

1. Wprowadzenie

Celem projektu jest analiza sentymentu wypowiedzi wybranych polityków w okresach, w których byli oni w opozycji oraz w rządzie. Zostały przyjęte dwa okresy badawcze: kadencja Sejmu w latach 2019-2023 oraz obecna kadencja Sejmu (tj. 2023-teraz). Analiza została przeprowadzona na 3 politykach: Radosławie Foglu (PiS), Wandzie Nowickiej (Lewica) oraz Donaldzie Tusku (KO). Chcemy zbadać, czy i ewentualnie jak bardzo zmienia się sentyment wypowiedzi polityków w zależności od tego, czy obecnie rządzą lub są w opozycji. W skrypcie R dokonujemy, na podstawie wczytanych z folderu dokumentów txt. zawierających wypowiedzi badanych polityków z obu okresów, analizy ich sentymentu, przy użyciu dwóch polskich słowników (odpowiednio dla słów pozytywnych jak i negatywnych). Ponadto generujemy dla każdego polityka 2 wizualizacje otrzymanych wyników (histogram rozkładu sentymentu oraz boxplot porównujący oba okresy).

2. Cele systemu

Celem systemu jest umożliwienie analizy postaw oraz języka wypowiedzi polityków w zależności od ich aktualnej roli – czy sprawują władzę (rząd), czy znajdują się w opozycji. System wspiera badanie zmian w sentymencie wypowiedzi w różnych kontekstach politycznych. Może być także wykorzystywany w analizach politologicznych umożliwiając analizę spójności przekazów poszczególnych polityków.

Celem systemu jest:

- wczytanie plików txt. zawierających wypowiedzi polityków z wybranego folderu.
- przetwarzanie i oczyszczanie tekstu (tokenizacja oraz lematyzacja przy pomocy pakietu UDPipe).
- Oznaczenie okresów (rząd/opozycja) wypowiedzi.
- Przeprowadzenie analizy sentymentu przy pomocy narzędzi języka R.
- Wizualizacja wyników w sposób umożliwiający porównanie sentymentu między politykami oraz w zależności od okresu (rząd vs opozycja) – stworzenie histogramów oraz wykresów typu boxplot.
- Stworzenie raportu podsumowującego wyniki analizy.

3. Wymagania funkcjonalne

1. System umożliwia wczytanie danych wejściowych zawierających wypowiedzi polityków.

2. System klasyfikuje wypowiedzi w zależności od okresu (rząd/opozycja) na podstawie konkretnych plików tekstowych.
3. System analizuje sentyment każdej wypowiedzi przy użyciu pakietów tidytext, udpipe i podobnych.
4. System umożliwia wczytanie polskich słowników dla słów pozytywnych i słów negatywnych.
5. System wizualizuje wyniki analizy sentymentu (histogramy i wykresy typu boxplot).
6. System generuje raport z analizy w formie przejrzystego i czytelnego dokumentu HTML.

4. Wymagania niefunkcjonalne

1. Kod powinien być napisany w języku R w sposób modularny i czytelny (z podziałem na funkcje i komentarze).
2. Program powinien działać na systemach Windows, macOS i Linux.
3. Średni czas analizy nie powinien przekraczać dla 3 polityków 3 minut.
4. Wymagania bibliotek: tidyverse, tidytext, udpipe, stopwords, stringr, tm.
5. Interfejs (jeśli obecny) powinien być responsywny i przyjazny użytkownikowi.

5. Interfejsy użytkownika i wymagania dotyczące danych

- **Wejście:** Pliki tekstowe .txt zawierające wypowiedzi badanych polityków, dwa pliki słowników sentymentów w języku polskim w formacie txt, model UDPipe dla języka polskiego.
- **Wyjście:** Wykresy ukazujące sentyment wypowiedzi w podziale na okresy (histogramy i wykresy typu boxplot).

Wymagania dotyczące danych

- Kodowanie UTF-8, które umożliwia obsługiwanie polskich znaków tekstowych.
- Dane tekstowe w postaci txt.

6. Słownictwo dokumentacji

- **Sentyment:** Wartość opisująca emocjonalny ton wypowiedzi (np. pozytywny, negatywny, neutralny).
- **Korpus:** Zbiór wypowiedzi tekstowych potrzebnych do analizy.

- **Okres rządzenia/opozycji:** Zależnie od przynależności partii politycznej danego polityka do obozu władzy.
- **Tokenizacja:** Proces dzielenia tekstu na mniejsze elementy (np. słowa).
- **Lematyzacja:** Proces, który grupuje różne formy jednego słowa i przypisuje im to samo znaczenie.

7. Przypadki użycia (use cases)

UC1: Wczytanie danych

Aktor: Użytkownik

Opis: Użytkownik wczytuje pliki txt. z wypowiedziami polityków.

Warunek początkowy: Program uruchomiony

Rezultat końcowy: Dane wczytane i gotowe do analizy.

UC2: Analiza sentymentu

Aktor: System

Opis: System wykonuje analizę sentymentu dla wypowiedzi danego polityka w wybranym okresie.

Warunek początkowy: Dane wczytane

Rezultat końcowy: Wyniki analizy dostępne w postaci wykresów i raportów.

8. Scenariusze użytkownika (user stories)

Story 1:

Jako analityk danych chcę porównać sentyment wypowiedzi Tuska, Fogla i Nowickiej w rządzie i opozycji, aby zbadać zmianę tonu w zależności od pozycji politycznej.

Story 2:

Jako użytkownik chcę wczytać dane z pliku txt., aby nie musieć ręcznie kopiować wypowiedzi.

Story 3:

Jako badacz języka chcę wizualnie zobaczyć rozkład emocji w wypowiedziach, aby łatwiej interpretować ich charakter.