

Zajęcia 6. Dokumentacja i specyfikacja wymagań

a) Notatki z zajęć (różnice między plikami 6.1, 6.2, 6.3 i 6.4):

Każdy z kodów ma różny cel i w efekcie daje różny rezultat. 6.1 daje najprostszą analizę tekstu - usuwa po prostu przerywniki i bada częstotliwość słów, następnie tworząc prostą chmurę słów. 6.2 podobnie - bierze jako podstawę tekst, wykreśla z niego przerywniki, ale stosuje stemming (tzn. usuwa końcówki wyrazów i sprowadza je do podstawowego członu) i tworzy po prostu listę najpopularniejszych słów i ich występowanie. 6.3 to połączenie 6.1 i 6.2 - tzn. funkcja stemminguje dany tekst, a następnie tworzy chmurę słów z uzyskanego wyniku. Z kolei 6.4 bada co innego - w pierwszej części oczyszczony jak zwykle tekst jest poddany badaniu analizy sentymentu, tzn. jak dużo słów ma skojarzenie negatywne, neutralne i pozytywne. Jak widać po wyniku analiz, jest to mocno zależne od przyjętego słownika (wzięto ich 4) - bowiem słowa i frazy pozytywne dla biznesu mogą być przez publikę odbierane negatywnie (i vice versa). Ostatnią częścią kodu jest z kolei analiza sentymentu "w czasie mówienia".

Generalnie łączy je analizowanie tekstu, dobry opis działania (kody są udokumentowane, zmienne nazwane i skomentowane), ich efekty łączy również to, że przybliżają nas do lepszego zrozumienia tekstu, nawet bez jego czytania, ale różni ich podejście do dokładności (stemming lub jego brak), a także generalna ocena całego przemówienia - analiza sentymentu jest tylko w 6.4.

b) Na podstawie dokumentacji SRS, która na końcu zawiera scenariusze użytkownika (user stories), napisz w Notatkach:

Scenariusz 4. Wczesne wykrywanie ryzyk reputacyjnych na podstawie analizy sentymentu opinii klientów z formularzy kontaktowych

- 1) **Jako:** Analityk danych w dziale Customer Experience
- 2) **Chcę:** Analizować na żywo opinie klientów
- 3) **Aby:** Wykrywać źródła negatywnego sentymentu i znajdować kluczowe słowa użyte w dużej liczbie negatywnych recenzji
- 4) **Kryteria akceptacji:**
 - (a) Użytkownik może wczytywać dane tekstowe z formularzy kontaktowych do pliku tekstowego
 - (b) Użytkownik może następnie analizować dane poprzez analizę sentymentu
 - (c) Program zwraca kluczowe słowa pojawiające się w negatywnych recenzjach wraz z ich częstotliwością i tym, czy pojawiały się razem (analiza sentymentu w czasie).