# Exercise 2

## a) Compute Mutual Information MI

The required probabilities can be easily derived from the given table:

$$\begin{aligned} P[T_k, C_i] &= \frac{50}{2'000} = 0.025, \\ P[T_k] &= \frac{130}{2'000} = 0.065, \\ P[C_i] &= \frac{950}{2'000} = 0.475. \end{aligned}$$

$$\begin{aligned} \mathrm{MI}(T_k, C_i) &= \log_2 \left( \frac{P[T_k, C_i]}{P[T_k]P[C_i]} \right) \\ &= \log_2 \left( \frac{0.025}{0.065 \cdot 0.475} \right) \\ &\approx -0.304511, \end{aligned}$$

## b) Compute Odds Ratio OR

Additionally to the ones from before, we need some more probabilities, namely the following conditional ones:

$$\begin{aligned} P[T_k|C_i] &= \frac{50}{950} \approx 0.05263, \\ P[T_k|\bar{C}_i] &= \frac{80}{1'050} \approx 0.07619 \end{aligned}$$

$$\begin{aligned} \mathrm{OR}(T_k, C_i) &= \frac{P[T_k|C_i]}{1 - P[T_k|C_i]} \cdot \frac{1 - P[T_k|\bar{C}_i]}{P[T_k|\bar{C}_i]} \\ &= \frac{0.05263}{0.94737} \cdot \frac{0.92381}{0.07619} \\ &\approx 0.6736 \end{aligned}$$

## c) Compute $\chi^2$

$$\begin{aligned} \chi^2(T_k, C_i) &= \overbrace{|T_r|}^{|\text{training set}|} \cdot \frac{\left( P[T_k, C_i] \cdot P[\bar{T}_k, \bar{C}_i] - P[T_k, \bar{C}_i] \cdot P[\bar{T}_k, C_i] \right)^2}{P[T_k] \cdot P[\bar{T}_k] \cdot P[C_i] \cdot P[\bar{C}_i]} \\ &= 2'000 \cdot \frac{\left( 0.025 \cdot \frac{970}{2'000} - \frac{80}{2'000} \cdot \frac{900}{2'000} \right)^2}{0.065 \cdot 0.935 \cdot 0.475 \cdot 0.525} \\ &\approx 4.5548 \end{aligned}$$

## d) Compute Information Gain IG

$$
\begin{aligned}
\mathrm{IG}(T_k, C_i) = \quad & P[T_k, C_i] \cdot \log_2\left(\frac{P[T_k, C_i]}{P[T_k]P[C_i]}\right) + P[\bar{T}_k, C_i] \cdot \log_2\left(\frac{P[\bar{T}_k, C_i]}{P[\bar{T}_k]P[C_i]}\right) \\
& + P[T_k, \bar{C}_i] \cdot \log_2\left(\frac{P[T_k, \bar{C}_i]}{P[T_k]P[\bar{C}_i]}\right) + P[\bar{T}_k, \bar{C}_i] \cdot \log_2\left(\frac{P[\bar{T}_k, \bar{C}_i]}{P[\bar{T}_k]P[\bar{C}_i]}\right) \\
= \quad & 0.025 \cdot (-0.304511) + \frac{900}{2000} \cdot \log_2\left(\frac{\frac{900}{2000}}{0.935 \cdot 0.475}\right) \\
& + \frac{80}{2000} \cdot \log_2\left(\frac{\frac{80}{2000}}{0.065 \cdot 0.525}\right) + \frac{970}{2000} \cdot \log_2\left(\frac{\frac{970}{2000}}{0.935 \cdot 0.525}\right) \\
\approx \quad & 0.00166
\end{aligned}
$$