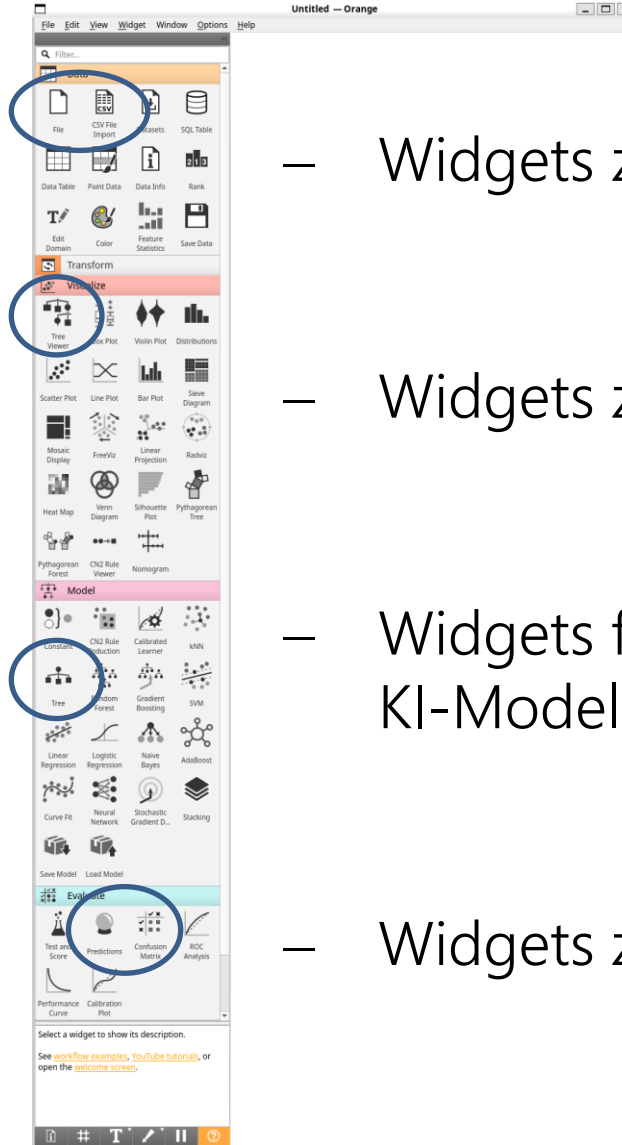




Künstliche Intelligenz – Data Mining mit Orange



Orange – Überblick über Widgets (gemeinsam)



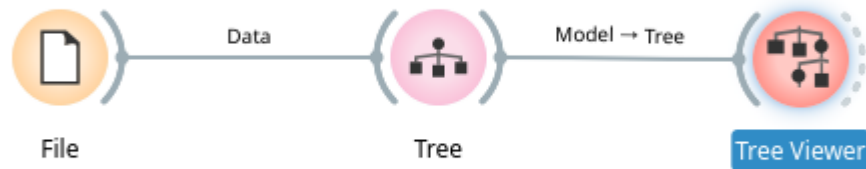
– Widgets zu Daten

– Widgets zum Anzeigen

– Widgets für verschiedene KI-Modelle

– Widgets zur Auswertung

Trainingsdaten/Entscheidungsbaum in Orange



File - Orange

Source

File: 2+3 Entscheidungsbaum\Fische_Beispi1_Trainingsdaten.csv

File Type

Automatically detect type

Info

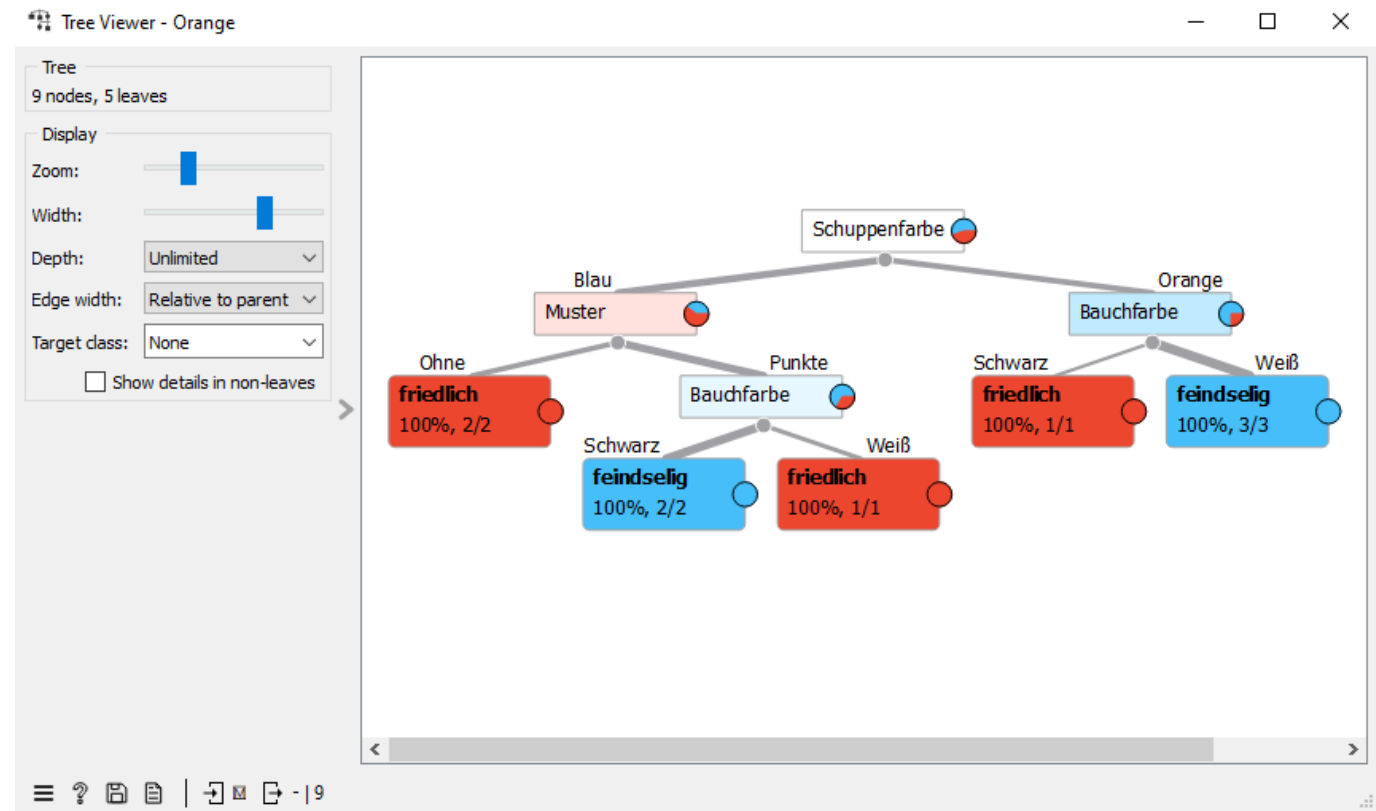
9 instances
5 features (no missing values)
Data has no target variable.
0 meta attributes

Columns (Double click to edit)

Name	Type	Role	Values
1 Muster	category	feature	Ohne, Punkte
2 Bauchfarbe	category	feature	Schwarz, Weiß
3 Schuppenfarbe	category	feature	Blau, Orange
4 Flossenfarbe	category	feature	Gelb, Rot
5 Label	category	target	feindselig, friedlich

Reset Apply

Browse documentation datasets



- feature $\triangleq$ Merkmale der Fische
- target $\triangleq$ Label/Beschriftung auf welches trainiert werden soll

Hyperparameter im Tree/Entscheidungsbaum

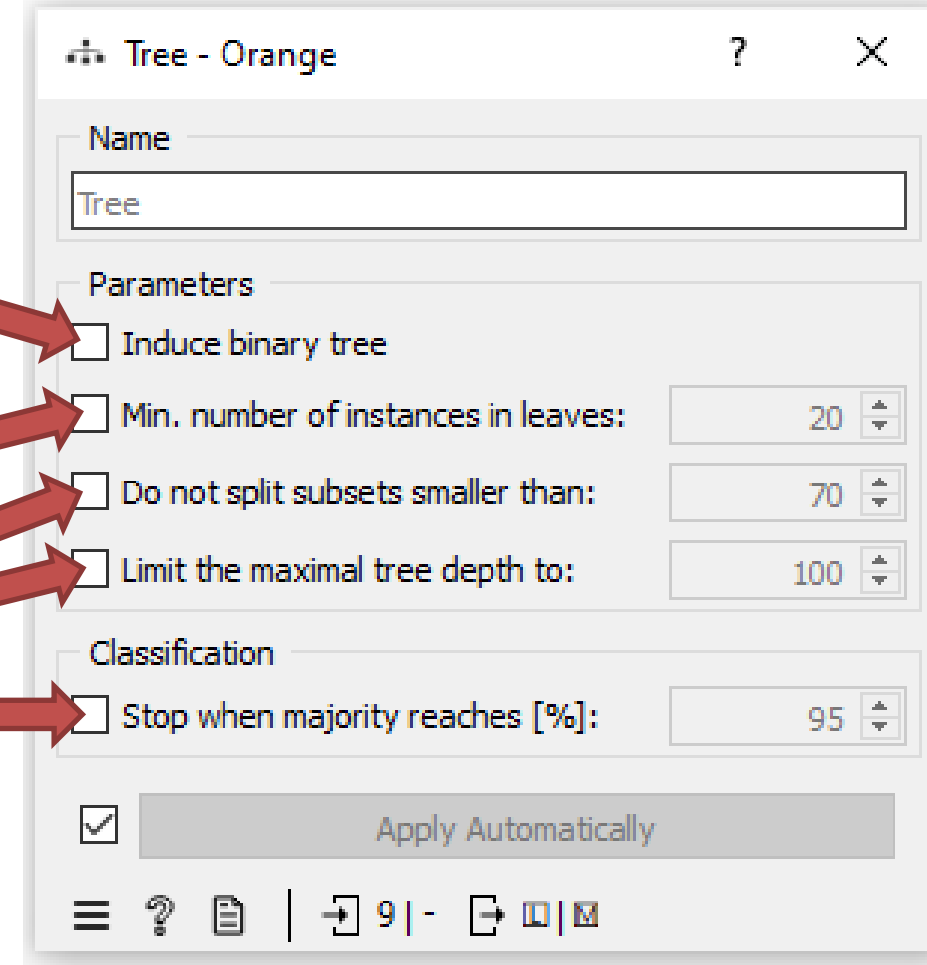
- **Hyperparameter:**

- Binärbaum mittels Zusammenfassung von Attributwerten zu einem Übergang

- **Pruning:**

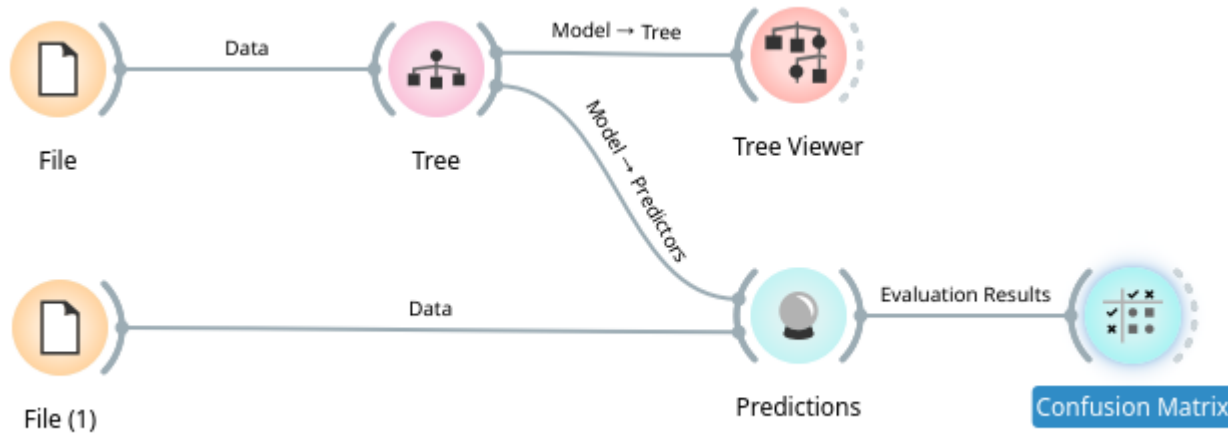
- Minimale Anzahl an Elementen im Blatt
 - Minimale Anzahl an Elementen im Knoten
 - Maximale Tiefe des Baums

- Erreichte Genauigkeit (optional)



- Passe deinen Entscheidungsbaum an, sodass dieser voll klassifiziert!

Testdaten/Konfusionsmatrix



Source

☒ File: 2+3 Entscheidungsbaum\Fische_Beispiel1_Testdaten.csv

☐ URL:

File Type

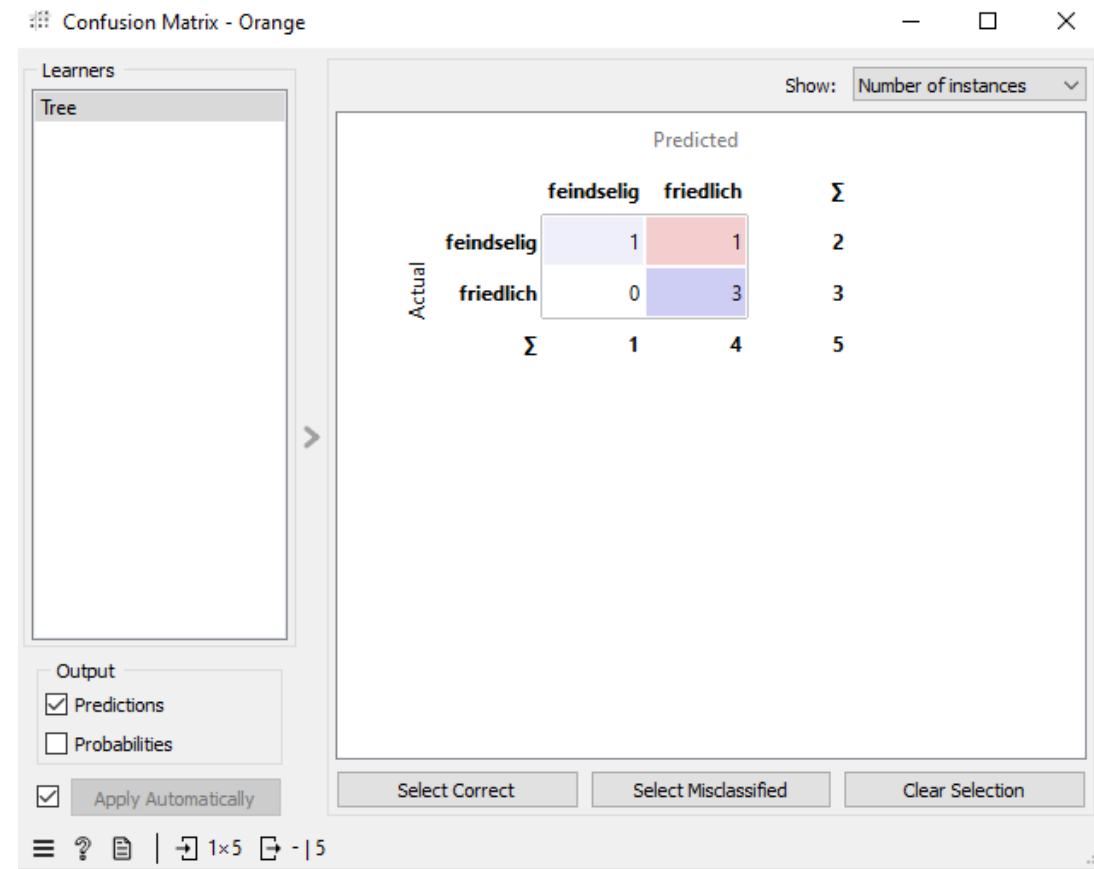
Automatically detect type

Info

5 instances
5 features (no missing values)
Data has no target variable.
0 meta attributes

Columns (Double click to edit)

	Name	Type	Role	Values
1	Muster	ⓐ categorical	feature	Ohne, Punkte
2	Bauchfarbe	ⓐ categorical	feature	Schwarz, Weiß
3	Schuppenfarbe	ⓐ categorical	feature	Blau, Orange
4	Flossenfarbe	ⓐ categorical	feature	Gelb, Rot
5	Label	ⓐ categorical	target	feindselig, friedlich



- feature $\triangleq$ Merkmale der Fische
- target $\triangleq$ Label/Beschriftung auf welches trainiert werden soll

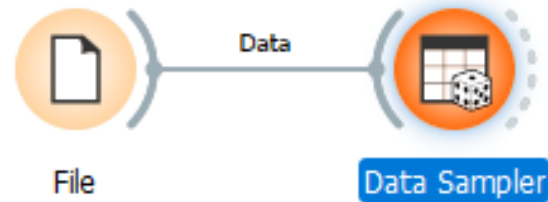
Aufgaben zur Übung

- Erstelle jeweils analog zum Beispiel der Präsentation in Orange ein neues OrangeWorkSheet für einen Entscheidungsbaum!
- Stelle die Hyperparameter für den Entscheidungsbaum passend ein und vergleiche jeweils die Entscheidungsbäume (TreeView-Widget) in Orange mit dem Baum, den Du per Hand in der vorherigen Aufgabe per Hand erstellt hast.
- Bewerte anschließend mithilfe der Konfusionsmatrix die Güte Deines (Orange-)Baums und vergleiche diese jeweils mit den Ergebnissen der vorherigen Aufgaben.
- Nutze dazu als Daten
 - a) [Hitzefrei Trainingsdaten.csv](#) und [Hitzefrei Testdaten.csv](#)
 - b) [Fische_Beispiel2 Trainingsdaten.csv](#) und [Fische_Beispiel2 Testdaten.csv](#)
 - c) Welches Problem ergibt sich bei der Klassifizierung von Fische2 ?

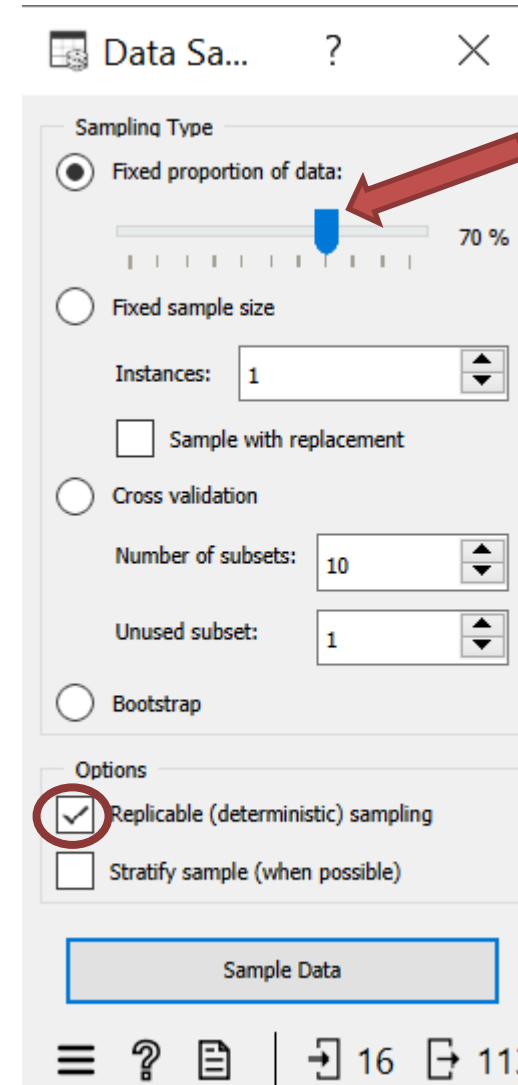
Hyperparameter

- Entscheidungsbaumalgorithmus erkennt Muster in Daten und erstellt daraus ein Modell.
 - Besonders geeignet für kategorische Daten
 - Keine Garantie für optimalen Baum (kein „bester“ oder „guter“ Baum sicher)
 - Sinnvoll nur bei vorhandenen Mustern in den Daten
- **Overfitting**: Zu stark angepasst, schlecht bei neuen/leicht geänderten Daten
 - Lösung: Baum stutzen (Pruning) → bessere Generalisierung
- **Underfitting**: Muster werden nicht oder kaum erkannt
 - Lösung: komplexeres Modell oder Feature Engineering (Transformation von Rohdaten in passende Attribute; statt volles Datum -> z. B. 4 Jahreszeiten)
- Testphase ist entscheidend zur Qualitätsbewertung des Modells
- **Aufgabe**: Ändere die Hyperparameter des Baums für Fische2?

Data Sampler – Datenaufteilung in Trainings-/Testdaten

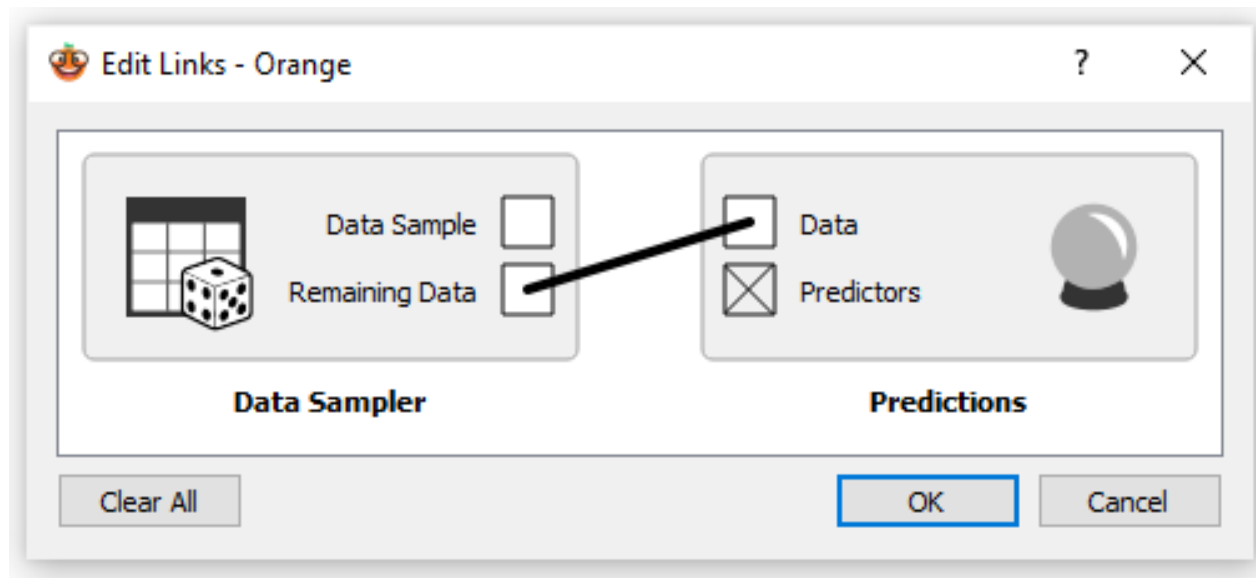
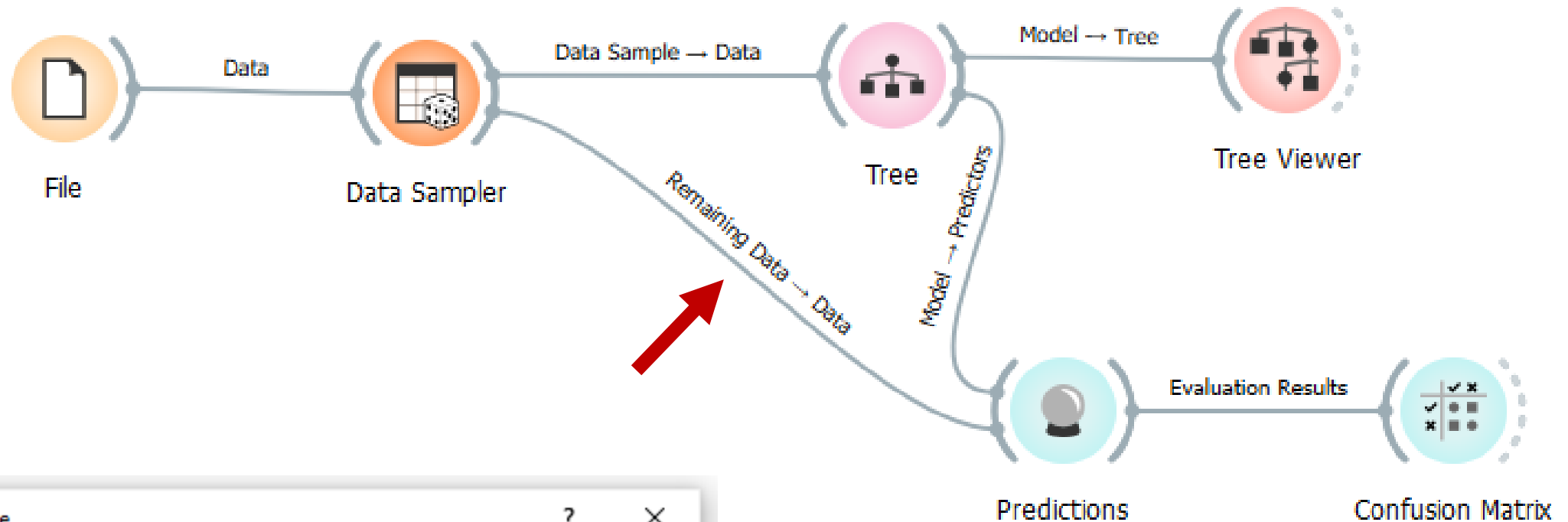


Reproduzierbares
Ergebnis, d.h. keine
zufällige Aufteilung
der Daten

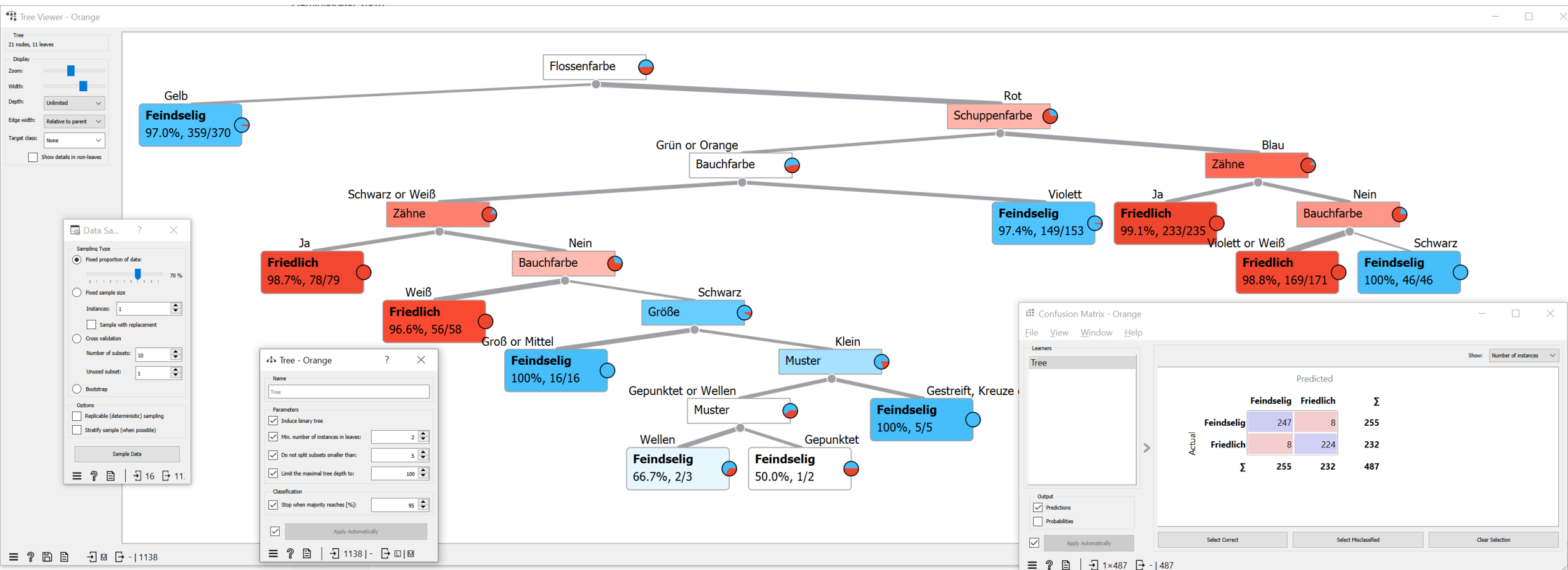


Aufteilung der
Daten in:
70% Trainingsdaten
30% Testdaten

Datenflussanpassung von Trainings-/Testdaten



(Eindeutiges?) Ergebnis zum großen Fischdatensatz



Aufgabe

- Erstelle mit Orange einen Entscheidungsbaum mittels geeigneter Einstellung der Hyperparameter zum Beißverhalten bei Affen.
- Bewerte anschließend mit Hilfe der Konfusionsmatrix die Güte Deines Baums.
- Die gelabelten Daten findest Du in der Datei **Affen-gesamt.csv**

Aufgabe (für Schnelle):

- Der Film Titanic (1997) zeigt den Untergang des Schiffs Titanic. Er zeigt das Schicksal der beiden Hauptdarsteller Jack und Rose (gespielt von Leonardo DiCaprio und Kate Winslet). Am Ende überlebt sie das Unglück, aber er nicht. Hätte ein KI-Modell das vorhersagen können?
- Die gelabelten Daten findest Du in der Datei **Titanic_Passagiere.xlsx**
- Erstelle mit Orange einen Entscheidungsbaum zur Überlebenschance auf der Titanic. Überlege dir hierbei, welche Daten überhaupt einen Einfluss auf die Überlebenschance haben können. Die irrelevanten Merkmale können im Widget „File“ als meta (sprich Metadaten, die keinen Einfluss haben) deklariert werden.
- **Stimmen KI-Ergebnis und Filmende überein?**