

Aufgabe 17:

Tabelle 4: Datensatz: „Soll ich Fußballspielen gehen?“

Temperatur / °C	Wettervorhersage	Luftfeuchtigkeit / %	Wind	Fußball
29,4	2	85	False	False
26,7	2	90	True	False
28,3	1	78	False	True
21,1	0	96	False	True
20	0	80	False	True
18,3	0	70	True	False
17,8	1	65	True	True
22,2	2	95	False	False
20,6	2	70	False	True
23,9	0	80	False	True
23,9	2	70	True	True
22,2	1	90	True	True
27,2	1	75	False	True
21,7	0	80	True	False

a) Entropie der Wurzel:

$$\text{Entropie: } H(Y) = - \sum_{z \in Z} P(Y=z) \log_2 (P(Y=z))$$

hier $Z = \{\text{True}, \text{False}\} = \{F, T\} = \{\text{Fußball}, \text{nicht Fußball}\}$

$Y \triangleq \text{Fußball}$

$$H(Y) = - \frac{n_T}{n} \cdot \log_2 \frac{n_T}{n} - \frac{n_F}{n} \log_2 \frac{n_F}{n}$$

$$n=14$$

$$n_T = 9 \quad n_F = 5$$

$$H(Y) = - \frac{9}{14} \cdot (\log \frac{9}{14}) - \frac{5}{14} \cdot \log_2 \frac{5}{14}$$

$$= 0.940286 \dots$$

b)

Informationsgewinn bei Schnitt auf Attribut Wind

Informationsgewinn:

$$IG(X, Y) = H(Y) - H(Y|X)$$

$$H(Y) = - \sum_{z \in Z} P(Y=z) \log_2 (P(Y=z))$$

$$H(Y|X) = \sum_{m \in M} P(X=m) H(Y|X=m)$$

$$= - \sum_{m \in M} (P(X=m) \cdot \sum_{z \in Z} P(Y=z | X=m) \log_2 (P(Y=z | X=m)))$$

hier (Wind): $M = \{W, \neg W\} \quad Z = \{F, \neg F\}$

hier (X_{Wind}): $M = \{W, \neg W\}$ $Z = \{F, \neg F\}$

$$H(Y|X_{\text{Wind}}) = -P(X=W) \cdot \left(P(Y=F|X=W) \log_2 \left(P(Y=F|X=W) + P(Y=\neg F|X=W) \log_2 \left(P(Y=\neg F|X=W) \right) \right) \right. \\ \left. - P(X=\neg W) \cdot \left(P(Y=F|X=\neg W) \log_2 \left(P(Y=F|X=\neg W) + P(Y=\neg F|X=\neg W) \log_2 \left(P(Y=\neg F|X=\neg W) \right) \right) \right) \right)$$

$$H(Y|X_{\text{Wind}}) = - \frac{n(W)}{n} \cdot \left(\frac{n(F \wedge W)}{n(W)} \cdot \log_2 \left(\frac{n(F \wedge W)}{n(W)} \right) + \frac{n(\neg F \wedge W)}{n(W)} \log_2 \left(\frac{n(\neg F \wedge W)}{n(W)} \right) \right) \\ - \frac{n(\neg W)}{n} \cdot \left(\frac{n(F \wedge \neg W)}{n(\neg W)} \log_2 \left(\frac{n(F \wedge \neg W)}{n(\neg W)} \right) + \frac{n(\neg F \wedge \neg W)}{n(\neg W)} \log_2 \left(\frac{n(\neg F \wedge \neg W)}{n(\neg W)} \right) \right)$$

$$n = 14$$

$$n(W) = 6 \quad n(\neg W) = 8$$

$$n(F \wedge W) = 3 \quad n(\neg F \wedge W) = 3$$

$$n(F \wedge \neg W) = 6 \quad n(\neg F \wedge \neg W) = 2$$

$$H(Y|X_{\text{Wind}}) = - \frac{1}{14} \left(3 \log_2 \left(\frac{3}{6} \right) + 3 \log_2 \left(\frac{3}{6} \right) \right. \\ \left. - \frac{1}{14} \left(6 \cdot \log_2 \left(\frac{6}{8} \right) + 2 \log_2 \left(\frac{2}{8} \right) \right) \right) \\ = 0.892159$$

$$G(Y|X_{\text{Wind}}) = H(Y) - H(Y|X_{\text{Wind}})$$

$$= 0.940286 - 0.892159$$

$$= 0.048127$$