

## Aufgabe 3.1

a)

Zusammenhang: Entscheidungsbäumen und Raumpartitionierungsverfahren

Beide Verfahren teilen den Raum Schrittweise auf.

Entscheidungsbäumen nach den ID3-Algorithmus versuchen in möglichst wenig Schritten die Punkte zu Klassifizieren. Neue Punkte durchlaufen den Baum und werden so zugeordnet.

Bei der Raumpartitionierung wird geschaut wo der Punkt liegt und wird so dem entsprechendem Raum zugeordnet.

b)

Siehe Abbildung.

Es kann nicht jeden Person eindeutig identifiziert werden da es Personen mit gleicher Merkmalsausprägung gibt. (z.B.: Ute und Tina [weiblich, keine Brille, jung, blond, keinen Bart, Ohringe])

c)

(siehe unten)

Jede Person ist eine eigene Klasse, daher 24 Klassen.

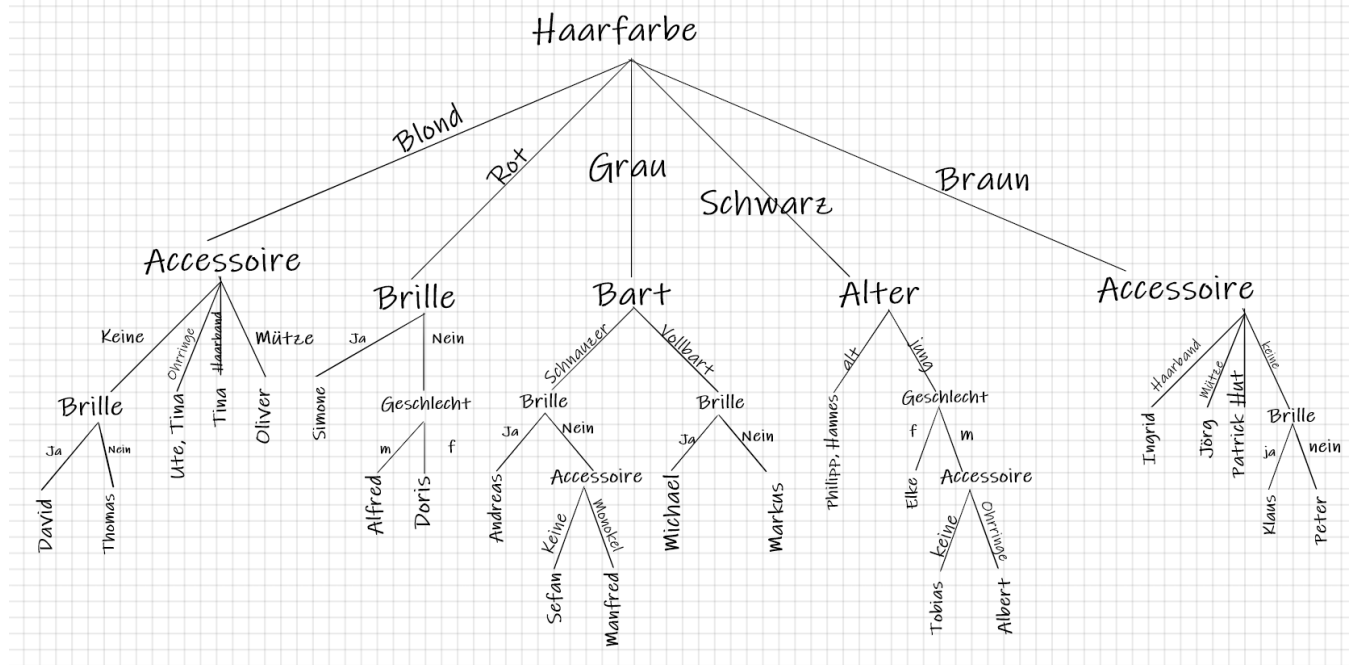
Es wird angenommen dass es jede Person nur einmal gibt. (auch Tina)

- Gesamtentropie = 4,58
- Teilentropien zu dem Merkmal Geschlecht =>  $m=4,08$        $f=2,8$

Erfahrungsgewinn/Informationsgewinn:

- Geschlecht = 0,878
- Brille = 0,9
- Alter = 0,95
- Haarfarbe = 2,29
- Bart = 1,46
- Accessoire = 2,04

Den höchsten Informationsgewinn liefert die Haarfarbe. Es ist daher sinnvoll die Gruppe zuerst nach der Haarfarbe zu unterteilen.



### Blatt 3

#### Aufgabe 3.1

a) Zusammenhang zwischen Entscheidungsbäumen und Raumpartitionierung

b) ID3-Algorithmus

Gesamtentropie:  $-\left(\frac{7}{24}\right)\log_2\left(\frac{7}{24}\right) - \left(\frac{17}{24}\right)\log_2\left(\frac{17}{24}\right) = 4,58$

Entropie:

Geschlecht: m:  $-\left(\frac{7}{17}\right)\log_2\left(\frac{7}{17}\right) = 4,08$

f:  $-\left(\frac{10}{17}\right)\log_2\left(\frac{10}{17}\right) = 2,8$

Informationsgewinn:

$4,58 - \left(\frac{17}{24}\right) \cdot 4,08 - \left(\frac{7}{24}\right) \cdot 2,8 = 0,878$

Brille: ja:  $-\left(\frac{3}{8}\right)\log_2\left(\frac{3}{8}\right) = 3$

nein:  $-\left(\frac{5}{16}\right)\log_2\left(\frac{5}{16}\right) = 4$

Informationsgewinn:

$4,58 - \left(\frac{8}{24}\right) \cdot 3 - \left(\frac{16}{24}\right) \cdot 4 = 0,9$

Alter: jung:  $-\left(\frac{7}{15}\right)\log_2\left(\frac{7}{15}\right) = 3,9$

alt:  $-\left(\frac{9}{15}\right)\log_2\left(\frac{9}{15}\right) = 3,17$

Informationsgewinn:

$4,58 - \left(\frac{15}{24}\right) \cdot 3,9 - \left(\frac{9}{24}\right) \cdot 3,17 = 0,95$

Bart: keinen:  $-\left(\frac{7}{14}\right)\log_2\left(\frac{7}{14}\right) = 3,8$

schnauzer:  $-\left(\frac{7}{7}\right)\log_2\left(\frac{7}{7}\right) = 2,8$

Vollbart:  $-\left(\frac{2}{2}\right)\log_2\left(\frac{2}{2}\right) = 1$

Informationsgewinn:

$4,58 - \left(\frac{14}{24}\right) \cdot 3,8 - \left(\frac{7}{24}\right) \cdot 2,8 - \left(\frac{2}{24}\right) \cdot 1 = 1,46$

Haarfarbe: blond:  $-\left(\frac{6}{24}\right)\log_2\left(\frac{6}{24}\right) = 2,58$

schwarz:  $-\left(\frac{5}{24}\right)\log_2\left(\frac{5}{24}\right) = 2,32$

Braun:  $-\left(\frac{5}{24}\right)\log_2\left(\frac{5}{24}\right) = 2,32$

Grau:  $-\left(\frac{5}{24}\right)\log_2\left(\frac{5}{24}\right) = 2,32$

Rot:  $-\left(\frac{3}{24}\right)\log_2\left(\frac{3}{24}\right) = 1,58$

Informationsgewinn:

$4,58 - \left(\frac{6}{24}\right) \cdot 2,58 - \left(\frac{5}{24}\right) \cdot 2,32 \cdot 3 - \left(\frac{3}{24}\right) \cdot 1,58 = 2,29$

Accessoire: kein:  $-\left(\frac{12}{24}\right)\log_2\left(\frac{12}{24}\right) = 3,58$

Ohrringe:  $-\left(\frac{4}{24}\right)\log_2\left(\frac{4}{24}\right) = 2$

Haarband:  $-\left(\frac{4}{24}\right)\log_2\left(\frac{4}{24}\right) = 2$

Mütze:  $-\left(\frac{2}{24}\right)\log_2\left(\frac{2}{24}\right) = 1$

Hut:  $-\left(\frac{1}{24}\right)\log_2\left(\frac{1}{24}\right) = 0$

Monokel:  $-\left(\frac{1}{24}\right)\log_2\left(\frac{1}{24}\right) = 0$

Informationsgewinn:

$4,58 - \left(\frac{12}{24}\right) \cdot 3,58 - \left(\frac{4}{24}\right) \cdot 2 \cdot 2 - \left(\frac{2}{24}\right) \cdot 1 = 2,04$