

Data Mining Übungsblatt 4

1.1

$$\text{supp}(\text{Buch1}) = \frac{6000}{10000} = \frac{3}{5} \Rightarrow 60\%$$

$$\text{supp}(\text{Buch2}) = \frac{8000}{10000} = \frac{4}{5} \Rightarrow 80\%$$

$$\text{supp}(\text{Beide}) = \frac{4000}{10000} = \frac{2}{5} \Rightarrow 40\%$$

Konfidenz Buch 1 \rightarrow Buch 2

$$\frac{\frac{2}{5}}{\frac{3}{5}} = \frac{2}{5} \cdot \frac{5}{3} = \frac{10}{15} = \frac{2}{3} \Rightarrow 66\%$$

Konfidenz Buch 2 \rightarrow Buch 1

$$\frac{\frac{2}{5}}{\frac{4}{5}} = \frac{2}{5} \cdot \frac{5}{4} = \frac{10}{20} = \frac{1}{2} \Rightarrow 50\%$$

① keine von beiden, weil die Konfidenz unter 75% liegt.

1.2 Lift Buch 1 \rightarrow Buch 2

$$\begin{aligned} \frac{\text{supp}(\text{Beide})}{\text{supp}(\text{Buch1}) \cdot \text{supp}(\text{Buch2})} &= \frac{\frac{2}{5}}{\frac{3}{5} \cdot \frac{4}{5}} = \frac{\frac{2}{5}}{\frac{12}{25}} = \frac{2}{5} \cdot \frac{25}{12} \\ &= \frac{5}{6} \Rightarrow 83\% \leftarrow 0.83 \end{aligned}$$

Durch den Lift mit 0.83 wird gezeigt, dass die Regel sehr schlecht ist, weil der Erwartungswert nicht erfüllt wird.