



Eksploracja danych

# Skuteczność leków na podstawie opinii

Grzegorz Łobza nr indeksu 22441

Hubert Łyszczarz nr indeksu 22861

Kamil Gruszczyński nr indeksu 22677

Program ma na celu wskazanie skuteczności leków na daną chorobę na podstawie wystawionych opinii

W celu analizy, postanowiliśmy sprowadzić komentarze do prostszej postaci.

Pierwszym krokiem jest wyczyszczenie komentarzy ze zbędnych znaków specjalnych. Następnie, program analizuje tekst i na podstawie wcześniej utworzonych słowników wyrazów pozytywnych i negatywnych, podejmuje zero-jedynkowe decyzje, czy dany komentarz jest pozytywny czy negatywny, zależnie od częstotliwości występowania wyrazów plusowych i minusowych.

W przypadku remisu, komentarz pozostaje neutralny.

Z założenia rozwiązania słownikowe, są na tyle skuteczne na ile obszerny jest zasób słów i wyrazów w słowniku, dlatego byliśmy zmuszeni wprowadzić także 4 typ komentarza – false.

Komentarze z tym typem charakteryzują się zerowym występowaniem wyrazów pozytywnych i negatywnych, przez co postanowiliśmy nie brać ich w ogóle pod uwagę.

Logika w skrócie:

-1 – wynik negatywny ( bad )

0 – wynik neutralny ( neutral )

1 – wynik pozytywny ( good )

False – wynik nie podlegający analizie ( usuwany )

Na tym etapie posiadamy nową kolumnę z oceną leków według naszego systemu oceniania.

Następnie dane zostają przygotowane pod utworzenie drzewa klasyfikacji i późniejszego wykorzystania w uczeniu maszynowym.

W skrócie, nazwy leków oraz chorób zostają zamienione w uniwersalne indeksy i umieszczane w liście – będą to nasze „features”.

Przykład: [ 13 , 23 ] gdzie 13 to lek, a 23 choroba

Wcześniej wygenerowane opinie na podstawie komentarzy to nasze „labels”, które posłużą do nauki algorytmu.

Na koniec, tworzymy diagram skuteczności leków na choroby, wykorzystując już wcześniej utworzone drzewo decyzji i testowy plik ( 65787 rekordów ).

Do nauki na wykorzystaliśmy 196431 rekordów.

Diagram posiada legendę, dzięki której możemy zweryfikować jak leki oceniał nasz algorytm, a jak algorytm Machine Learning.