

## Arbeitsblatt – Trainieren Sie einen Entscheidungsbaum

Dieses Arbeitsblatt gehört zum einführenden Themenbereich des Kurses KI und Maschinelles Lernen, und zwar zu den Themen A2-1 Maschinelles Lernen sowie A2-2 KI-Modelle. Sie sollen „von Hand“ ein Klassifikationsmodell trainieren, und zwar einen Entscheidungsbaum. Bitte sehen Sie sich vorher das Video Maschinelles Lernen in Abschnitt A2-1 an, oder lassen Sie sich von Ihrer Lehrkraft entsprechend informieren.

Bei dieser Aufgabe geht es um Eisverkaufsdaten von einem Eisstand – siehe dazu auch das einführende Spiel zur Prognose von Eisverkäufen. Auf dem nächsten Blatt finden Sie Daten dazu, ob der Eisstand aus Rentabilitätsgründen offen oder geschlossen sein sollte. Diese Daten sind Ihre **Trainingsdaten**, in Form von Datenpunkten für verschiedene Tage und verschiedene Wetterbedingungen samt der jeweils korrekten Klasse.

Ihre Aufgabe ist es, einen Entscheidungsbaum zu trainieren, der diese Daten gut **modelliert**, der also für möglichst viele der Datenpunkte die korrekte Entscheidung trifft. Gehen Sie dabei vor wie im Video beschrieben – allerdings wird Ihr Entscheidungsbaum komplexer werden als im Video.

Bitte beachten Sie, dass in den Trainingsdaten auch ein wenig „Zufall“ enthalten ist, dass es also vermutlich nicht möglich sein wird, einen Baum zu finden, der die Daten perfekt modelliert und trotzdem auch für weitere Daten dieser Art plausibel wäre. Versuchen Sie daher nicht, den perfekten Baum zu finden, sondern ihn so zu gestalten, dass maximal ein paar wenige Datenpunkte falsch klassifiziert werden.

Anschließend evaluieren Sie Ihren Baum auf sogenannten Testdaten aus. Das sind Daten, die Sie vorher noch nicht gesehen haben. So kann überprüft werden, ob Ihr Baum ein sinnvolles Modell der Eisverkäufe ist und nicht nur ein extrem spezielles Modell, das genau auf die Trainingsdaten passt. Sie erhalten die korrekten Klassen für die Testdaten am Ende der Stunde von Ihrer Lehrkraft.



Sie benötigen ein Blatt Papier, um darauf Ihren Entscheidungsbaum zu entwerfen, sowie eine Schere.

### Schritte

- Schneiden Sie als erstes die Datenpunkte auf dem nächsten Blatt aus, so dass Sie lauter Kärtchen mit jeweils einem Datenpunkt haben.
- Versuchen Sie, Merkmale zu finden, die die Datenpunkte möglichst gut in homogene Gruppen (lauter Fälle der Klasse „offen“ bzw. lauter Fälle der Klasse „geschlossen“) trennen. Wenn Sie ein Merkmal gefunden haben, sortieren Sie die Kärtchen in zwei Gruppen – eine, die das Merkmal erfüllt,

und eine, die es nicht erfüllt. Notieren Sie auf ihrem Blatt das Merkmal als obersten Knoten des Entscheidungsbaums (z.B.: „Temperatur < 5 Grad“). Zeichnen Sie von diesem Knoten aus einen Zweig nach links mit der Beschriftung „ja“ und einen nach rechts mit der Beschriftung „nein“.

- Fangen Sie jetzt für beide Untergruppen von vorne an: einmal für alle Kärtchen, die das Merkmal erfüllen und einmal für alle, die es nicht erfüllen. Suchen sie für beide ein weiteres Merkmal, das die verbliebenen Kärtchen gut aufteilt, notieren Sie es in Ihrem Baum als neuen Knoten im linken Zweig (für die Gruppe von Kärtchen, die das Merkmal erfüllt) bzw. im rechten Zweig (für die, die es nicht erfüllt). Falls in der Gruppe (fast) ausschließlich Fälle einer Klasse vorkommen, können Sie den Zweig mit einem Blatt der entsprechenden Klasse beenden, ansonsten fangen Sie auch hier wieder für die Untergruppen von vorne an. Falls Sie ein Blatt einfügen, bei dem ein bis zwei Datenpunkte falsch klassifiziert werden, notieren Sie sich für später, welche das sind. Dieser Trainingsprozess ist auch im Video A2-1: Maschinelles Lernen beschrieben.
- Wenn Sie alle Zweige mit Blättern beendet haben, ziehen sie ein paar Kärtchen und kontrollieren Sie, ob der Baum für diese auch tatsächlich die korrekte Vorhersage macht. Falls Sie unsicher sind, wie die Vorhersage funktioniert, schauen Sie sich das Video KI-Modelle – Beispiel Entscheidungsbäume an.
- Optional: Lassen Sie sich anschließend den Baum Ihrer Nachbargruppe geben und probieren Sie deren Baum aus.
- Evaluieren Sie abschließend Ihren Baum auf Testdaten: Wenden Sie Ihren Baum auf die unten abgedruckten Testdaten an und notieren Sie die Entscheidungen.
- Sie erhalten am Ende von Ihrer Lehrkraft die korrekten Vorhersagen und können damit evaluieren, wie viel Prozent der Testdaten Ihr Baum korrekt modelliert.

### Testdaten:

Tag	Bewölkung	Temperatur
Dienstag, 2.1.2018	stark bewölkt	7
Mittwoch, 24.1.2018	bewölkt	13
Donnerstag, 15.2.2018	stark bewölkt	5
Freitag, 9.3.2018	stark bewölkt	11
Samstag, 31.3.2018	stark bewölkt	12
Sonntag, 22.4.2018	mäßige Bewölkung	28
Montag, 14.5.2018	stark bewölkt	15
Dienstag, 5.6.2018	heiter bis wolzig	30
Mittwoch, 27.6.2018	schwach bewölkt	25
Donnerstag, 19.7.2018	heiter bis wolzig	29

Montag, 9 Grad, bewölkt	geschlossen	Freitag, 23 Grad, mäßige Bewölkung	offen	Dienstag, 3 Grad, stark bewölkt	geschlossen
Freitag, 4 Grad, stark bewölkt	geschlossen	Dienstag, 22 Grad, bewölkt	offen	Samstag, 9 Grad, stark bewölkt	offen
Dienstag, 9 Grad, stark bewölkt	geschlossen	Samstag, 27 Grad, mäßige Bewölkung	offen	Mittwoch, 7 Grad, stark bewölkt	geschlossen
Samstag, 3 Grad, stark bewölkt	geschlossen	Mittwoch, 28 Grad, schwach bewölkt	offen	Sonntag, 6 Grad, stark bewölkt	geschlossen
Mittwoch, 4 Grad, strahlend blauer Himmel	geschlossen	Sonntag, 31 Grad, mäßige Bewölkung	offen		
Sonntag, -1 Grad, strahlend blauer Himmel	geschlossen	Donnerstag, 34 Grad, mäßige Bewölkung	offen		
Donnerstag, 12 Grad, bewölkt	offen	Montag, 29 Grad, strahlend blauer Himmel	offen		
Montag, 0 Grad, stark bewölkt	geschlossen	Freitag, 16 Grad, stark bewölkt	offen		
Freitag, 15 Grad, bewölkt	offen	Dienstag, 28 Grad, mäßige Bewölkung	offen		
Dienstag, 20 Grad, stark bewölkt	offen	Samstag, 19 Grad, stark bewölkt	offen		
Samstag, 27 Grad, schwach bewölkt	offen	Mittwoch, 16 Grad, bewölkt	offen		
Mittwoch, 17 Grad, stark bewölkt	geschlossen	Sonntag, 26 Grad, strahlend blauer Himmel	offen		
Sonntag, 17 Grad, stark bewölkt	offen	Donnerstag, 14 Grad, stark bewölkt	geschlossen		
Donnerstag, 22 Grad, bewölkt	offen	Montag, 15 Grad, mäßige Bewölkung	offen		
Montag, 30 Grad, heiter bis wolkig	offen	Freitag, 10 Grad, schwach bewölkt	offen		