



## Zadanie testowe - NLP

### Opis zadania

Celem zadania jest przedstawienie koncepcji na rozwiązanie problemu oceny sentymentu konwersacji w języku angielskim, a także zaprezentowanie procesu analizy i czyszczenia danych.

Do zadania dołączone zostają dwa pliki .csv. Oba pliki posiadają przecinek “,” jako znak podziału komórek, a dane które są ciągami znaków są dodatkowo otoczone cudzysłowem.

ratings.csv

Opis struktury pliku:

- Id odwiedzającego
- ocena (1 pozytywna, 0 negatywna)

conversations.csv

Opis struktury pliku:

- Id konwersacji (liczba)
- Id projektu (liczba)
- Id odwiedzającego (ciąg znaków)
- Kanał (ciąg znaków)
- Nadawca (ciąg znaków)
- Id operatora (liczba - w przypadku wiadomości gdzie nadawca jest operator lub 0)
- Id typu wiadomości
- treść

Kluczem wspólnym obu plików jest “id odwiedzającego” co pozwala na dopasowanie sentymentu rozmowy do jej treści.

### Kryteria oceny

- Analiza danych
  - Sposób analizy danych i zwrócenie uwagi na ewentualne problemy z przedstawionym zbiorem danych
  - Określenie czy jest możliwość wykorzystania tych danych do treningu modelu

- Czyszczenie danych
  - Jeśli analiza wykaże jakieś problemy, ocenie będzie podlegać propozycja automatycznego procesu oczyszczania danych
- Propozycja rozwiązania
  - Konceptyjny opis proponowanego rozwiązania problemu automatycznego oceniania sentymentu rozmów w języku angielskim
  - Zwrócenie uwagi na mocne i słabe strony zaproponowanego rozwiązania

Dodatkowym atutem będzie dostarczenie kodu (w formie repozytorium na github) do przeprowadzania etapu analizy i czyszczenia danych.

UWAGA: Nie wymagamy przedstawienia wytrenowanego modelu