

Program 6

Wykonaj regresję logistyczną dla danych z załączonego w zadaniu pliku. Zrób to w następujący sposób:

1. Zaimportuj potrzebne biblioteki.
2. Wczytaj dane.
3. Usuń w zmiennej *Churn?* kropki po *True* i *False*.
4. Utwórz nową ramkę danych zawierającą tylko zmienne *Churn?* oraz *CustServ Calls*.
5. Utwórz nową zmienną *CSC* w następujący sposób:
 - 0 lub 1 rozmowa z biurem obsługi klienta \iff *CSC* przyjmuje wartość *małe*
 - 2 lub 3 rozmowy z biurem obsługi klienta \iff *CSC* przyjmuje wartość *średnie*
 - 4 lub więcej rozmów z biurem obsługi klienta \iff *CSC* przyjmuje wartość *wysokie*
6. Utwórz tabelę krzyżową zmiennej *Churn?* względem *CSC*.
7. Utwórz dwie nowe zmienne wskaźnikowe *CSC_sr* i *CSC_wys*, gdzie dla małej liczby rozmów *CSC_sr* = 0, *CSC_wys* = 0, dla średniej *CSC_sr* = 1, *CSC_wys* = 0, a dla wysokiej *CSC_sr* = 0, *CSC_wys* = 1.
8. Używając *CSC* o małej wartości jako kategorii odniesienia, oblicz ilorazy szans *OR_nis*, *OR_sr* i *OR_wys* dla *CSC* o małej wartości, *CSC* o średniej wartości i *CSC* o wysokiej wartości odpowiednio.
9. Oblicz b_0, b_1, b_2 , gdzie $b_0 = \ln(\text{OR_nis})$, $b_1 = \ln(\text{OR_sr})$, $b_2 = \ln(\text{OR_wys})$.
10. Zapisz funkcję logitową $g(X) = b_0 + b_1x + b_2y$, gdzie $X = [x, y]$ (zostawiając x y w postaci symboli).
11. Zapisz prawdopodobieństwo przeniesienia się $\pi(X) = \frac{e^{g(X)}}{1+e^{g(X)}}$.
12. Oblicz prawdopodobieństwa przeniesienia się w zależności od tego, czy klient odbył małe, średnio czy dużo rozmów z biurem obsługi klienta.
13. Policz dodatkowo prawdopodobieństwo przeniesienia się bez podziału na liczbę rozmów.
14. Napisz wnioski.

Opisane zadanie z wynikami znajdziesz w podrozdziale 4.5.2. książki D. Larose'a [*Metody i modele eksploracji danych*](#).