

LU02.L01 - Fallstudio Machine Learning

Auftrag A: - Grundlagen (Theorie)

1. Supervised Learning VS Unsupervised Learning

- **Supervised Learning:** Das Modell lernt mit gelabelten Daten, d. h. Eingabedaten und die richtige Antwort sind bekannt. Klassischer Frontalunterricht.
- **Unsupervised Learning:** Das Modell bekommt keine Zielvariable und versucht selbst, Muster oder Strukturen zu finden. Gruppenarbeit ohne Aufgabenstellung.

2. Handelt es sich bei diesem Problem um (mit Begründung): Klassifikation oder Regression?

- **Klassifikation**
- Begründung: Die Zielvariable Kaufentscheidung hat diskrete Werte (Ja / Nein). Zahlen vorhersagen wäre Regression, hier geht es um Kategorien.

3. Nenne Sie zwei typische Algorithmen, die für dieses Problem geeignet wären, und begründen Sie jeweils in einem Satz.

- **Logistische Regression:** Gut geeignet für binäre Klassifikationsprobleme und leicht interpretierbar (im Gegensatz zu manchem Buzzword-Modell).
- **Decision Tree:** Trifft Entscheidungen anhand klarer Regeln und ist auch für Menschen nachvollziehbar – ein seltenes Gut.

Auftrag B: - Datenverständnis

4. Welche der oben genannten Merkmale sind numerisch, welche kategorisch?

- **Numerisch:**
 - Alter
 - Anzahl der bisherigen Einkäufe
 - Verweildauer auf der Produktseite
- **Kategorisch:**
 - Kaufentscheidung (Ja / Nein)

5. Warum ist es problematisch, die Zielvariable als Feature zu verwenden?

- Weil das Modell sonst die Antwort bereits kennt.
- Das nennt man Data Leakage und führt zu beeindruckend guten Ergebnissen, die in der Realität genau gar nichts taugen.

Auftrag c: - Praxisdenken

6. Beschreiben Sie in 4-5 Schritten den typischen Ablauf eines Machine-Learning-Projekts (von Daten bis Modellbewertung).

1. Problem definieren (was soll vorhergesagt werden?)
2. Daten sammeln und vorbereiten (bereinigen, skalieren, encoden)
3. Daten in Trainings- und Testdaten aufteilen
4. Modell trainieren
5. Modell evaluieren und verbessern
6. Optionaler Schritt 6: Erkennen, dass alles komplizierter ist als gedacht.

7. Nennen Sie zwei Risiken, die auftreten koennen, wenn:

- zu wenige Trainingsdaten vorhanden sind
 - Modell lernt keine stabilen Muster
 - Schlechte Generalisierung auf neue Daten
- das Modell zu gut auf die Trainingsdaten passt
 - Sehr gute Trainingsgenauigkeit
 - Katastrophale Leistung bei neuen Daten

8. Was ist *Overfitting* - erklart so, dass es auch jemand versteht, der glaubt, KI sei Bewusstsein.

- Overfitting bedeutet, dass ein Modell die Trainingsdaten auswendig lernt, inklusive Zufall und Rauschen, anstatt allgemeine Muster zu erkennen.
- Es ist wie jemand, der eine Matheaufgabe perfekt loest - solange sie exakt gleich aussieht wie im Heft.



Volkan Demir

From:
<https://wiki.bzz.ch/> - **BZZ - Modulwiki**

Permanent link:
<https://wiki.bzz.ch/de/modul/m245/learningunits/lu02/loesungen/l01?rev=1767612874>

Last update: **2026/01/05 12:34**

