

DTL.03: Machine Learning mit Weka

Übungsblatt: Entscheidungsbäume (Decision Tree Learner DTL)

9. November 2025

Aufgabe 3.1 – Training mit J48

Restaurant-Datensatz

Modell: J48 (C4.5) – Parameter: -C 0.25 -M 2

Attributmenge: 11 Attribute, 12 Instanzen.

Ergebnis:

- Baumstruktur: Nur Attribut `patrons` wird verwendet.
- Anzahl der Blätter: 3, Baumgröße: 4.
- Korrekt klassifizierte Instanzen: 10 von 12 (83.33 %).
- Fehlklassifizierte Instanzen: 2 (16.67 %).

Confusion Matrix:

```
a b    <-- classified as
6 0 | a = No
2 4 | b = Yes
```

Interpretation:

Der Baum klassifiziert 83.3 % der Trainingsbeispiele korrekt. Alle `No`-Fälle wurden richtig erkannt, zwei `Yes`-Fälle jedoch fälschlich als `No` klassifiziert. Das Modell nutzt nur das Attribut `patrons`, wodurch eine einfache, nachvollziehbare Entscheidungsregel entsteht.

Aufgabe 3.2 – ARFF-Format

Die ARFF-Dateien `zoo.arff` und `restaurant.arff` wurden erstellt. Dabei wurden alle Attribute als `nominal` definiert, da Entscheidungsbaum-Algorithmen wie ID3 und J48 nominale Attribute bevorzugen.

Attributtypen:

- **nominal**: endliche, ungeordnete Werte (z. B. {`yes`, `no`}).
- **numeric/ordinal**: geordnete, numerische Werte (z. B. Temperatur, Alter).
- **string**: Freitext, nicht für ID3 geeignet.

Aufgabe 3.3 – Vergleich von ID3 und J48

Restaurant-Datensatz

ID3:

- Korrekt klassifiziert: 12 von 12 (100 %)
- Baumattribute: `patrons`, `type`, `fri_sat`, `alt`

J48:

- Korrekt klassifiziert: 10 von 12 (83.33 %)
- Baumattribute: nur `patrons`

Vergleich:

ID3 erreicht auf den Trainingsdaten 100 % Genauigkeit, da es keinen Pruning-Schritt durchführt. J48 hingegen schneidet den Baum zurück und erzielt dadurch eine etwas geringere Genauigkeit, vermeidet aber Überanpassung und generalisiert besser.

Zoo-Datensatz (J48)

- Korrekt klassifiziert: 100 von 101 (99.01 %)
- Fehlklassifikationen: 1 (ein Amphibium als Reptil klassifiziert)

Confusion Matrix:

a	b	c	d	e	f	g	<-- classified as
3	0	0	0	0	1	0	a = amphibian
0	20	0	0	0	0	0	b = bird
0	0	13	0	0	0	0	c = fish
0	0	0	8	0	0	0	d = insect
0	0	0	0	41	0	0	e = mammal
0	0	0	0	0	5	0	f = reptile
0	0	0	0	0	0	10	g = shellfish

Interpretation:

Der Zoo-Datensatz zeigt eine nahezu perfekte Klassifikation (99 % Genauigkeit). Nur ein Amphibium wurde fälschlich als Reptil klassifiziert. Der J48-Baum ist differenziert und nutzt u. a. die Attribute `feathers`, `milk`, `fins` und `tail` zur Klassentrennung.

Fazit

- ID3 liefert perfekte Ergebnisse auf Trainingsdaten, neigt aber zur Überanpassung (kein Pruning).
- J48 wendet Pruning an, ist kompakter und robuster gegenüber Rauschen.
- In beiden Datensätzen zeigen sich nachvollziehbare und interpretierbare Entscheidungsbäume.