

Jak wiatr zawieje, czyli wykrywanie ludzkich chorągiewek



Tło

Zmienność zdania wśród ludzi jest zjawiskiem naturalnym i powszechnym. Zmienność ta jest szczególnie istotna, gdy dotyczy osób publicznych w mediach społecznościowych – chcemy wiedzieć, czy osoba zrewidowała swoje nastawienie do pewnej sprawy, by móc w rezultacie spytać ją o przyczynę, bądź ocenić, czy jest wiarygodna. Automatyzacja wykrywania zmian w postawie ułatwi to zadanie.

Gromadzenie danych

Tweety zostały zescrapowane w języku polskim. Spośród 300 tys. tweetów wyselekcjonowano 7 tys. dla 92 unikalnych kont z lat 2018-2023. Oczyszczone dane korpusowe z Sejmu i Senatu zawierały 67 tys. wypowiedzi dla 3 tys. unikalnych polityków z lat 2007-2023. Czyszczenie obejmowało m.in. usunięcie hashtagów, zawężenie przedziału czasowego czy usunięcie tekstów charakterystycznych dla początku / końca obrad.



Hipoteza

Na podstawie wypowiedzi parlamentarnych oraz postów na Twitterze jesteśmy w stanie stwierdzić czy dany polityk zmienił swoje zdanie.

Przetwarzanie danych

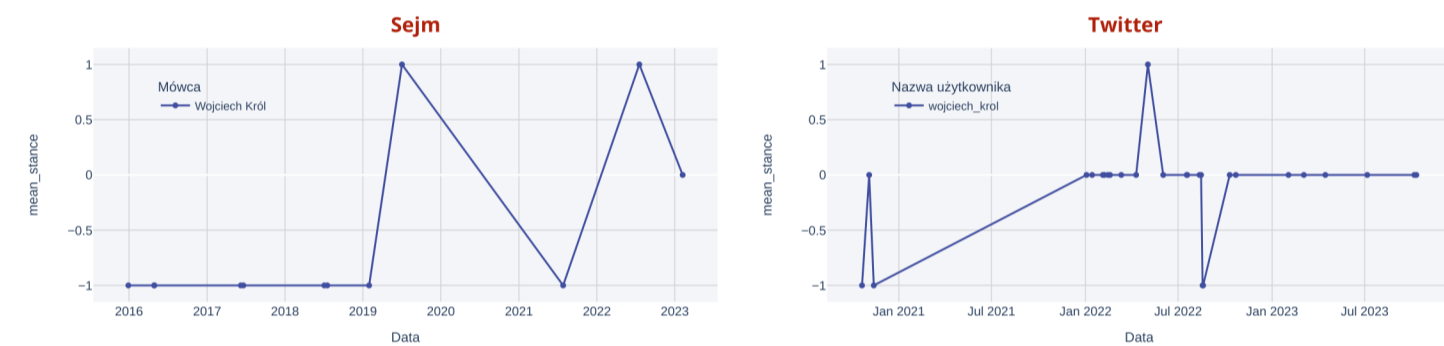
Wyselekcjonowane dane poddano klasteryzacji tematycznej z wykorzystaniem BERTopic, a wynikowe klastry ręcznie dostosowano (ustalono deksyptywne nazwy na bazie najczęstszych słów, złączono podobne tematy, usunięto teksty, dla których nie sposób było jednoznacznie określić tematu). W następnym kroku określono postawę autora wobec poruszanego tematu za pomocą odpowiedniego modelu typu stance detection dla języka polskiego. Model był wytrenowany na zbiorze ręcznie zaanotowanych przez 5 anotorów próbek z zasobów CLARIN-PL (2500) oraz próbkę losowej tweetów zaanotowanej przez gpt-3.5-turbo (1900). Zbiorem testowym dla modelu był Varied Stance Topics przetłumaczony automatycznie na język polski, dla którego f1 macro wynosiło 0.683.



Analiza

Analizowano 57 tematów w Sejmie i Senacie oraz 42 na Twitterze. Za pojedynczą zmianę zdania przyjmowano sytuację, w której dana osoba zmieniła swoją postawę z negatywnej "-1" na pozytywną "1" lub odwrotnie (bez uwzględnienia neutralnych stanów pośrednich "0").

Nastawienie do tematu w czasie 'media publiczne'



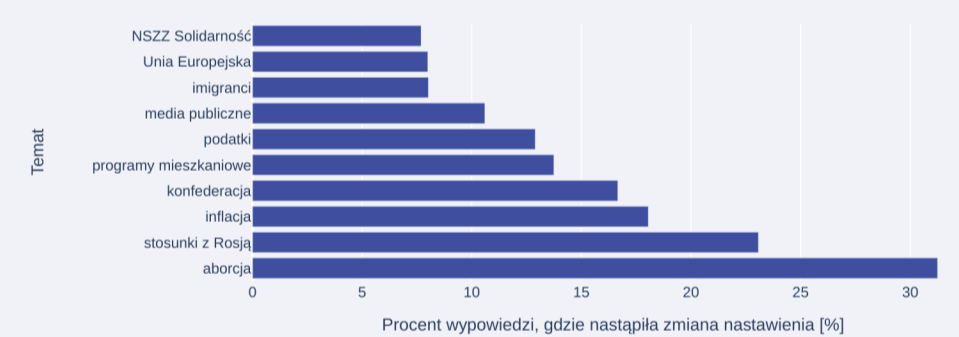
Najczęściej obserwowane tematy w parlamencie: opieka zdrowotna (16.37% wypowiedzi) i wsparcie rolników (15.56% wypowiedzi), na Twitterze: sądownictwo (11.20% wypowiedzi) oraz imigranci (10.78% wypowiedzi).

Sejm

Twitter

VS.

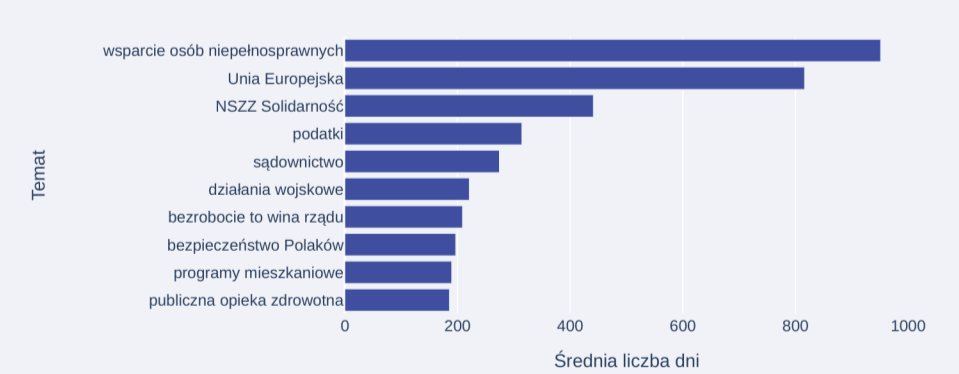
W których tematach najczęściej zmieniano zdanie?



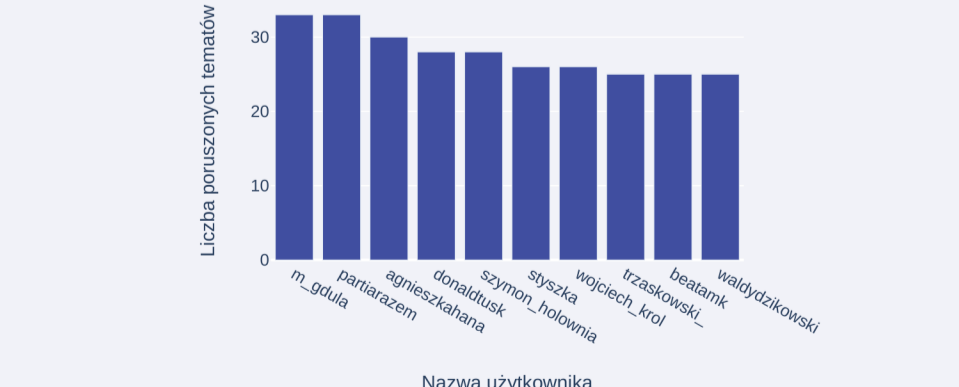
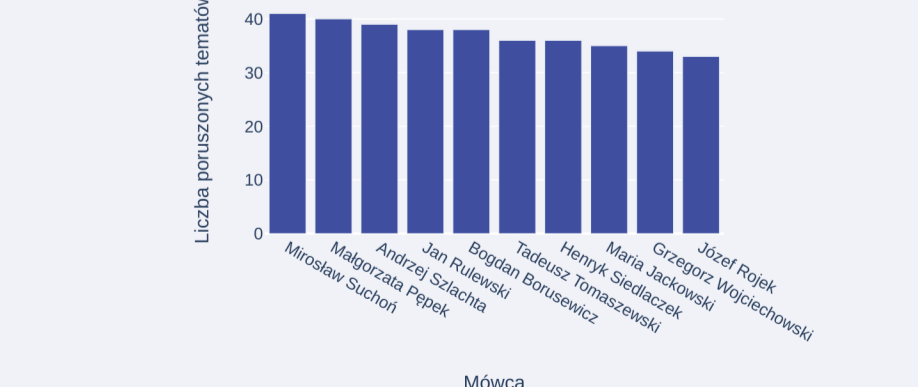
Kto najczęściej zmienia swoje poglądy?



Ile czasu średnio trzeba na zmianę poglądu?



Kto porusza najwięcej tematów?



Wnioski i uwagi

Aktywność na Twitterze oraz w Sejmie i Senacie różnią się od siebie. Dodatkowo analizując próbki tekstów z obu platform, zaobserwowano, że często mówca odnosi się do opinii publicznej, co nie jest w rzeczywistości jego personalną opinią, a model traktuje to jako postawa autora wypowiedzi, co w efekcie w sposób istotny mogło wpłynąć na prezentowane rezultaty. Wyniki modelu typu stance detection nie były korygowane ręcznie.