

Przykład 2

Zadanie: Mamy następujący zbiór treningowy złożony z atrybutów określających warunki pogodowe i decyzji o wyjściu na spacer. Mamy sklasyfikować przypadek $X = (\text{deszczowo}, 21^{\circ}\text{C}, \text{słaby})$.

id	pogoda	temperatura	wiatr	decyzja
1	słonecznie	23°C	umiarkowany	tak
2	deszczowo	15°C	mocny	nie
3	pochmurno	17°C	słaby	tak
4	pochmurno	21°C	umiarkowany	nie
5	słonecznie	20°C	mocny	tak
6	słonecznie	25°C	słaby	tak
7	deszczowo	22°C	słaby	tak
8	słonecznie	14°C	mocny	nie
9	pochmurno	19°C	mocny	nie
10	deszczowo	16°C	słaby	nie

Przykład 2 – Rozwiązanie

Najpierw potrzebujemy dokonać pewnego podziału atrybutu drugiego (temperatury) w zależności o jego wartości. Atrybut przyjmuje za dużo różnych wartości, dlatego dokonamy umownego podziału na 3 grupy w taki sposób, że

- do grupy *chłодно* zaliczymy temperatury poniżej 16°C ,
- do grupy *ciepło* zaliczymy temperatury pomiędzy 16°C a 20°C ,
- do grupy *gorąco* zaliczymy temperatury powyżej 20°C .

Przykład 2 – Rozwiązanie

Najpierw potrzebujemy dokonać pewnego podziału atrybutu drugiego (temperatury) w zależności o jego wartości. Atrybut przyjmuje za dużo różnych wartości, dlatego dokonamy umownego podziału na 3 grupy w taki sposób, że

- do grupy *chłодно* zaliczymy temperatury poniżej 16°C ,
- do grupy *ciepło* zaliczymy temperatury pomiędzy 16°C a 20°C ,
- do grupy *gorąco* zaliczymy temperatury powyżej 20°C .

Dodatkowo niech C_1 oznacza decyzję pozytywną, a C_2 – negatywną.

Przykład 2 – Rozwiązanie

Po tej modyfikacji zbiór treningowy przedstawia się następująco

Przykład 2 – Rozwiązanie

Po tej modyfikacji zbiór treningowy przedstawia się następująco

id	pogoda	temperatura	wiatr	decyzja
1	słonecznie	gorąco	umiarkowany	tak
2	deszczowo	chłodno	mocny	nie
3	pochmurno	ciepło	słaby	tak
4	pochmurno	gorąco	umiarkowany	nie
5	słonecznie	ciepło	mocny	tak
6	słonecznie	gorąco	słaby	tak
7	deszczowo	gorąco	słaby	tak
8	słonecznie	chłodno	mocny	nie
9	pochmurno	ciepło	mocny	nie
10	deszczowo	ciepło	słaby	nie

Przykład 2 – Rozwiązanie

Po tej modyfikacji zbiór treningowy przedstawia się następująco

id	pogoda	temperatura	wiatr	decyzja
1	słonecznie	gorąco	umiarkowany	tak
2	deszczowo	chłodno	mocny	nie
3	pochmurno	ciepło	słaby	tak
4	pochmurno	gorąco	umiarkowany	nie
5	słonecznie	ciepło	mocny	tak
6	słonecznie	gorąco	słaby	tak
7	deszczowo	gorąco	słaby	tak
8	słonecznie	chłodno	mocny	nie
9	pochmurno	ciepło	mocny	nie
10	deszczowo	ciepło	słaby	nie

Natomiast obiekt do sklasyfikowania jest teraz postaci $X = (\text{deszczowo}, \text{gorąco}, \text{słaby})$.

Przykład 2 – Rozwiązanie

Musimy ustalić, które z poniższych prawdopodobieństw jest większe

- $P(C = C_1|X)$ czyli
 $P(\text{decyzja=tak}|\text{pogoda=deszczowo, temperatura=gorąco, wiatr=słaby})$,
- $P(C = C_2|X)$ czyli
 $P(\text{decyzja=nie}|\text{pogoda=deszczowo, temperatura=gorąco, wiatr=słaby})$.

Przykład 2 – Rozwiązanie

Musimy ustalić, które z poniższych prawdopodobieństw jest większe

- $P(C = C_1|X)$ czyli
 $P(\text{decyzja=tak}|\text{pogoda=deszczowo, temperatura=gorąco, wiatr=słaby})$,
- $P(C = C_2|X)$ czyli
 $P(\text{decyzja=nie}|\text{pogoda=deszczowo, temperatura=gorąco, wiatr=słaby})$.

Więc wystarczy, że obliczymy $P(C = C_1) \cdot P(C = C_1|X)$ oraz $P(C = C_2) \cdot P(C = C_2|X)$ i porównamy te wyniki.

Przykład 2 – Rozwiązanie

Dla ułatwienia dalszych obliczeń przypadki, w których końcowa decyzja jest pozytywna zaznaczymy na zielono, a tam gdzie negatywna – na czerwono.

id	pogoda	temperatura	wiatr	decyzja
1	słonecznie	gorąco	umiarkowany	tak
2	deszczowo	chłodno	mocny	nie
3	pochmurno	ciepło	słaby	tak
4	pochmurno	gorąco	umiarkowany	nie
5	słonecznie	ciepło	mocny	tak
6	słonecznie	gorąco	słaby	tak
7	deszczowo	gorąco	słaby	tak
8	słonecznie	chłodno	mocny	nie
9	pochmurno	ciepło	mocny	nie
10	deszczowo	ciepło	słaby	nie

Przykład 2 - Rozwiązanie

Obliczamy składowe szukanych prawdopodobieństw

Przykład 2 - Rozwiązanie

Obliczamy składowe szukanych prawdopodobieństw

$$\textcircled{1} P(C = C_1) = \frac{5}{10} = \frac{1}{2},$$

Przykład 2 - Rozwiązanie

Obliczamy składowe szukanych prawdopodobieństw

$$\textcircled{1} P(C = C_1) = \frac{5}{10} = \frac{1}{2},$$

$$\textcircled{2} P(a_1 = \text{deszczowo} | C = C_1) = \frac{1}{5},$$

Przykład 2 - Rozwiązanie

Obliczamy składowe szukanych prawdopodobieństw

- 1 $P(C = C_1) = \frac{5}{10} = \frac{1}{2},$
- 2 $P(a_1 = \text{deszczowo} | C = C_1) = \frac{1}{5},$
- 3 $P(a_2 = \text{gorąco} | C = C_1) = \frac{3}{5},$

Przykład 2 - Rozwiązanie

Obliczamy składowe szukanych prawdopodobieństw

- ① $P(C = C_1) = \frac{5}{10} = \frac{1}{2},$
- ② $P(a_1 = \text{deszczowo} | C = C_1) = \frac{1}{5},$
- ③ $P(a_2 = \text{gorąco} | C = C_1) = \frac{3}{5},$
- ④ $P(a_3 = \text{słaby} | C = C_1) = \frac{3}{5},$

Przykład 2 - Rozwiązanie

Obliczamy składowe szukanych prawdopodobieństw

$$\textcircled{1} P(C = C_1) = \frac{5}{10} = \frac{1}{2},$$

$$\textcircled{2} P(a_1 = \text{deszczowo} | C = C_1) = \frac{1}{5},$$

$$\textcircled{3} P(a_2 = \text{gorąco} | C = C_1) = \frac{3}{5},$$

$$\textcircled{4} P(a_3 = \text{słaby} | C = C_1) = \frac{3}{5},$$

$$\textcircled{5} P(C = C_2) = \frac{5}{10} = \frac{1}{2},$$

Przykład 2 - Rozwiązanie

Obliczamy składowe szukanych prawdopodobieństw

$$① P(C = C_1) = \frac{5}{10} = \frac{1}{2},$$

$$② P(a_1 = \text{deszczowo} | C = C_1) = \frac{1}{5},$$

$$③ P(a_2 = \text{gorąco} | C = C_1) = \frac{3}{5},$$

$$④ P(a_3 = \text{słaby} | C = C_1) = \frac{3}{5},$$

$$⑤ P(C = C_2) = \frac{5}{10} = \frac{1}{2},$$

$$⑥ P(a_1 = \text{deszczowo} | C = C_2) = \frac{2}{5},$$

Przykład 2 - Rozwiązanie

Obliczamy składowe szukanych prawdopodobieństw

$$① P(C = C_1) = \frac{5}{10} = \frac{1}{2},$$

$$② P(a_1 = \text{deszczowo} | C = C_1) = \frac{1}{5},$$

$$③ P(a_2 = \text{gorąco} | C = C_1) = \frac{3}{5},$$

$$④ P(a_3 = \text{słaby} | C = C_1) = \frac{3}{5},$$

$$⑤ P(C = C_2) = \frac{5}{10} = \frac{1}{2},$$

$$⑥ P(a_1 = \text{deszczowo} | C = C_2) = \frac{2}{5},$$

$$⑦ P(a_2 = \text{gorąco} | C = C_2) = \frac{1}{5},$$

Przykład 2 - Rozwiązanie

Obliczamy składowe szukanych prawdopodobieństw

$$① P(C = C_1) = \frac{5}{10} = \frac{1}{2},$$

$$② P(a_1 = \text{deszczowo} | C = C_1) = \frac{1}{5},$$

$$③ P(a_2 = \text{gorąco} | C = C_1) = \frac{3}{5},$$

$$④ P(a_3 = \text{słaby} | C = C_1) = \frac{3}{5},$$

$$⑤ P(C = C_2) = \frac{5}{10} = \frac{1}{2},$$

$$⑥ P(a_1 = \text{deszczowo} | C = C_2) = \frac{2}{5},$$

$$⑦ P(a_2 = \text{gorąco} | C = C_2) = \frac{1}{5},$$

$$⑧ P(a_3 = \text{słaby} | C = C_2) = \frac{1}{5},$$

Przykład 2 – Rozwiązanie

Podstawiamy obliczone wartości do wzorów i otrzymujemy

Przykład 2 – Rozwiązanie

Podstawiamy obliczone wartości do wzorów i otrzymujemy

- $$P(C = C_1) \cdot \prod_{j=1}^3 P(a_j = x_j | C = C_1) = \frac{1}{2} \cdot \frac{1}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} = \frac{9}{250}$$

Przykład 2 – Rozwiązanie

Podstawiamy obliczone wartości do wzorów i otrzymujemy

- $P(C = C_1) \cdot \prod_{j=1}^3 P(a_j = x_j | C = C_1) = \frac{1}{2} \cdot \frac{1}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} = \frac{9}{250}$
- $P(C = C_2) \cdot \prod_{j=1}^3 P(a_j = x_j | C = C_2) = \frac{1}{2} \cdot \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{1}{5} = \frac{2}{250}$

Przykład 2 – Rozwiązanie

Podstawiamy obliczone wartości do wzorów i otrzymujemy

- $P(C = C_1) \cdot \prod_{j=1}^3 P(a_j = x_j | C = C_1) = \frac{1}{2} \cdot \frac{1}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} = \frac{9}{250}$
- $P(C = C_2) \cdot \prod_{j=1}^3 P(a_j = x_j | C = C_2) = \frac{1}{2} \cdot \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{1}{5} = \frac{2}{250}$

Zatem $P(C = C_1 | X) > P(C = C_2 | X)$.

Przykład 2 – Rozwiązanie

Podstawiamy obliczone wartości do wzorów i otrzymujemy

- $P(C = C_1) \cdot \prod_{j=1}^3 P(a_j = x_j | C = C_1) = \frac{1}{2} \cdot \frac{1}{5} \cdot \frac{3}{5} \cdot \frac{3}{5} = \frac{9}{250}$
- $P(C = C_2) \cdot \prod_{j=1}^3 P(a_j = x_j | C = C_2) = \frac{1}{2} \cdot \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{1}{5} = \frac{2}{250}$

Zatem $P(C = C_1 | X) > P(C = C_2 | X)$.

Wniosek

Obiekt $X=(\text{deszczowo}, 21^{\circ}\text{C}, \text{słaby})$ powinien zostać przydzielony do klasy C_1 oznaczającej pozytywną decyzję o wyjściu na spacer.