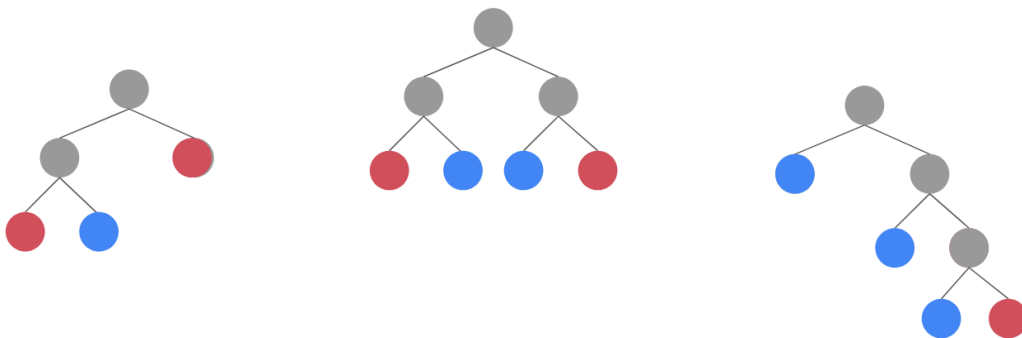


Erläuterung KNN bzw. Random Forest:

Random Forest	Random-Forest ist ein Verfahren für überwachtes Lernen, bei dem mehrere Entscheidungsbäume generiert und Entscheidungen nach dem Mehrheitsprinzip getroffen werden.
k-Nächste-Nachbarn	k-Nächste-Nachbarn ist ein Verfahren für überwachtes Lernen, bei dem die Zugehörigkeit zu einer Klasse bestimmt wird, indem die nächsten k Nachbarn eines Datenpunktes betrachtet werden und die Klassenzugehörigkeit nach dem Mehrheitsprinzip entschieden wird.

Optional: Klassifikation mittels Random Forest-Methode

Variiert man das Lernen von Entscheidungsbäumen, können je nach zur Entscheidung herangezogenen Merkmalen verschiedene Entscheidungsbäume resultieren. Anstelle eines einzelnen Entscheidungsbaumes kann es hilfreich sein, die Entscheidung mehrerer Entscheidungsbäume zu kombinieren, um die Klassifikation zu verbessern. Häufig wird eine Art von Zufall bei der Erstellung der Bäume gewählt, weshalb man dann bei diesem "Wald von Entscheidungsbäumen" von einem Random Forest¹ spricht



Optional: Klassifikation mittels k-Nearest-Neighbour

Die verschiedenen Merkmale der Daten lassen sich als Vektoren interpretieren, bei dem jede Dimension ein Merkmal darstellt. Bei der k-Nearest-Neighbour-Klassifikation (oder k-Nächste-Nachbarn-Klassifikation) wird die Klasse eines Datensatzes bestimmt, indem seine nächsten k Nachbarn in diesem Raum betrachtet werden und die Klassenzugehörigkeit nach dem Mehrheitsprinzip entschieden wird: Gehört die Mehrheit der k nächsten Nachbarn der Klasse A an, erhält auch der zu klassifizierende Datenpunkt die Klasse A. Das Lernen beschränkt sich in diesem Fall auf ein Abspeichern der Trainingsbeispiele.

¹ Englisch für: Zufälliger Wald