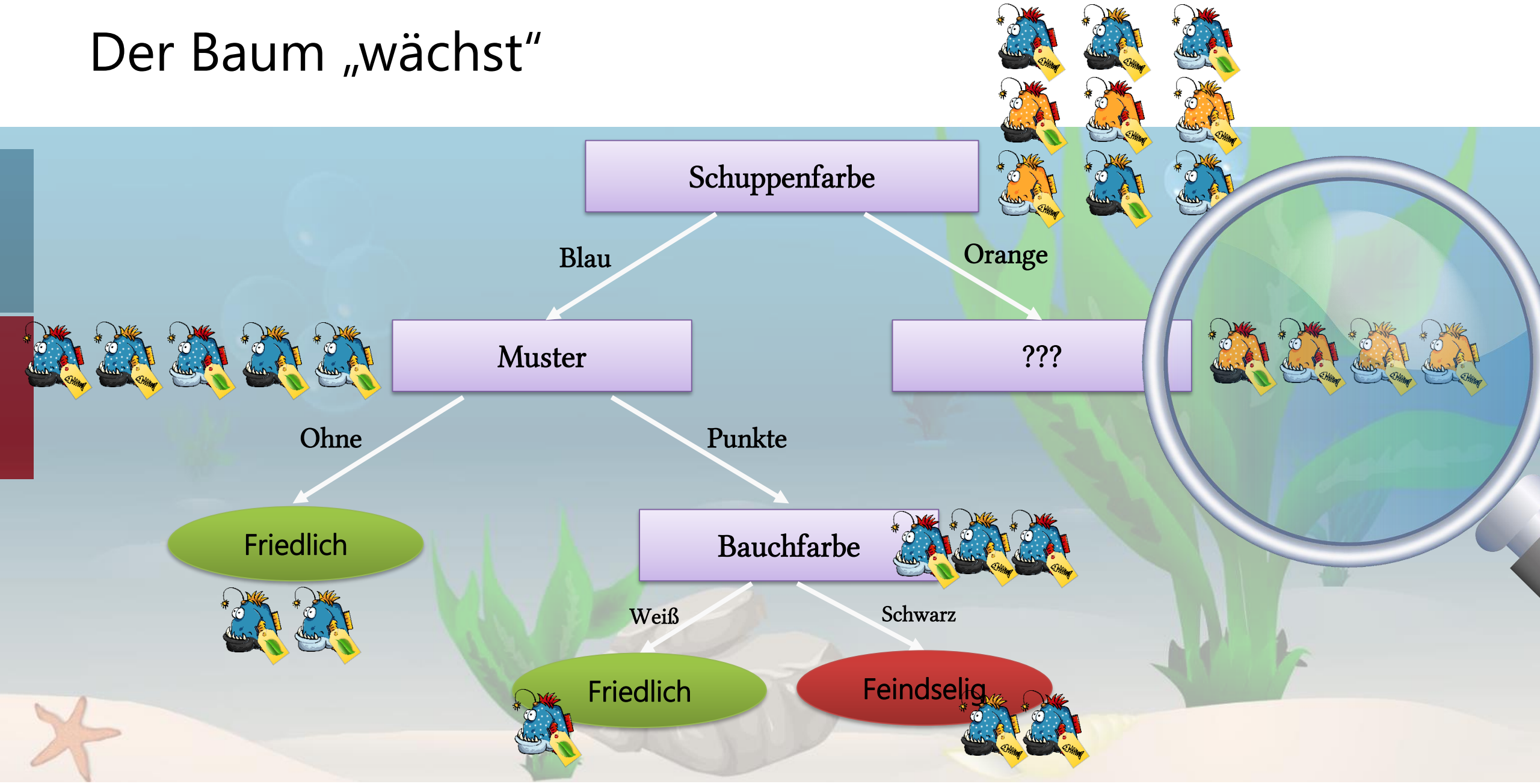
Weiß

Schwarz

Friedlich

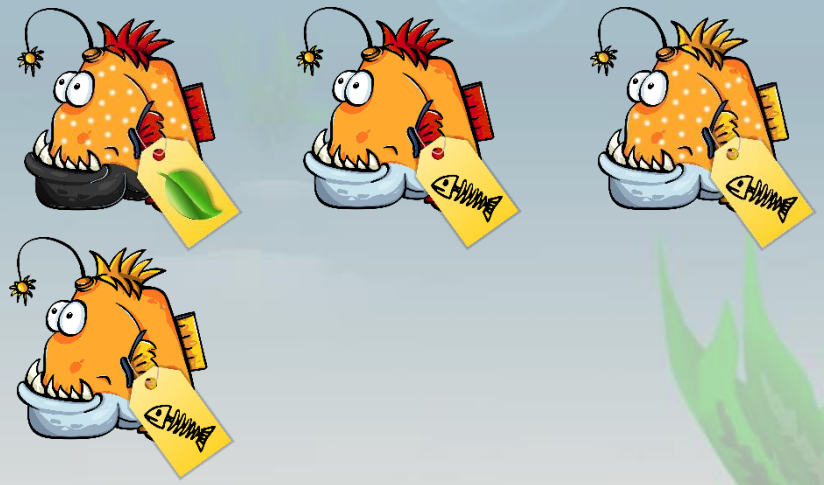
Feindselig

Der Baum „wächst“



Suche nach dem nächsten wichtigsten „Attribut“

Trainingsdaten

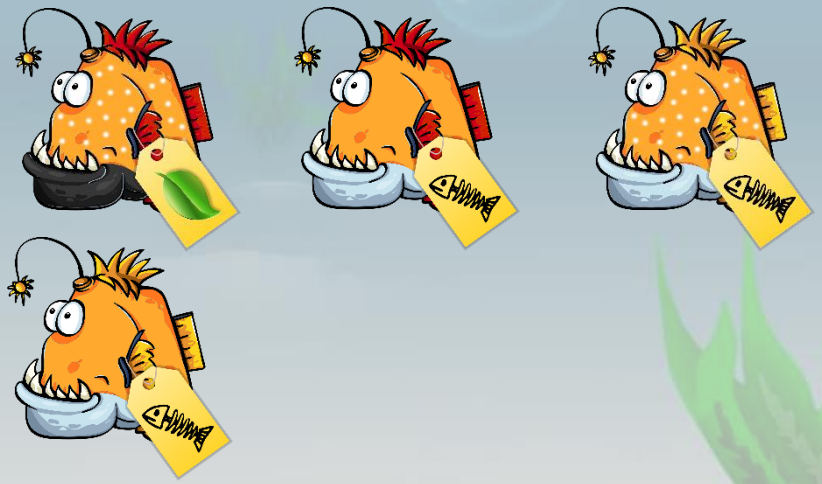


Muster

Muster			
Fehler vor Aufteilen der Datenmenge:			1 Fehler
Attributwert / Label	*Friedlich*	*Feindselig*	Fehler
Ohne	0	2	0 Fehler
Punkte	1	1	1 Fehler
		Gesamt:	1 Fehler
		Informationsgewinn:	1 Fehler – 1 Fehler = 0 Fehler

Suche nach dem nächsten wichtigsten „Attribut“

Trainingsdaten

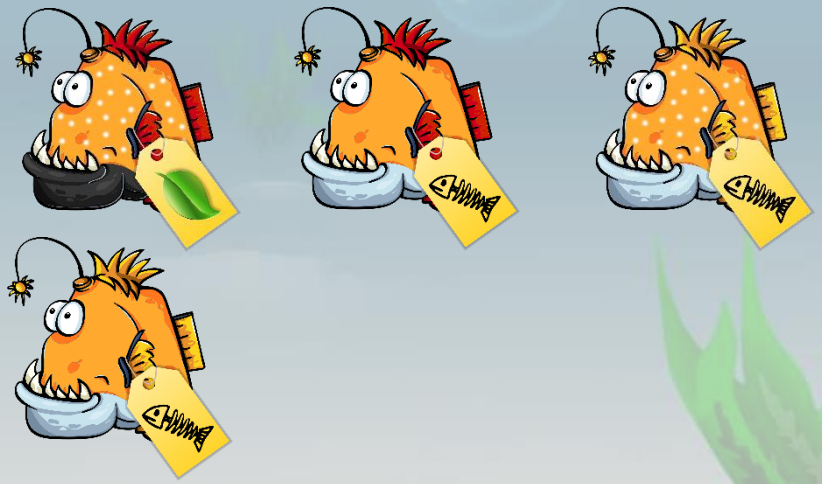


Bauchfarbe

Bauchfarbe			
Fehler vor Aufteilen der Datenmenge:			1 Fehler
Attributwert / Label	*Friedlich*	*Feindselig*	Fehler
Weiß	0	3	0 Fehler
Schwarz	1	0	0 Fehler
		Gesamt:	0 Fehler
		Informationsgewinn:	1 Fehler – 0 Fehler = 1 Fehler

Suche nach dem nächsten wichtigsten „Attribut“

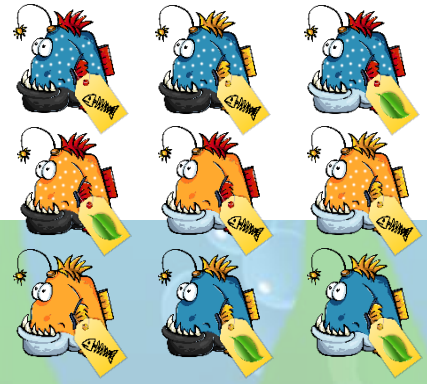
Trainingsdaten

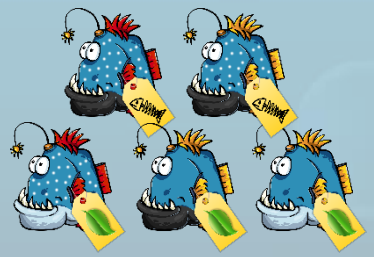


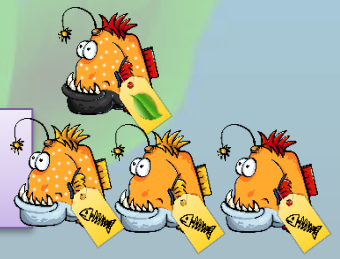
Flossenfarbe

Flossenfarbe			
Fehler vor Aufteilen der Datenmenge:			1 Fehler
Attributwert / Label	*Friedlich*	*Feindselig*	Fehler
Gelb	0	2	0 Fehler
Rot	1	1	1 Fehler
		Gesamt:	1 Fehler
		Informationsgewinn:	1 Fehler – 1 Fehler = 0 Fehler

Der fertige Entscheidungsbaum







Schuppenfarbe

Blau

Orange

Muster

Bauchfarbe

Ohne

Punkte

Weiß

Schwarz

Friedlich

Bauchfarbe

Feindselig

Friedlich

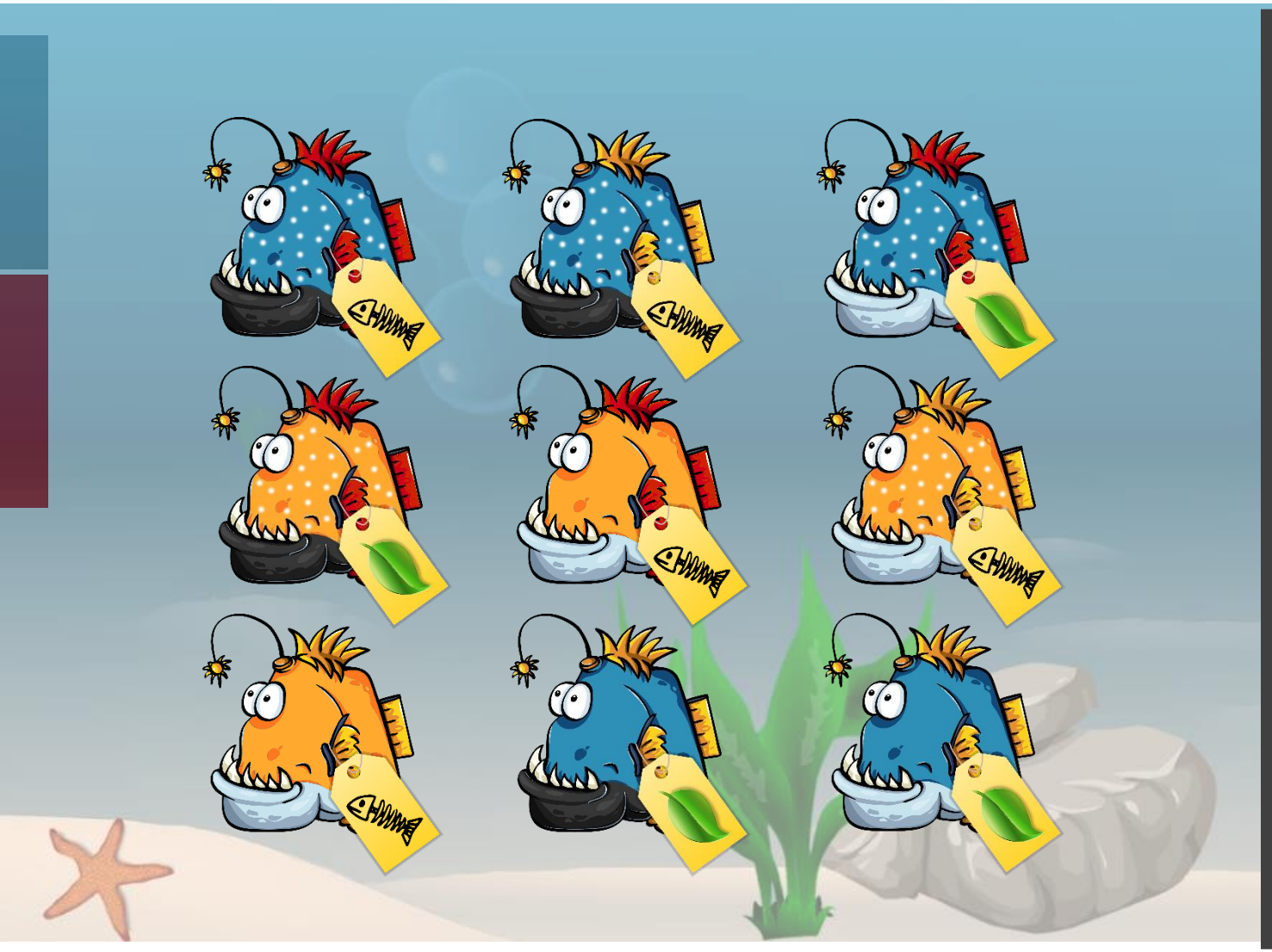
Weiß

Schwarz

Friedlich

Feindselig

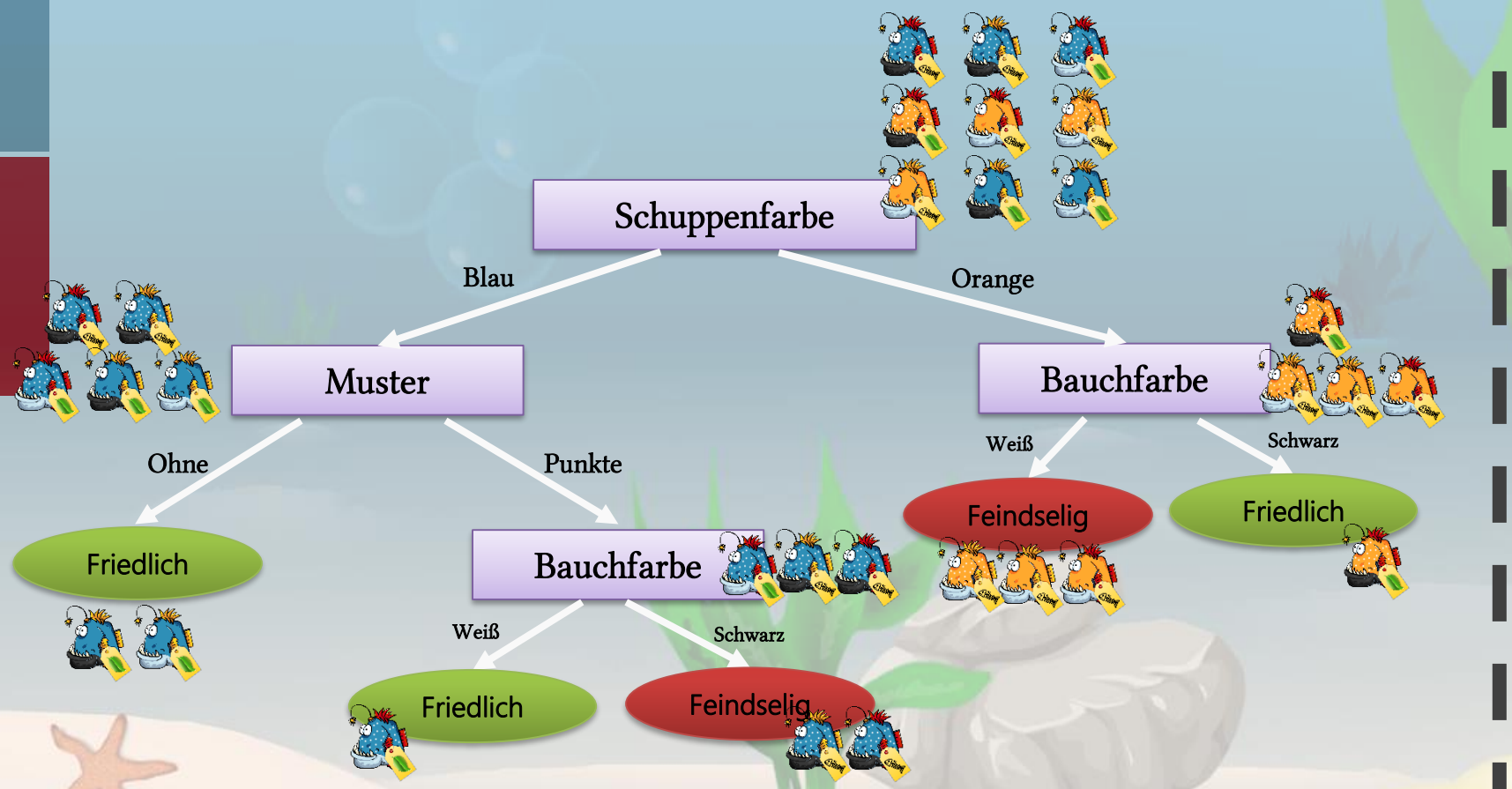
Zurück zur Anfangsfrage



Ist dieser Fisch bissig oder nicht?
Versuchen Sie dies aus den anderen Fischen abzuleiten?

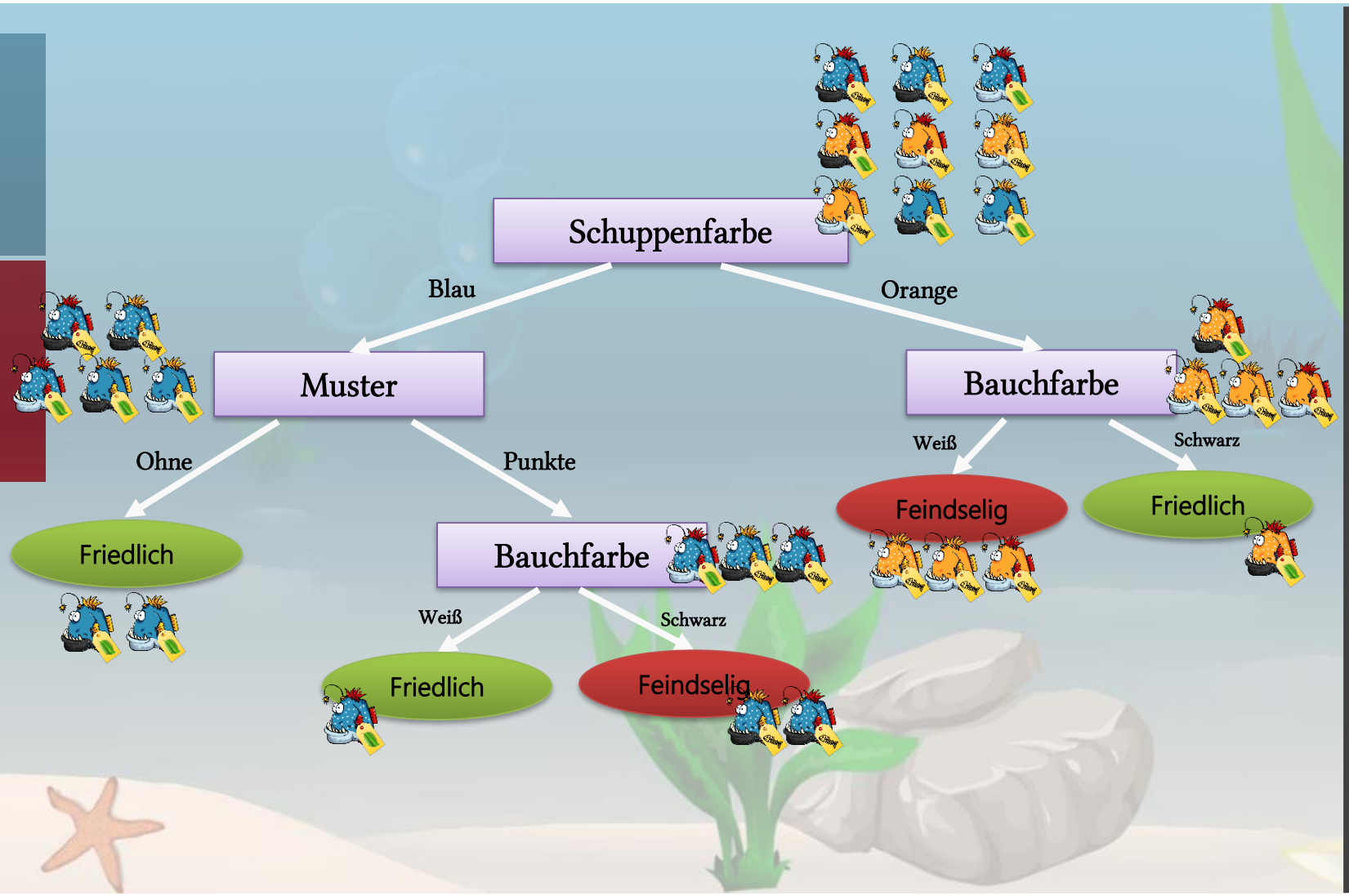
Zurück zur Anfangsfrage

- Bewertung des Baums anhand von Testdaten



Friedlich

Ein Schritt zurück: Wie ist dieser Baum entstanden?



Formulieren Sie einen Algorithmus zur Erstellung eines Entscheidungsbaums ausgehend von einer gegebenen Menge an gelabelten Trainingsdaten!

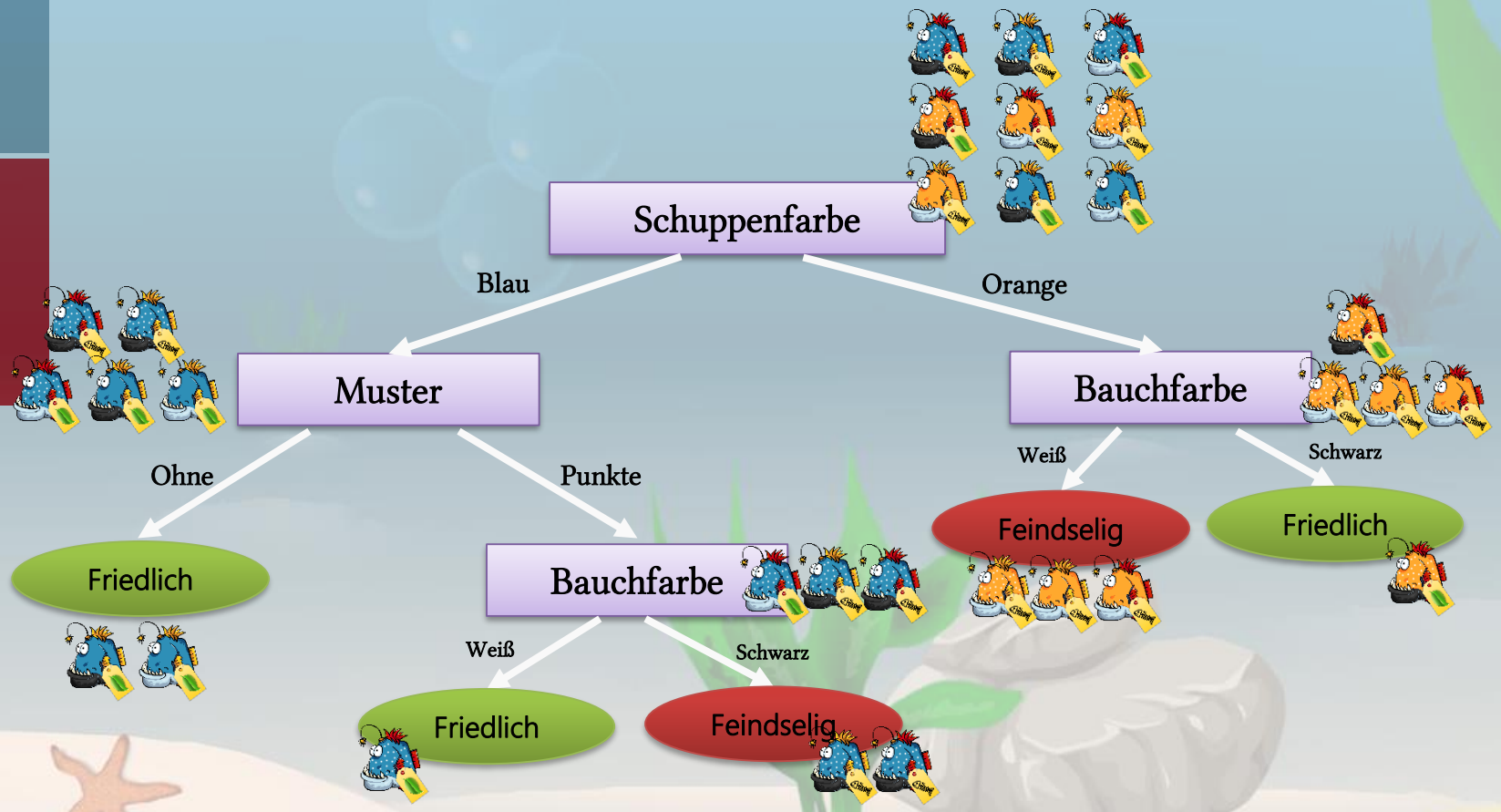
Algorithmus zur Erstellung eines Entscheidungsbaumes

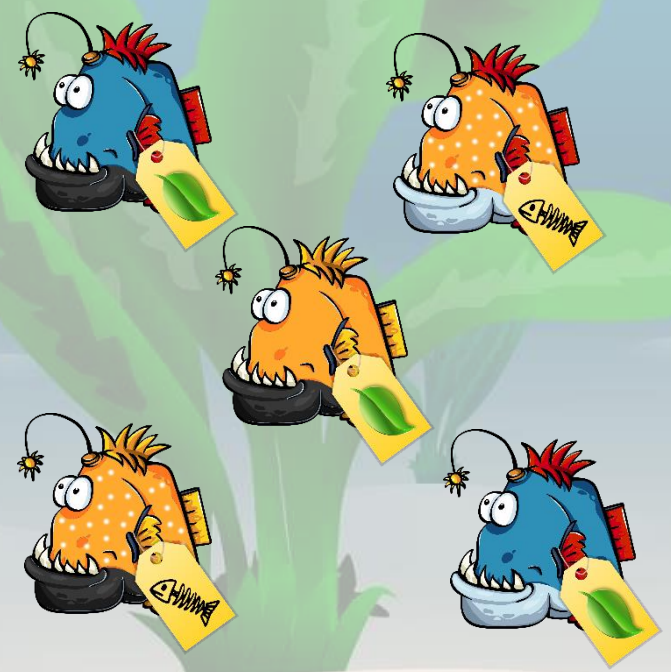
- Wiederhole bis alle Elemente in einem Knoten das gleiche Label haben oder kein Attribut mehr zum Aufsplitten vorhanden ist: der aktuelle Knoten wird zu einem Blattknoten mit dem Label, das mehrheitlich vorkommt

1. Bestimme das Attributs mit dem größten Informationsgewinn!
2. Teile die Trainingsdaten entsprechend der Attributwerte in Teilmengen auf!
3. Wende den Algorithmus beginnend mit Schritt 1 rekursiv auf die einzelnen Teilmengen an!

Wie gut ist der Baum: Testen

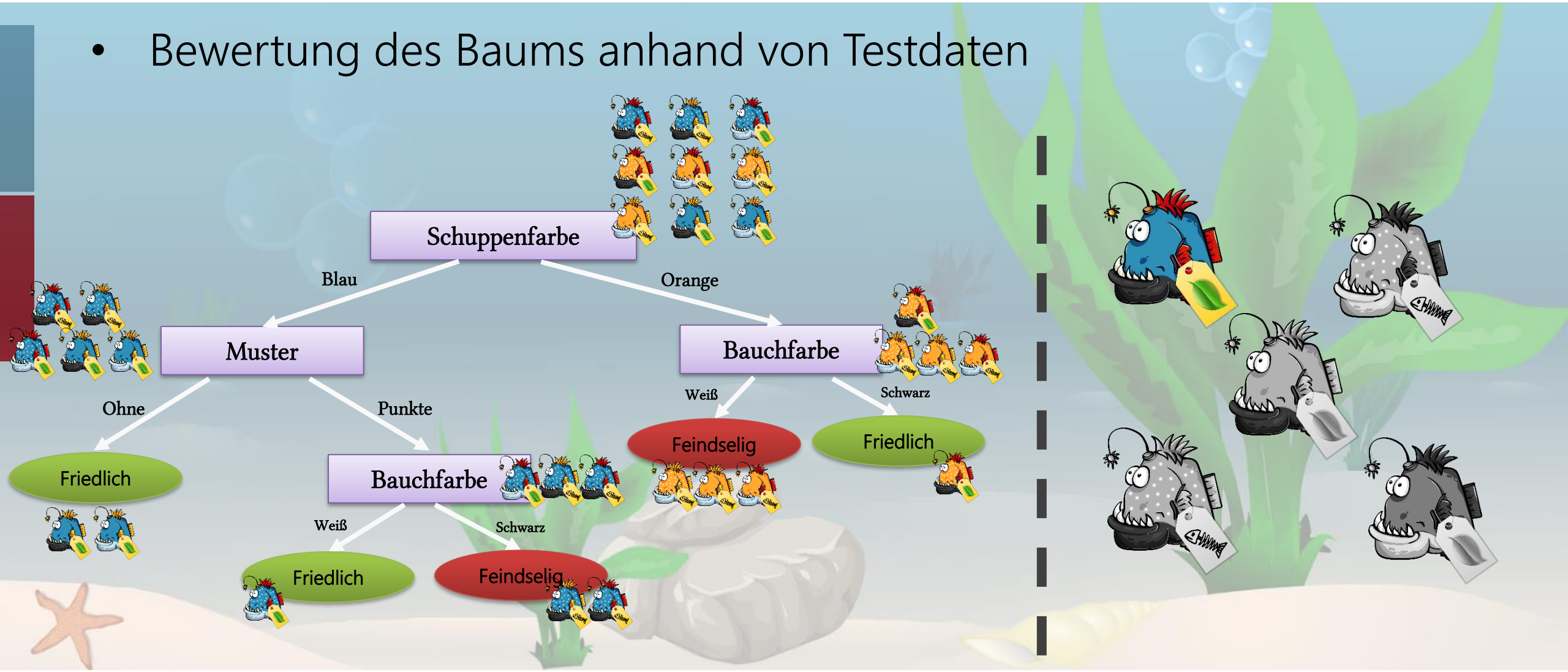
- Bewertung des Baums anhand von Testdaten





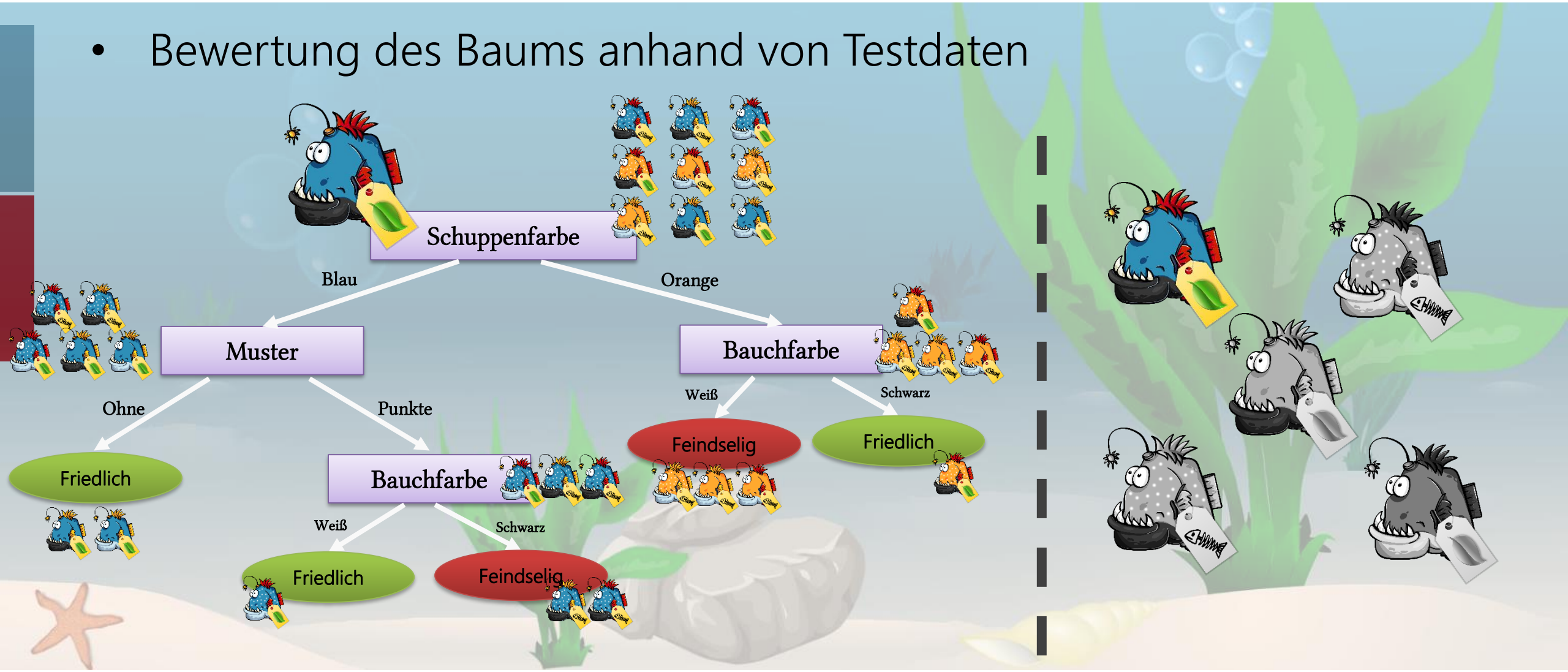
Wie gut ist der Baum: Testen

- Bewertung des Baums anhand von Testdaten



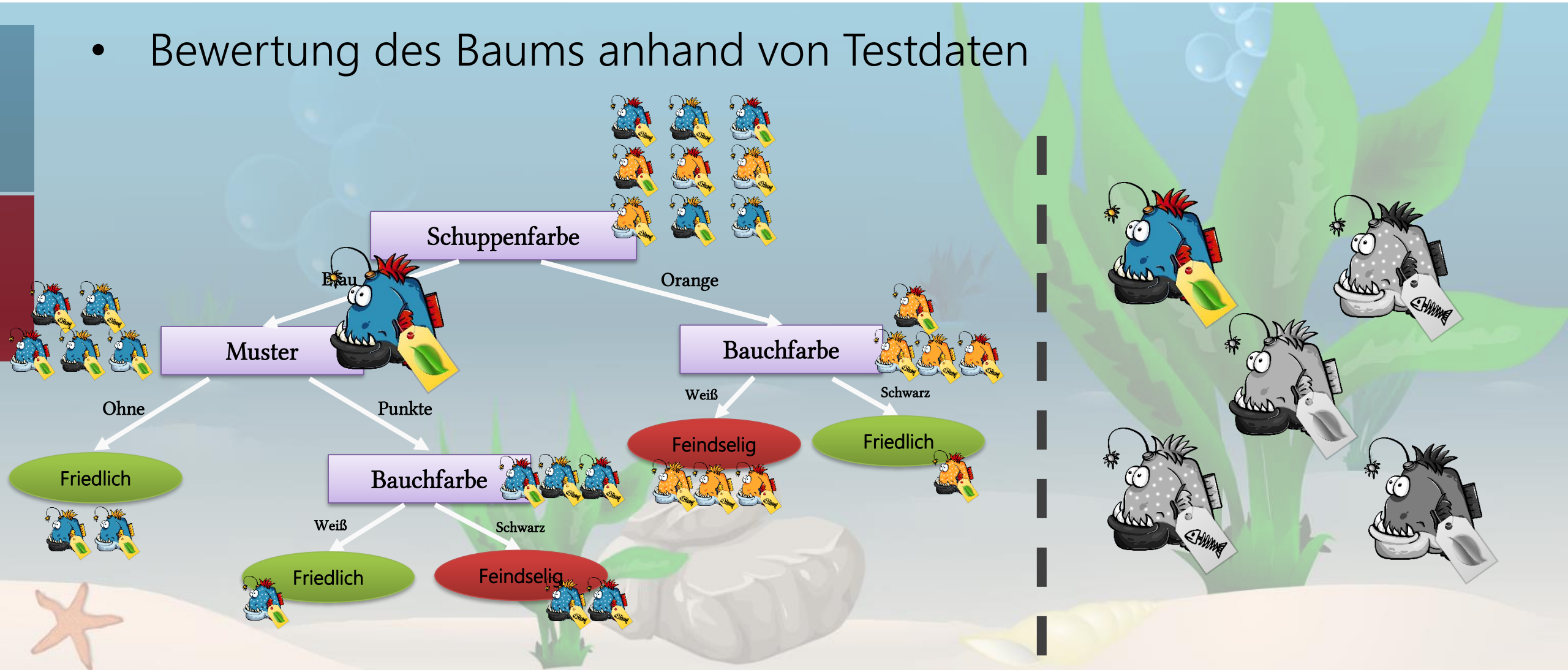
Wie gut ist der Baum: Testen

- Bewertung des Baums anhand von Testdaten



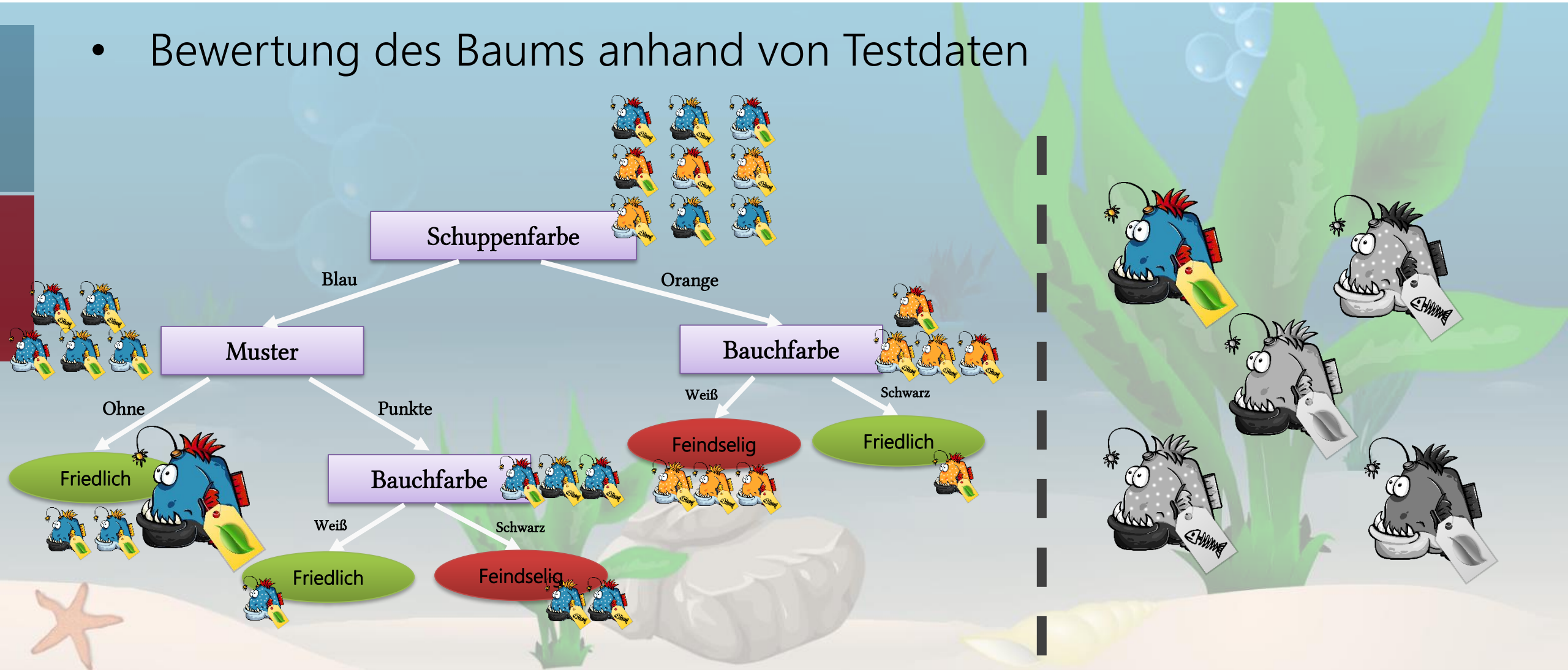
Wie gut ist der Baum: Testen

- Bewertung des Baums anhand von Testdaten



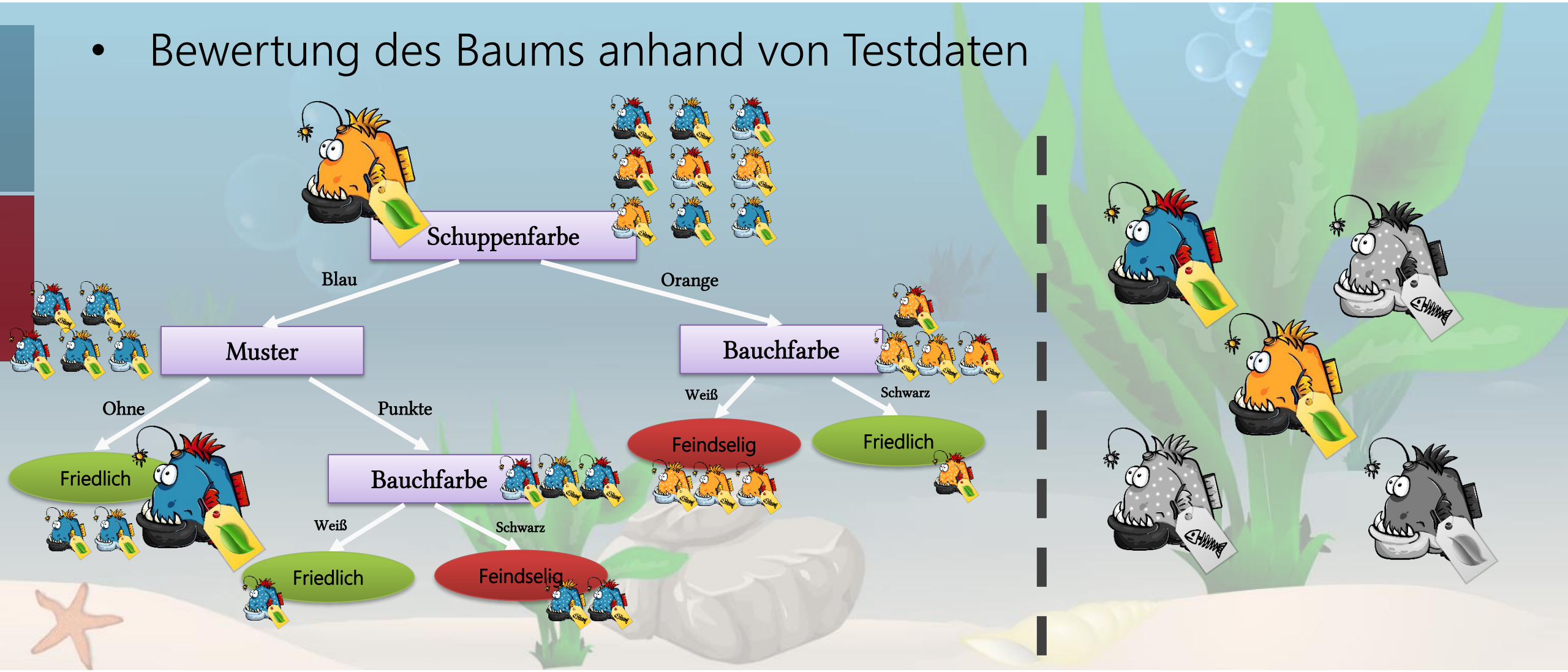
Wie gut ist der Baum: Testen

- Bewertung des Baums anhand von Testdaten



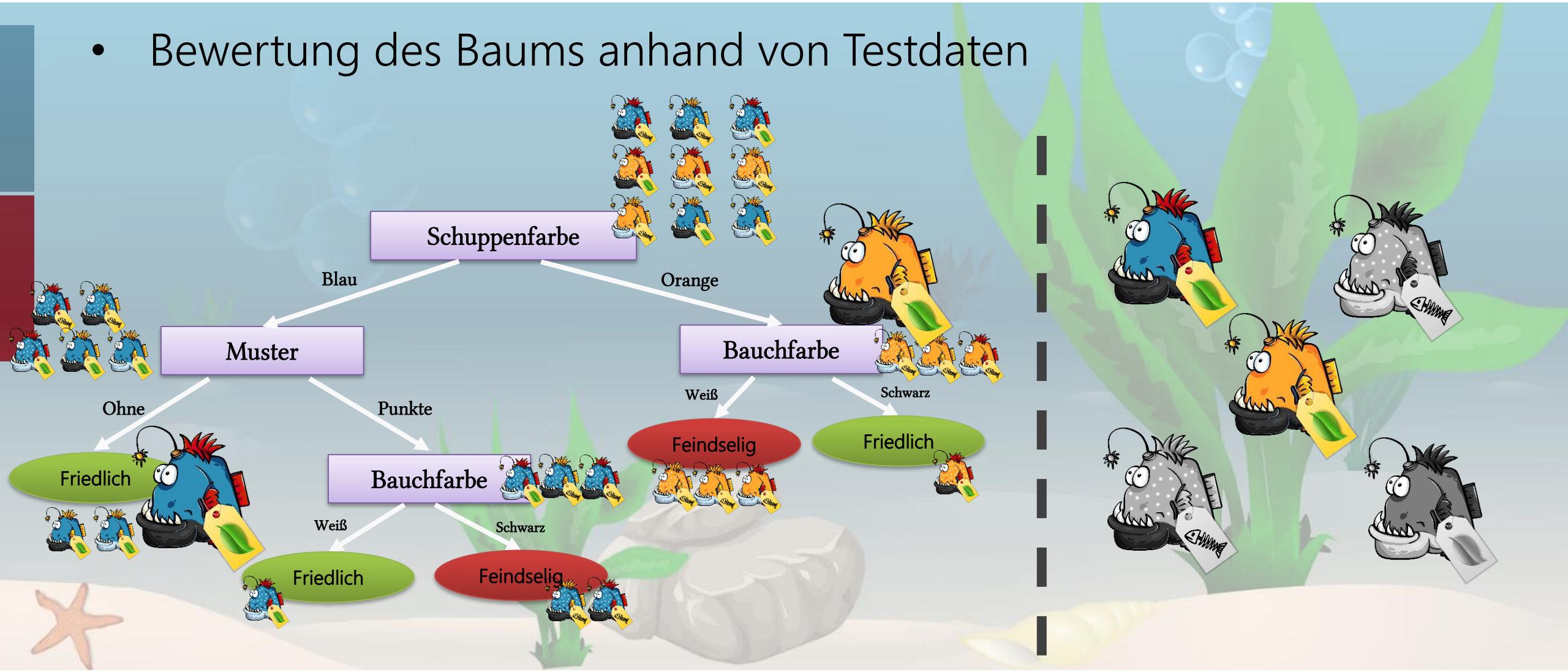
Wie gut ist der Baum: Testen

- Bewertung des Baums anhand von Testdaten



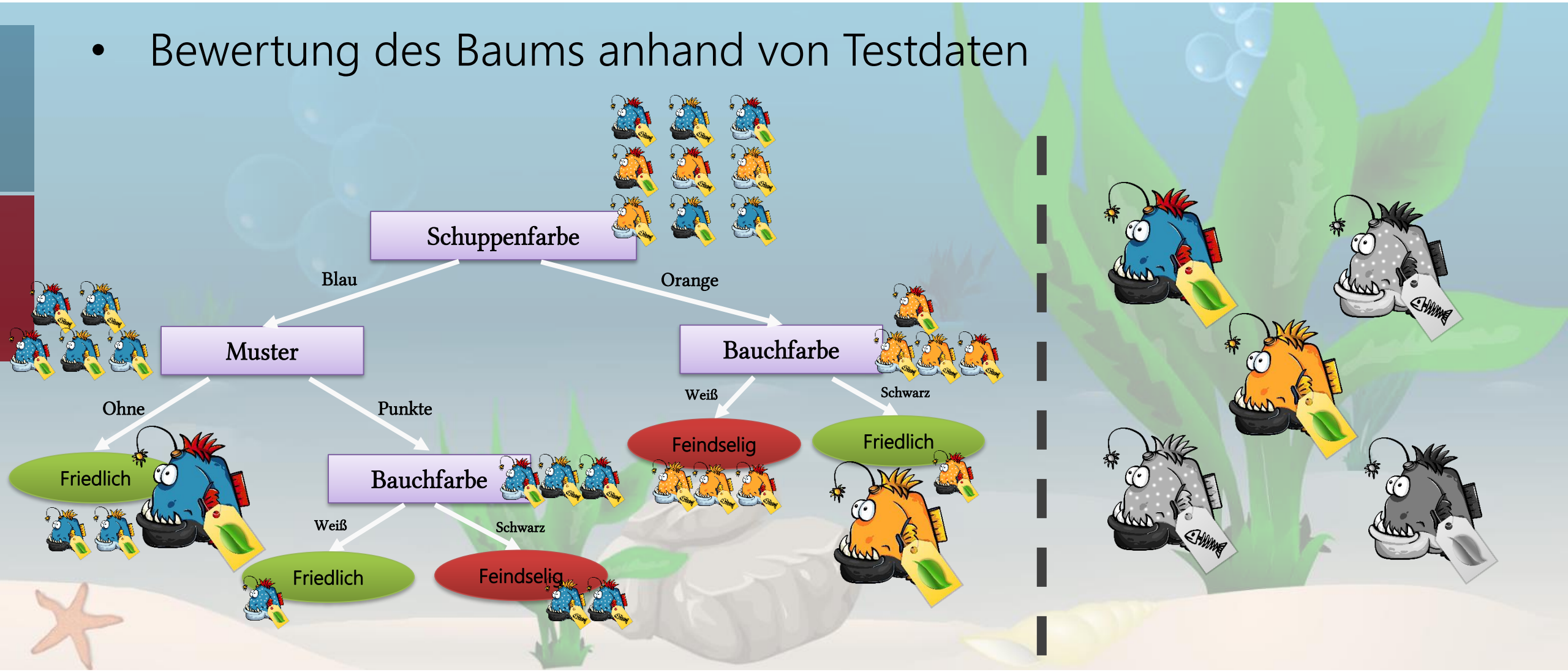
Wie gut ist der Baum: Testen

- Bewertung des Baums anhand von Testdaten



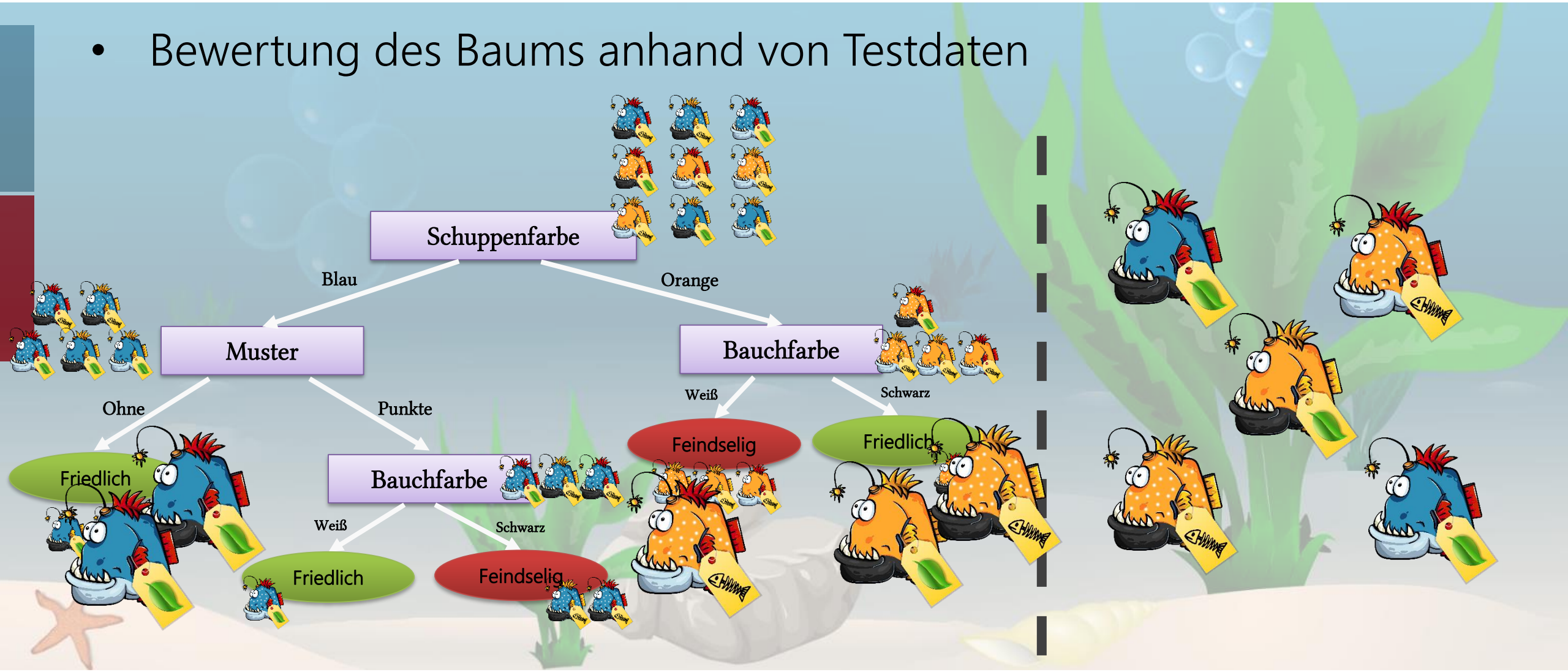
Wie gut ist der Baum: Testen

- Bewertung des Baums anhand von Testdaten



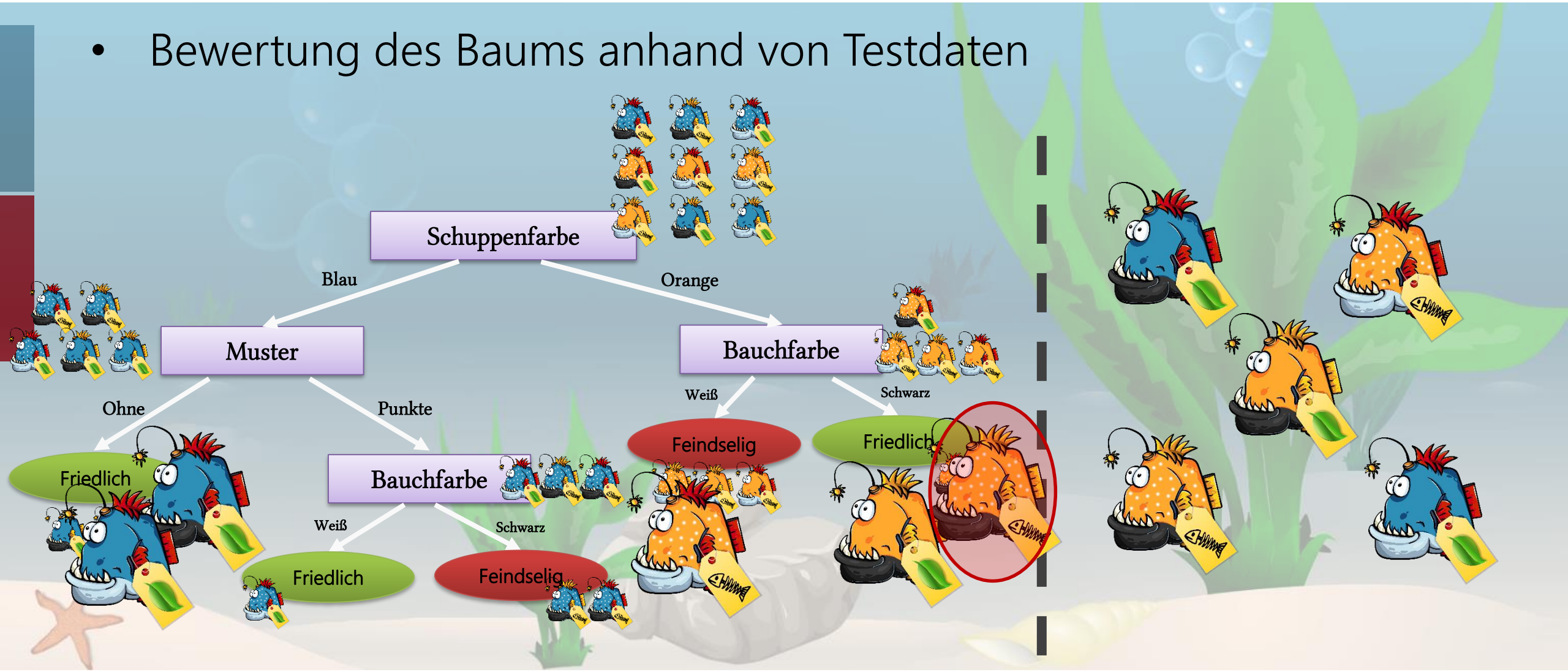
Wie gut ist der Baum: Testen

- Bewertung des Baums anhand von Testdaten

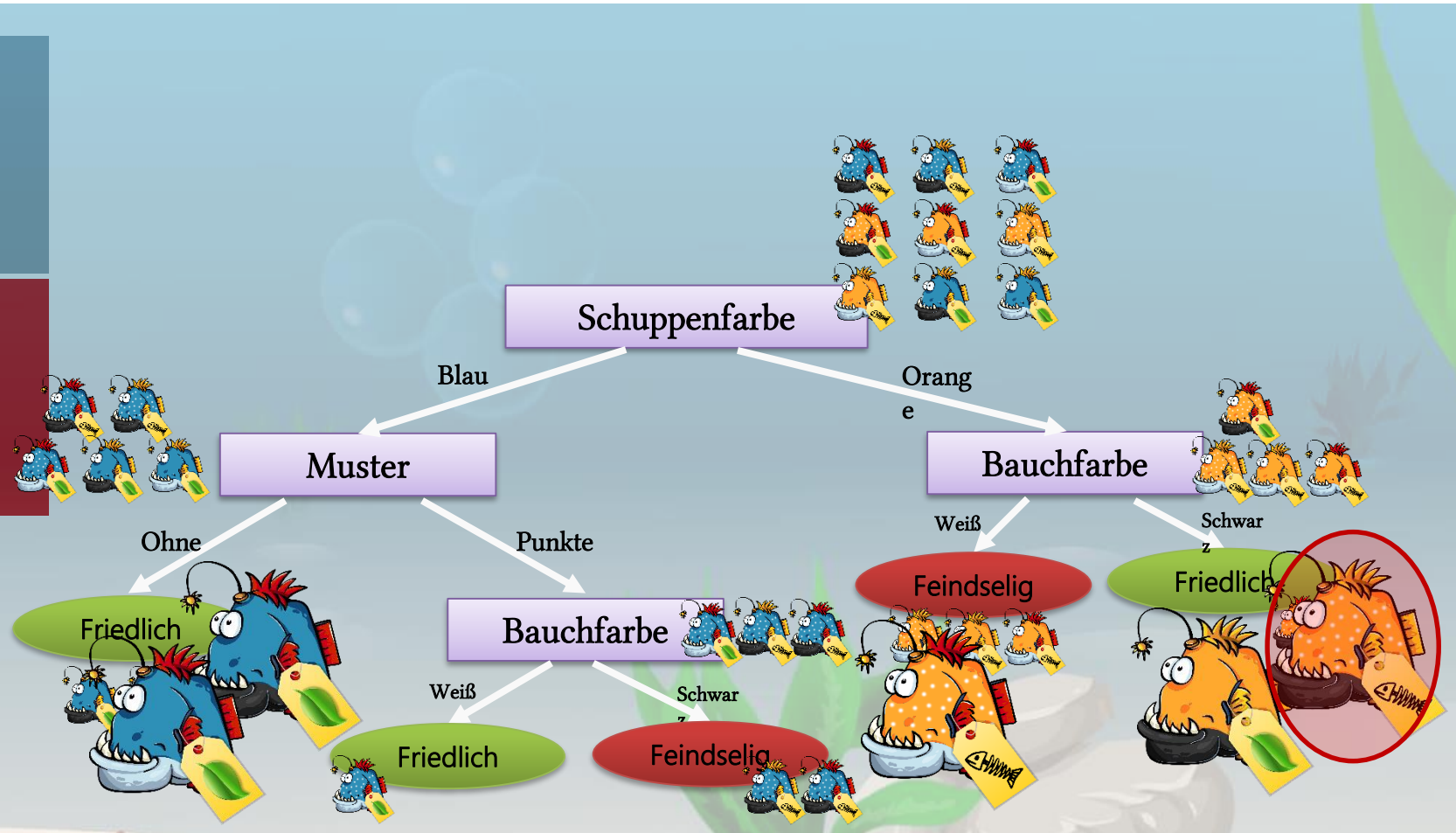


Wie gut ist der Baum: Testen

- Bewertung des Baums anhand von Testdaten



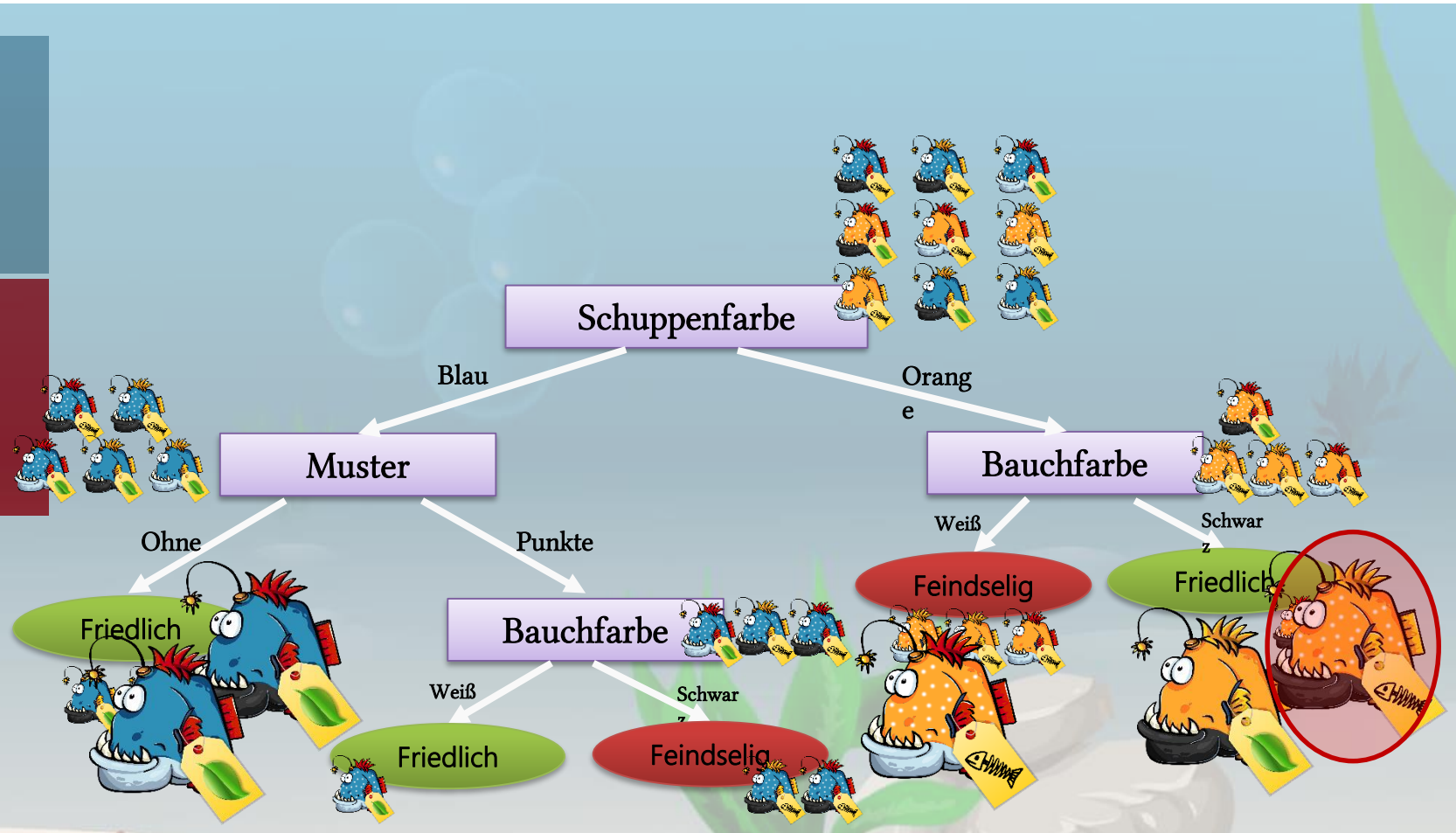
Bewertung der gelabelten Testdaten



Konfusionsmatrix

Erwartetes Label		Berechnetes Label		
		Friedlich	Feindselig	Summe
	Friedlich			
	Feindselig			
	Summe			

Bewertung der gelabelten Testdaten

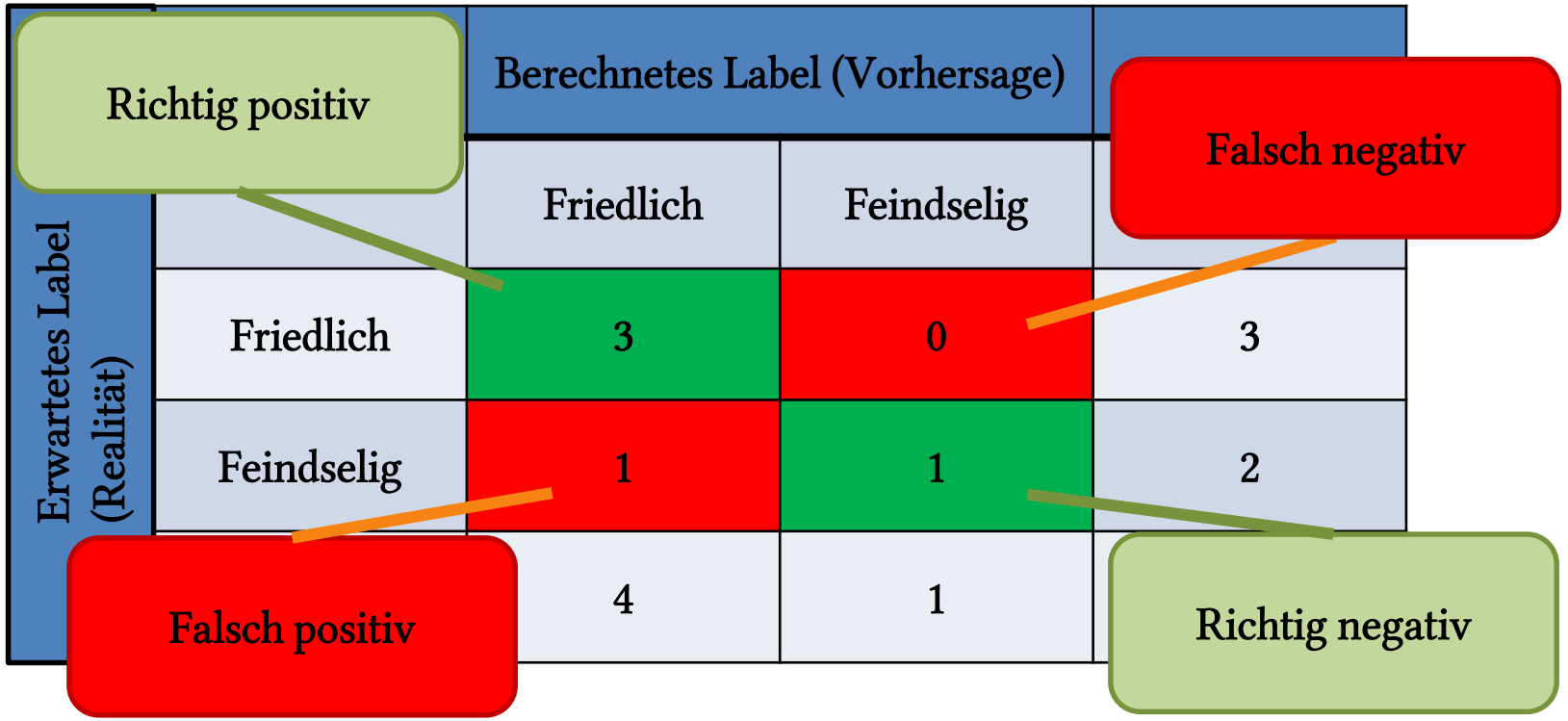


Konfusionsmatrix

Erwartetes Label		Berechnetes Label		
		Friedlich	Feindselig	Summe
	Friedlich	3	0	3
	Feindselig	1	1	2
	Summe	4	1	5

Bewertung der gelabelten Testdaten

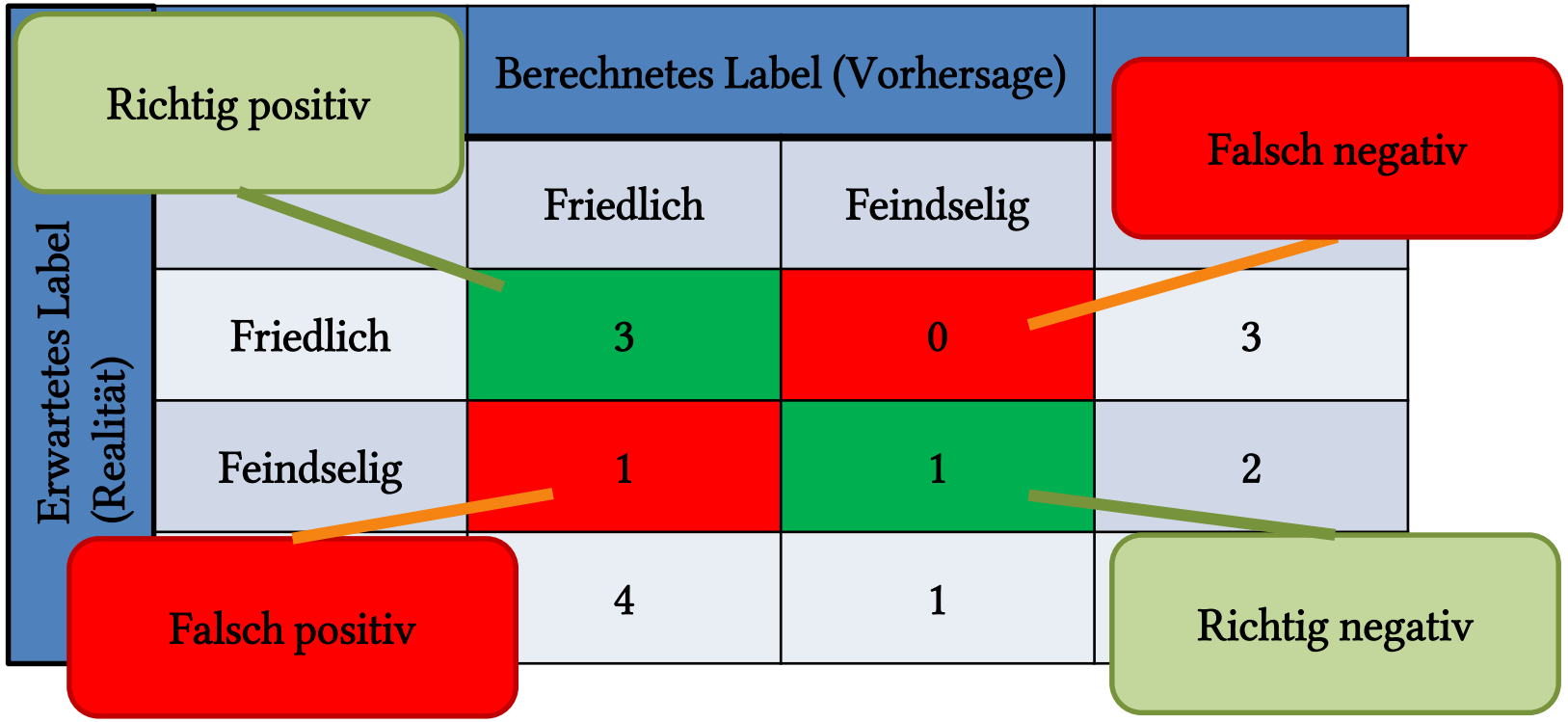
Konfusionsmatrix



Erwartetes Label (Realität)	Berechnetes Label (Vorhersage) Friedlich	Feindselig	
Friedlich	3	0	3
Feindselig	1	1	2
	4	1	

Bewertung der gelabelten Testdaten

Konfusionsmatrix



Erwartetes Label (Realität)	Berechnetes Label (Vorhersage): Friedlich	Berechnetes Label (Vorhersage): Feindselig	
Friedlich	3	0	3
Feindselig	1	1	2
	4	1	

Genauigkeit: $\frac{1+3}{1+1+3+0} = \frac{4}{5} = 80\%$