Decision Tree - Buy Computer

1. Calculate entropy of the target:

Entropy (Buys_computer) = Entropy (12,8) = Entropy (0,6 , 0,4)
= - (0,6 $\log_2$ 0,6) - (0,4 $\log_2$ 0,4)
≈ 0,97

2. Calculate Entropy for each branch (t and then the gain)

Column Credit Rating: 10 Fair — yes 7, no 3
10 Excellent — yes 5, no 5

→ Entropy (Credit_Rating, Fair) =
Entropy (7,3) = Entropy (0,7 , 0,3) = - (0,7 $\log_2$ 0,7) -
(0,3 $\log_2$ 0,3) ≈ 0,88

Entropy (Credit_Rating, Excellent) = Entropy (5,5) = Entropy (0,5 0,5)
= - (0,5 $\log_2$ 0,5) - (0,5 $\log_2$ 0,5) = 1

↳ Gain (Buys_computer, Credit_rating) = 0,97 - $\frac{10}{20}$ · 0,88 -
$\frac{10}{20}$ · 1 = 0,03

Column Student: 9 yes — yes 8, no 1
11 no — yes 7, no 4

→ Entropy (Student, Yes) = Entropy (8,1) = Entropy (0,88; 0,11)
= - (0,88 · $\log_2$ 0,88) - (0,11 $\log_2$ 0,11) ≈ 0,51

Entropy (Student, No) = Entropy (7,4) = Entropy (0,64 , 0,36)
= - (0,64 · $\log_2$ 0,64) - (0,36 $\log_2$ 0,36) ≈ 0,94

↳ Gain (Buys_computer, Student) = 0,97 - $\frac{9}{20}$ · 0,51 - $\frac{11}{20}$ · 0,94
≈ 0,22

Column Income: 5 high — yes 3, no 2
8 medium — yes 5, no 3
7 low — yes 4, no 3

→ Entropy (Income, High) = Entropy (3,2) = Entropy (0,6; 0,4)
≈ 0,97

Entropy (Income, Medium) = Entropy (5,3) = Entropy (0,625 , 0,375)
= - (0,625 $\log_2$ 0,625) - (0,375 $\log_2$ 0,375)
≈ 0,95

Entropy (Income, Low) = Entropy (4,3) = Entropy (0,57, 0,43)
= 0,99

↳ Gain (Buys-computer, Income) = 0,97 - $\frac{5}{20}$ · 0,97 -
$\frac{8}{20}$ · 0,95 - $\frac{7}{20}$ · 0,99
= 0,001

Column Age: 8 <= 30 — yes - 2, no - 6
6 31...40 — yes - 4, no - 0
6 >40 — yes - 4, no - 2

→ Entropy (Age, <=30) = Entropy (2,6) = Entropy (0,25, 0,75)
= 0,188

→ Entropy (Age, 31...40) = Entropy (4,0) = 0

→ Entropy (Age, >40) = Entropy (4,2) = Entropy (0,67, 0,33)
≈ 0,92

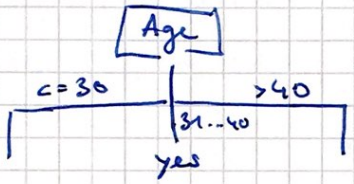
↳ Gain (Buys_computer, Age) = 0,97 − $\frac{8}{20}$ · 0,188 − $\frac{6}{20}$ · 0 −
$\frac{6}{20}$ · 0,92 = 0,6188

↳ Übrlich:

Gain (Credit_ranking): 0,03 Gain (Income): 0,001
Gain (Student): 0,22 Gain (Age): 0,6188
Gain (REC): nicht betrachtet, da 20 verschied. Werte

↳ Age hat den höchsten Gain

↳ erste Ebene:



3. Iterativ weiter für jeden Branch:

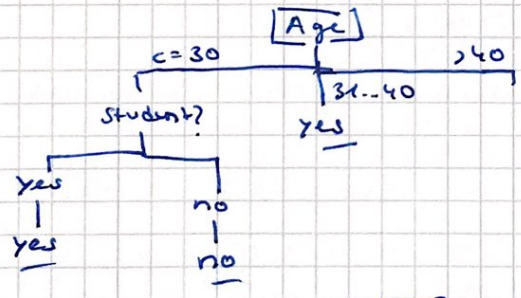
- <=30 Branch: Target Entropy hier: 0,188

Column Student: 2 yes — yes 2
6 no — no 0
yes 0
no 6

Entropy = H

Entropy (Student, yes) = H(2,0) = H(1,0) = 0 → Leaf
Entropy (Student, no) = H(0,6) = H(0,1) = 0 → Leaf

↳ da 2 Leaf direkt expandieren:



- >40 Branch: Entropy hier: 0,92

Column Student: 4 yes — yes 3
no 1
2 no — yes 1
no 1

Entropy (Student, Yes) = H(3,1) = H(0,75, 0,25) ≈ 0,81
Entropy (Student, No) = H(1,1) = H(0,5, 0,5) = 1

Gain (Student) = 0,92 − $\frac{4}{6}$ · 0,81 − $\frac{2}{6}$ · 1 ≈ 0,047

Column Credit rating: 3 Fair — yes 3
3 Excellent — no 0
1
2

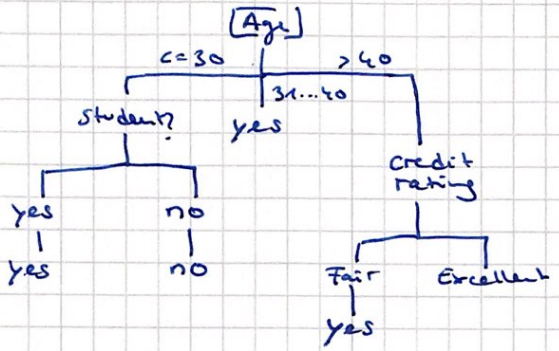
Entropy (Credit_rating, Fair) = H(3,0) = 0 → Leaf
Entropy (Credit_rating, Excellent) = H(1,2) = H(0,33, 0,67) ≈ 0,46

Column Income: 0 High → yes 3, no 1
4 Medium → yes 1, no 1
2 Low → yes 1, no 1

Entropy(Income, Medium) = $H(3,1) = H(0{,}75, 0{,}25) \approx 0{,}81$
Entropy(Income, Low) = $H(1,1) = H(0{,}5, 0{,}5) = 1$

Gain(Income) = $0{,}92 - \frac{4}{6} \cdot 0{,}81 - \frac{2}{6} \cdot 1 \approx 0{,}047$

↳ Highest Gain: Credit_rating:



4. Letzter Branch: >40 & credit rating Excellent:

Entropy(Income) → high, medium 2 → yes 1, no 1; low 1 → no 1

Entropy(Student) → yes 2 → yes 1, no 1; no 1 → yes 0, no 0

⇒ Nur noch sehr wenige Datensätze greifen hier & diese sind sehr durchmischt. Keine klare Aussage treffbar hier