


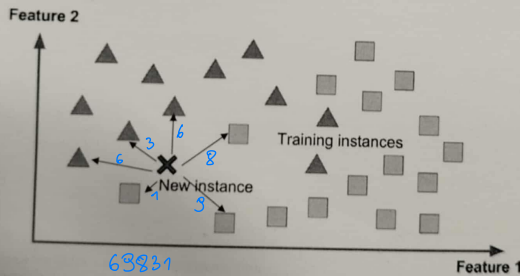



















1. [P_U01, 6pkt] Stosując metodę kNN, wyznacz kategorię nieznanego obiektu dla:

1 sąsiada	3 sąsiadów	5 sąsiadów
		

Przyjmij, jako odległość cyfry z Twojego nr albumu. Wpisz odległości na strzałkach na rysunku. Niezbędne obliczenia i wyjaśnienia powinny znajdować się na kartce.



d	KLASA		plon. większość
1		$k=1$	
3		$k=3$	   = 
6		$k=5$	     = 
6			
8			
9			

LP	Etat	Zgoda przełożonego	Szkolenie BHP	Decyzja
1	Pełny etat	Tak	Tak	Tak
2	YYYYYY <i>Pół</i>	Tak	Tak	Tak
3	Zlecenie	Tak	Nie	Nie
4	Pełny etat	Nie	Tak	Nie
5	Brak	Tak	Tak	Nie
6	Zlecenie	Tak	Tak	Tak

YYYYYY - gdy ostatnia cyfra z numeru indeksu jest: 7

0-5 - to wstawiamy Pół etatu

6-9 - to wstawiamy Zlecenie

2. [P_U02, 6 pkt] Wyznacz najbardziej informacyjny atrybut, który maksymalizuje przyrost informacji, a więc najlepiej redukuje entropię (niepewność) w zbiorze danych po dokonaniu podziału.

ETAT

	PEŁNY	PÓŁ	ZLEC.	BRAK
TAK	1	1	1	0
NIE	1	0	1	1
INFO	1	0	1	0

$$\text{INFO dla etatu: } \frac{2}{6} \cdot 1 + 0 + \frac{2}{6} \cdot 1 + 0 = \frac{4}{6}$$

$$\text{GAIN} = 1 - \frac{4}{6} = \frac{2}{6} \approx 0,33 \dots$$

2 GODA : S2K.BHP — rozłożenie jest takie same!

	TAK	NIE
TAK	3	0
NIE	2	1
INFO	$\sim 0,97$	

$$\text{INFO Tak: } -\frac{3}{5} \cdot \log_2 \frac{3}{5} - \frac{2}{5} \cdot \log_2 \frac{2}{5} \approx 0,970$$

$$\text{INFO NIE: } 0$$

$$\text{INFO dla klasy: } \frac{5}{6} \cdot 0,97 + 0 \approx 0,809$$

$$\text{GAIN} = 1 - 0,809 = 0,191$$

ETAT ma największy GAIN,
ten attribute jest najbardziej informacyjny.

③ jest w temacie