Auxiliary AI GmbH

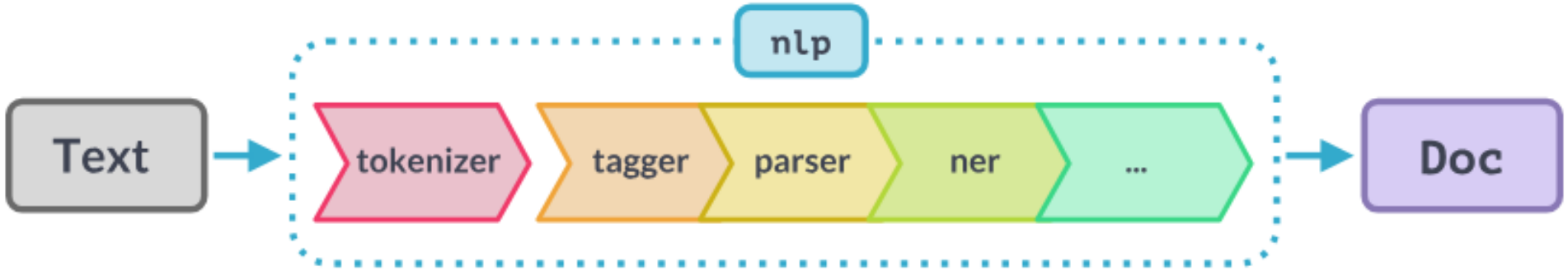
preprocessing pipeline & classification

- Block 03 -

Auxiliary AI GmbH

Preprocessing Pipeleine

- Wir werden hierfür spaCy nutzen, wobei auch andere Word Embeddings möglich sind.



Picture from ojosujono96 (Freepik)

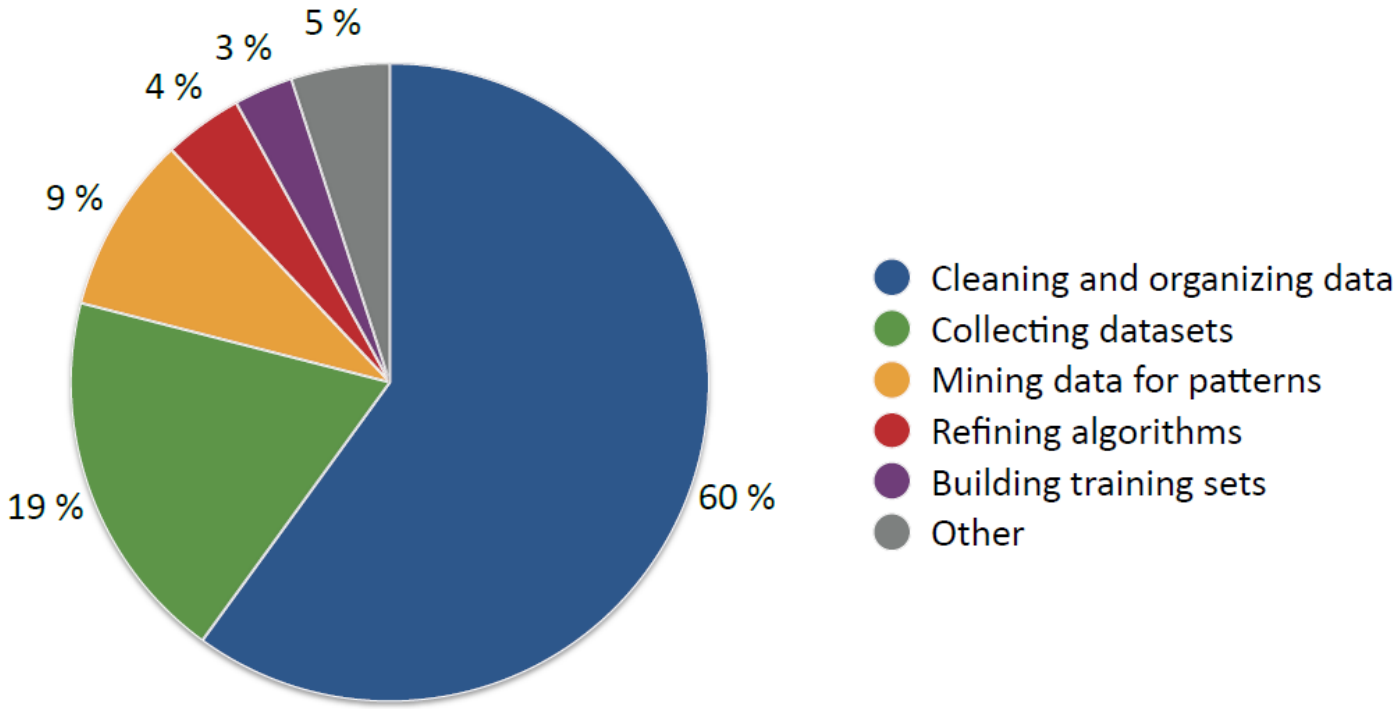
Auxiliary AI GmbH

Preprocessing Pipeleine

- Warum Preprocessing?

- Sprache ist etwas schönes und Kontextabhängig. Wörter sind aber nicht immer alle gleich relevant und es gibt zudem viele grammatikalische Fälle, die syntaktisch Texte voneinander entfernen, auch wenn diese semantisch dich beieinander liegen.

Erinnerung: Data Science Aufwendungsverteilung



http://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#532bc2887f75

Auxiliary AI GmbH

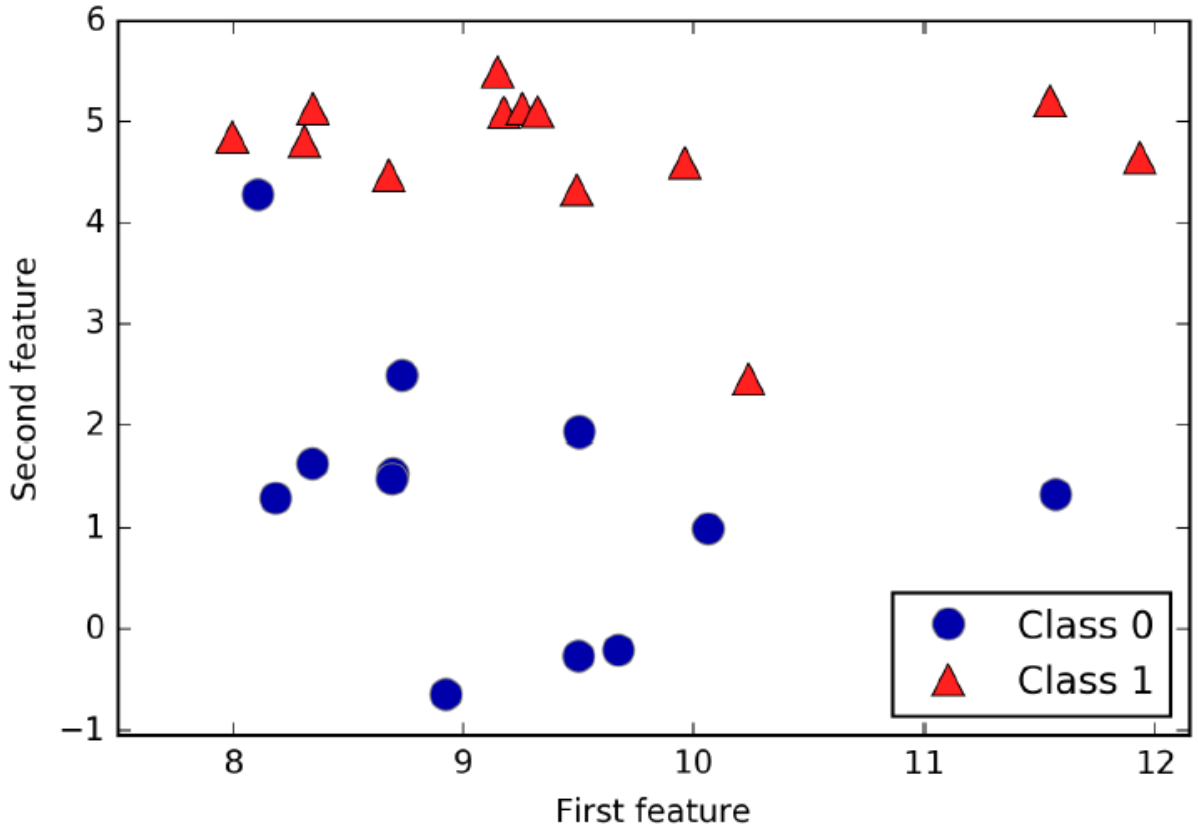
Klassifikation

- Vorhersage eines Wertes oder einer Klasse
- Binäre Klassifizierung: y ist ein Element von {-1, +1}
- Mehrfachklassifizierung: y ist ein Element von {0,1,2,3, ... ,n}

- Die Klassifizierung verwendet Merkmale (X) und Bezeichnungen (y) und ist Teil des überwachten Lernens, bei dem wir die Klassen kennen.

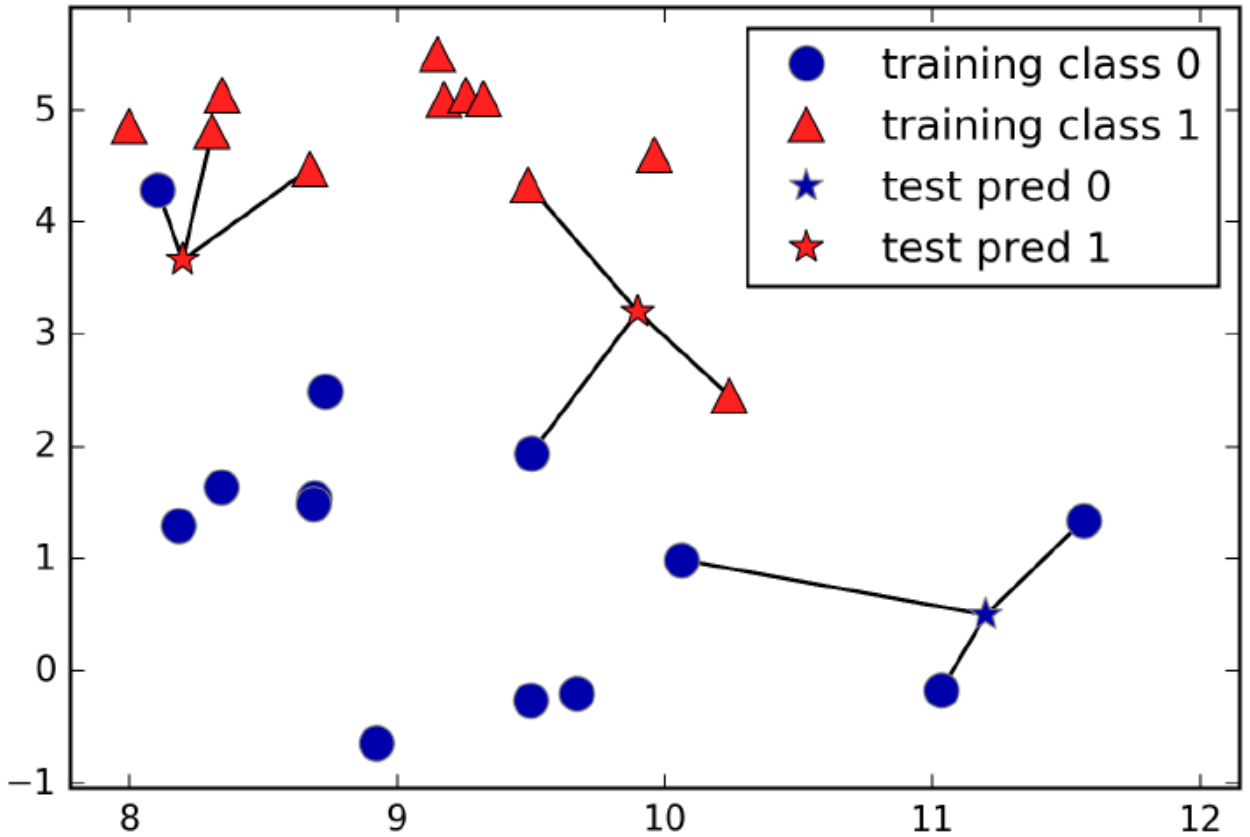
Auxiliary AI GmbH

Klassifikation



Auxiliary AI GmbH

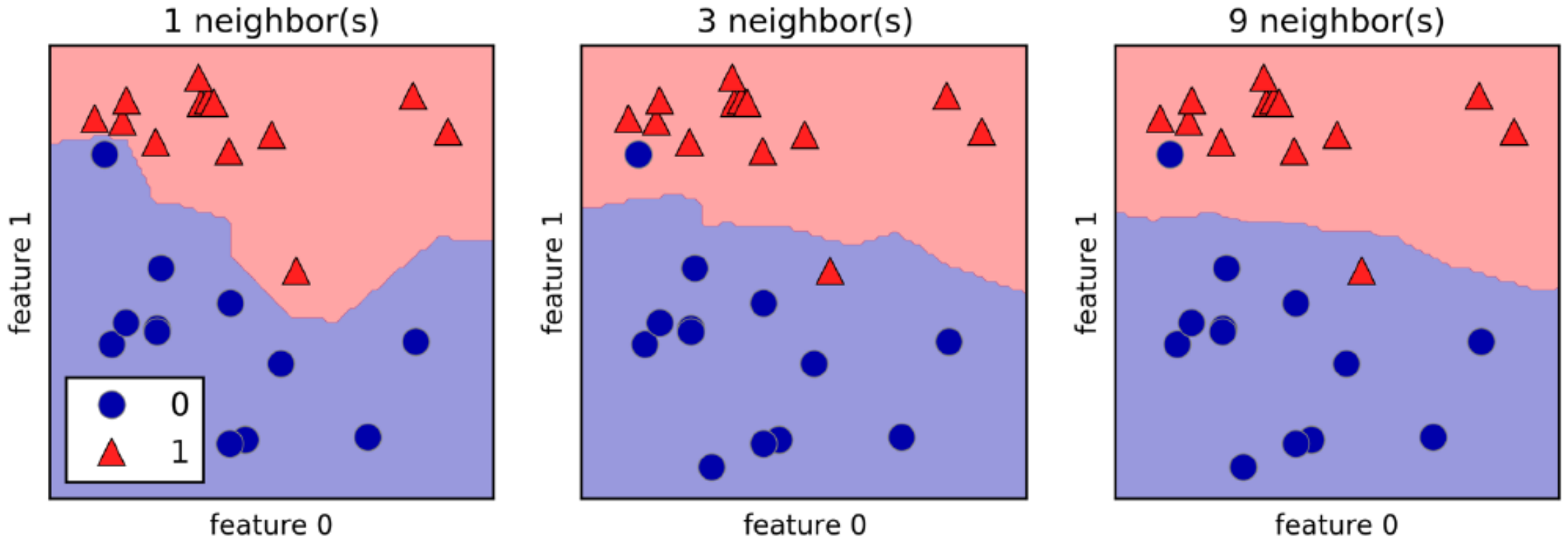
Klassifikation – Nearest Neighbours (n=3)



Picture from ojosujono96 (Freepik)

Auxiliary AI GmbH

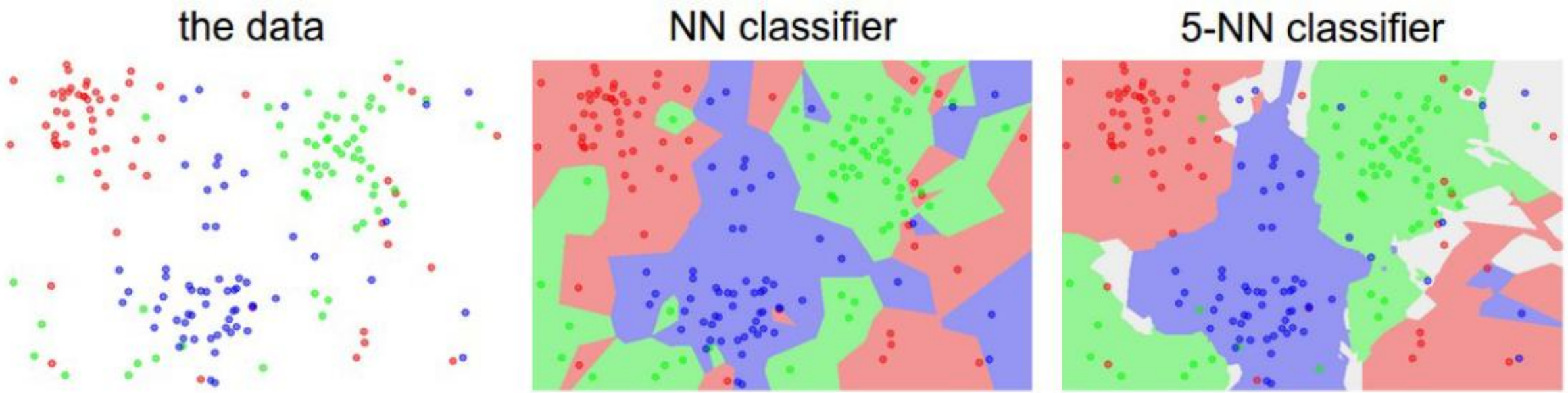
Klassifikation – Nearest Neighbours (n=1-9)



Picture from ojosujono96 (Freepik)

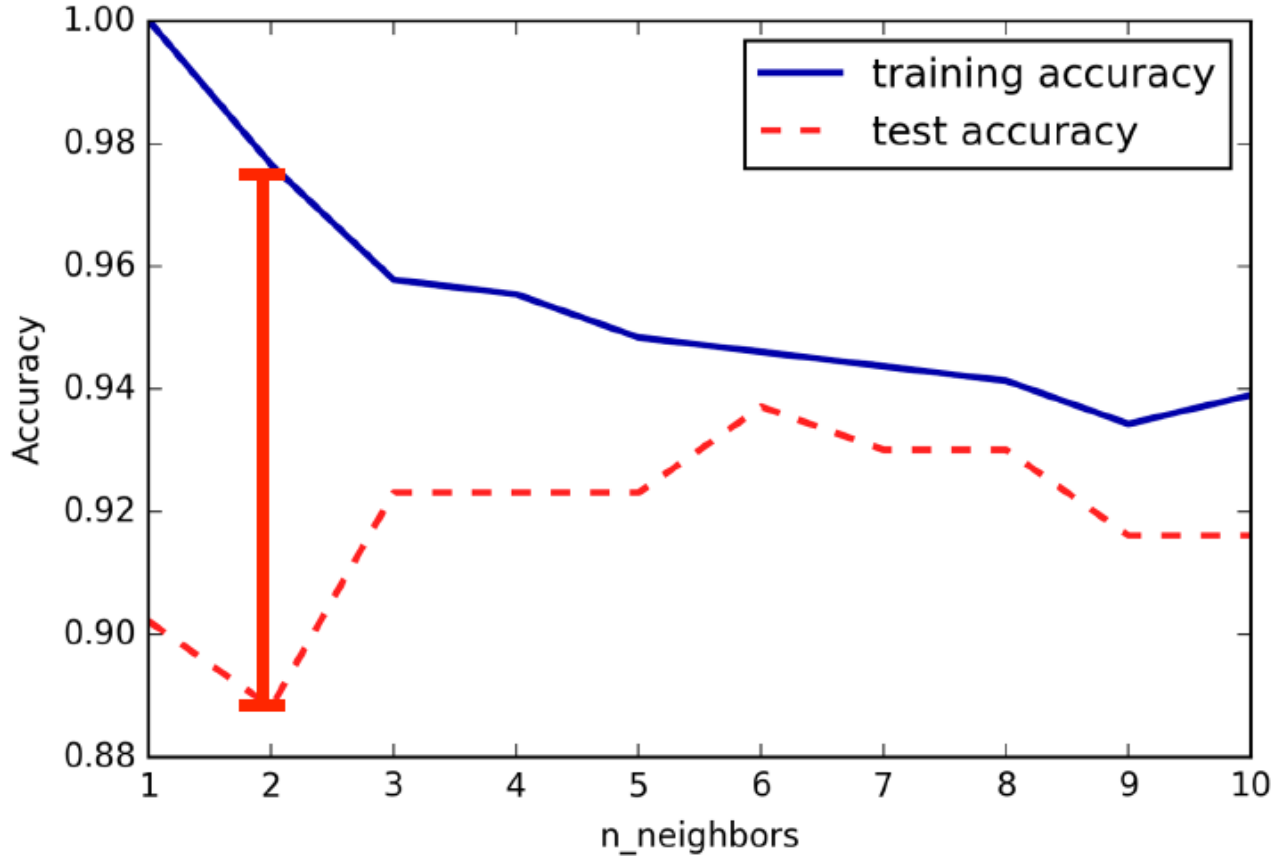
Auxiliary AI GmbH

Klassifikation - Nearest Neighbours



Auxiliary AI GmbH

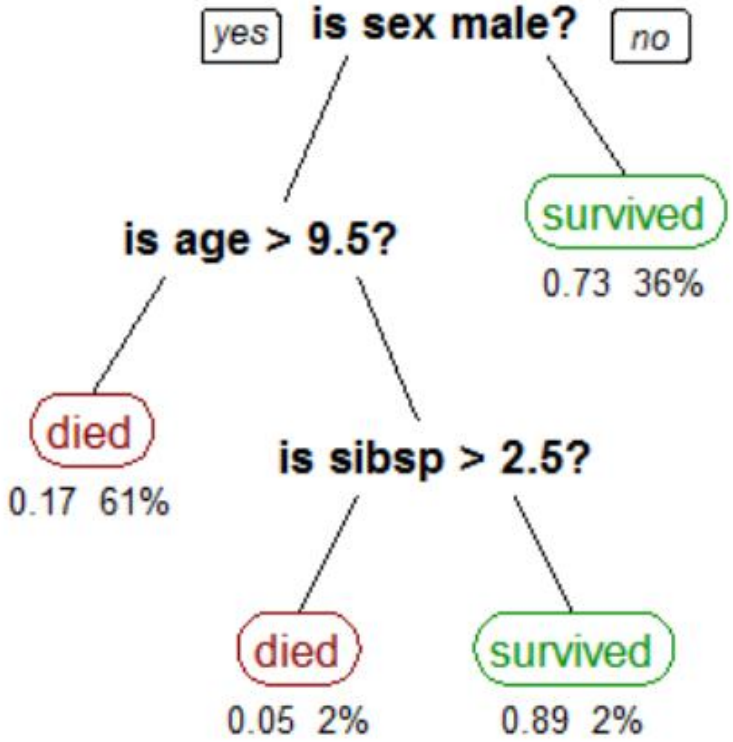
Klassifikation - Nearest Neighbours



Picture from ojosujono96 (Freepik)

Auxiliary AI GmbH

Klassifikation – Decision Tree



Picture from ojosujono96 (Freepik)

Auxiliary AI GmbH

Klassifikation - Evaluation

- **True Positives (TP):** Dies sind die korrekt vorhergesagten positiven Werte, die bedeuten, dass der Wert der tatsächlichen Klasse „ja“ ist und der Wert der vorhergesagten Klasse ebenfalls „ja“ ist. Wenn z. B. der Wert der tatsächlichen Klasse anzeigt, dass dieser Passagier überlebt hat, und die vorhergesagte Klasse das Gleiche aussagt.
- **True Negative (TN):** Dies sind die korrekt vorhergesagten negativen Werte, die bedeuten, dass der Wert der tatsächlichen Klasse nein ist und der Wert der vorhergesagten Klasse ebenfalls nein ist. Wenn z. B. die tatsächliche Klasse besagt, dass dieser Passagier nicht überlebt hat, und die vorhergesagte Klasse das Gleiche aussagt.

Auxiliary AI GmbH

Klassifikation - Evaluation

- **False Positives (FP):** Wenn die tatsächliche Klasse „Nein“ und die vorhergesagte Klasse „Ja“ lautet. Wenn z. B. die tatsächliche Klasse besagt, dass dieser Passagier nicht überlebt hat, die vorhergesagte Klasse aber besagt, dass dieser Passagier überleben wird.
- **Falsch Negative (FN):** Wenn die tatsächliche Klasse „ja“ ist, die vorhergesagte Klasse aber „nein“ ist. Wenn z. B. der tatsächliche Klassenwert angibt, dass dieser Passagier überlebt hat, die vorhergesagte Klasse aber besagt, dass der Passagier sterben wird.

Auxiliary AI GmbH

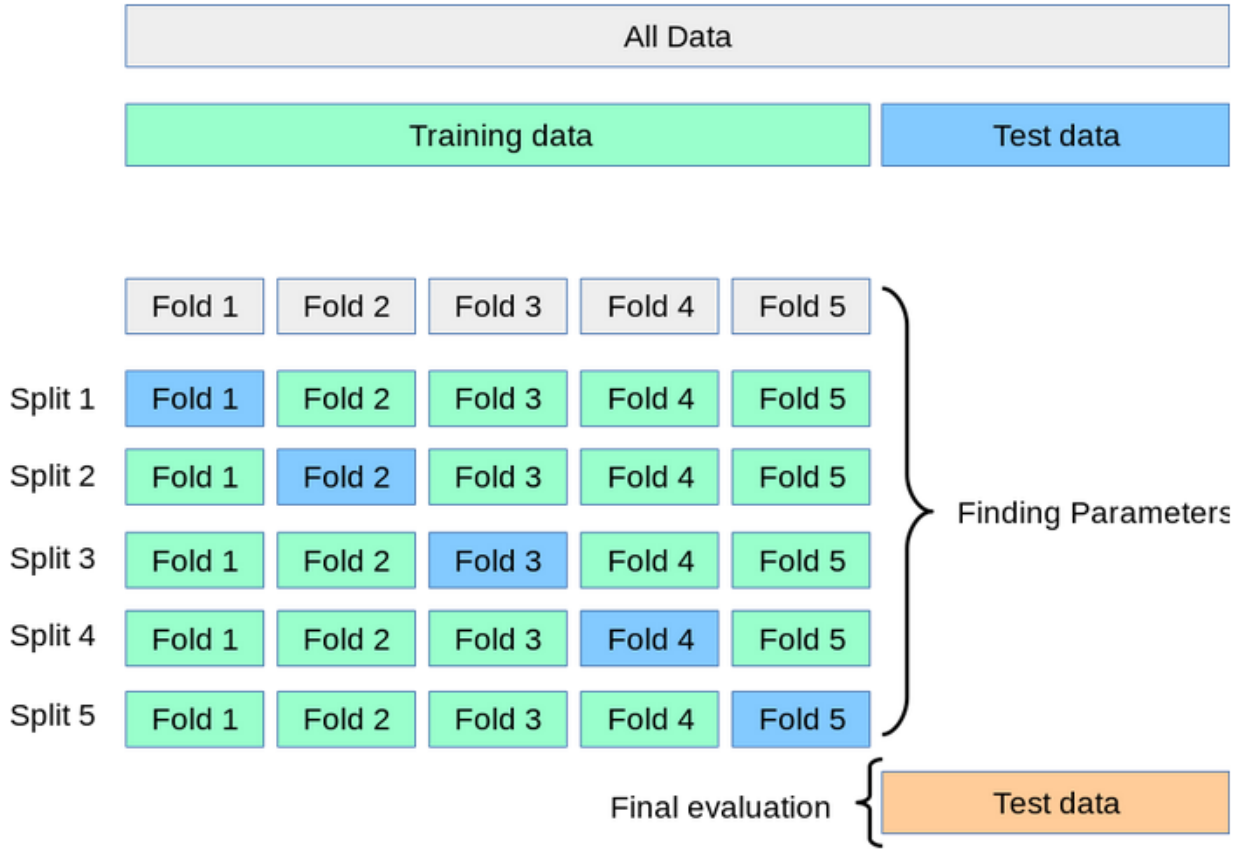
Klassifikation – Evaluation (Confusion Matrix)

- Accuracy (Genauigkeit) = (TP + TN) / (TP + FP + FN + TN)
- Precision (Präzision) = TP / (TP + FP)
- Recall (Sensititvität/Trefferquote) = TP / (TP + FN)

	Predicted class		
Actual Class		Class = Yes	Class = No
	Class = Yes	True Positive	False Negative
	Class = No	False Positive	True Negative

Auxiliary AI GmbH

Klassifikation – Cross-Validation



Klassifikation

- F1 Score = (2 * TP) / (2* TP + FP + FN) = 2* (Precision*Recall)/(Precision+Recall)
- Der F1-Score ist das harmonische Mittel von Präzision und Recall und wird verwendet, um ein ausgewogenes Maß zu erhalten, das sowohl die Präzision als auch den Recall berücksichtigt.

Auxiliary AI GmbH

Auxiliary AI GmbH

Geschäftsführer	**Marten Borchers & Benjamin Klinkigt**
Anschrift	Am Ziegelteich 74 22525 Hamburg Deutschland
Handelsregister	HR B 185519
Registergericht	Amtsgericht Hamburg
Umsatzsteuer-ID	DE 366 814 276
Kontakt	info@auxiliary-ai.de
Webseite	https://auxiliary-ai.de/