

ID3

Ich werde nicht alle Rechnungen hier aufschreiben, die ich per Hand gemacht habe, sondern nur die Zwischenergebnisse, damit es etwas übersichtlicher wird. Und ich schreibe für den ersten Durchlauf die Schritte auf, danach wiederhole ich sie nur.

Durchlauf 1:

Entropie der Gesamtmenge berechnen:

$H(S)$ von O4 und M3 = 0,99

Rest Entropie für x_1 (Alter) berechnen:

\geq sind 4/7, davon O2 M2

$<$ sind 3/7, davon O2 M1

Ergebnis: 0,96

Rest Entropie für x_2 (Einkommen) berechnen:

hoch sind 4/7, davon O3 M1

niedrig sind 3/7, davon O1 M2

Ergebnis: 0,66

Rest Entropie für x_3 (Bildung) berechnen:

Abi sind 3/7, davon O1 M2

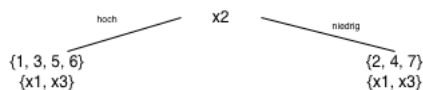
Bachelor sind 2/7, davon O1 M1

Master sind 2/7, davon O2

Ergebnis: 0,68

Gain berechnen: $H(S) - R(S,A)$ der höchste Wert wird für Split genommen. Hier x_2 mit 0,33

Vorläufiger Baum:



Neue Mengen nach hoch und niedrig:

$\{1, 3, 5, 6\}$, und $\{2, 4, 7\}$

Erst rechte Seite:

hoch:

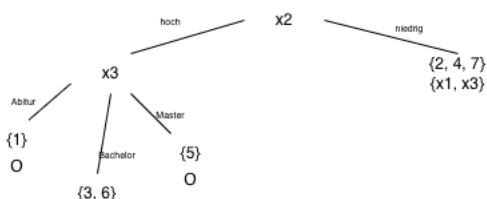
Gesamt-Entropie: 0,66

x_1 -Entropie: 0,64,

x_3 -Entropie: 0,5

Höchster Gain: 0,16 bei x_3 .

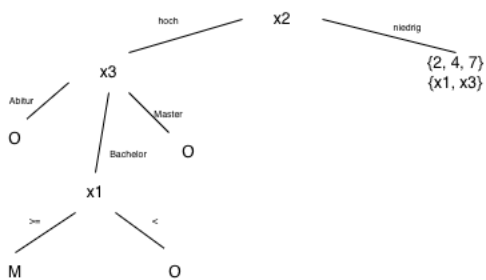
Split in die Mengen: Abitur, Bachelor, Master $\{1\}, \{3,6\}, \{5\}$, Wir haben also schon zwei Blätter mit O.



Die Menge $\{x_3\}$ habe ich in dem Diagramm vergessen.

3 und 6 müssen nochmal gesplittet werden, nach letztem Attribut alter. Da nur noch zwei in der Menge sind, wird nach der Ausprägung gesplittet.

Vorläufiger Baum:



Damit ist die rechte Seite abgearbeitet.

Linke Seite:

Gesamt-Entropie: 0,92

x1-Entropie: 0,033,

x3-Entropie: 0 (perfekter Split)

Höchster Gain: 0,92 bei x3.

Split in {2} und {4, 7}

Damit haben die Mengen jeweils dieselben Klassen und es wird beendet.

