POLARYZUJĄCA POLITYKA

czyli temat, który sprawia, że **hejtujemy** ponad **2 razy częściej**

Przemysław Mikluszka | Jan Kulbiński

Politechnika Wrocławska

CEL

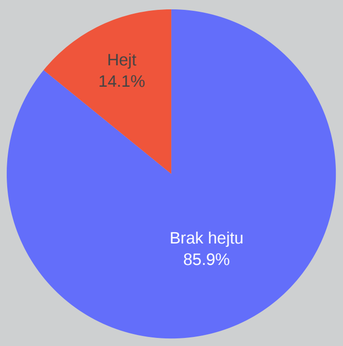
Hejt to szerzenie mowy nienawiści oraz treści obraźliwych i atakujących. Naszym celem było stworzenie automatycznej metody do jego rozpoznawania, a następnie zbadanie jego występowania w postach pod profilami polskich polityków na Twitterze.

ZBIÓR DANYCH

Przygotowaliśmy listę **292 kont** na Twitterze należących do aktywnych polskich polityków. Dla każdego z nich zebraliśmy posty wraz z komentarzami innych użytkowników. W ten sposób uzyskaliśmy ponad **300,000 unikalnych postów i komentarzy**. Dodatkowo zebraliśmy 10,000 postów i komentarzy dotyczących innych tematów.

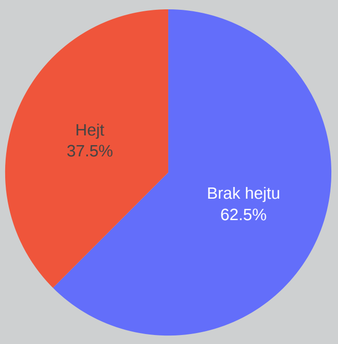
ANALIZA ZBIORU

POSTY O RÓŻNEJ TEMATYCE



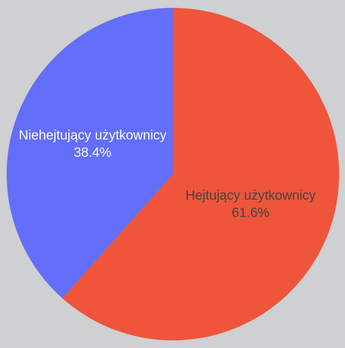
Zbadaliśmy różnice w częstości występowania hejtu w postach i komentarzach dotyczących różnych, popularnych tematów oraz tych pojawiających się pod postami polityków. W postach o tematyce politycznej hejt pojawia się **ponad 2.5 razy częściej** niż w postach, dotyczących innych tematów.

POSTY POLITYCZNE



Policzyliśmy liczbę tekstów ofensywnych napisanych przez każdego użytkownika. Okazało się, że **ponad 60 % użytkowników co najmniej raz użyło mowy nienawiści** pisząc post lub komentarz.

CZY WSZYSCY HEJTUJĄ?



MODEL

Do budowy modelu pozwalającego na analizę ofensywności treści wykorzystaliśmy pretrenowany model semantyki dystrybucyjnej **FastText** oraz samodzielnie trenowany klasyfikator oparty na architekturze **BiLSTM**.

ZBIÓR TRENINGOWY

Do wytrenowania modelu wykorzystaliśmy zbiór dla analogicznego zadania z konkursu **PolEval 2019**. Zawiera on ponad **11,000 zaanotowanych postów**, przy czym tylko **8.92 %** z nich jest ofensywnych.

OCENA JAKOŚCI MODELU

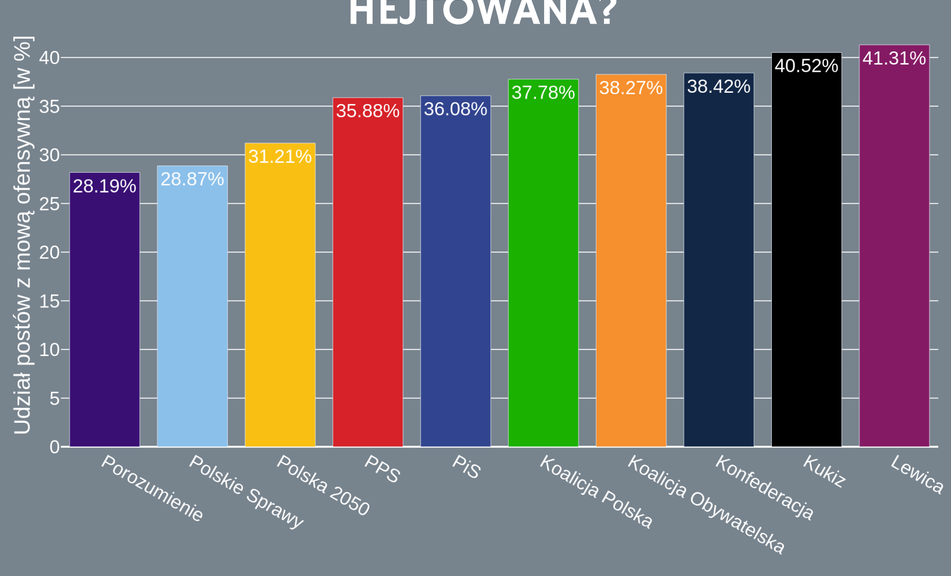
Do oceny jakości utworzonego modelu wykorzystano zbiór testowy wchodzący w skład zbioru treningowego. Ostatecznie udało nam się osiągnąć **89.10 % dokładności** oraz **60.6 % F1**.

UŻYTE NARZĘDZIA

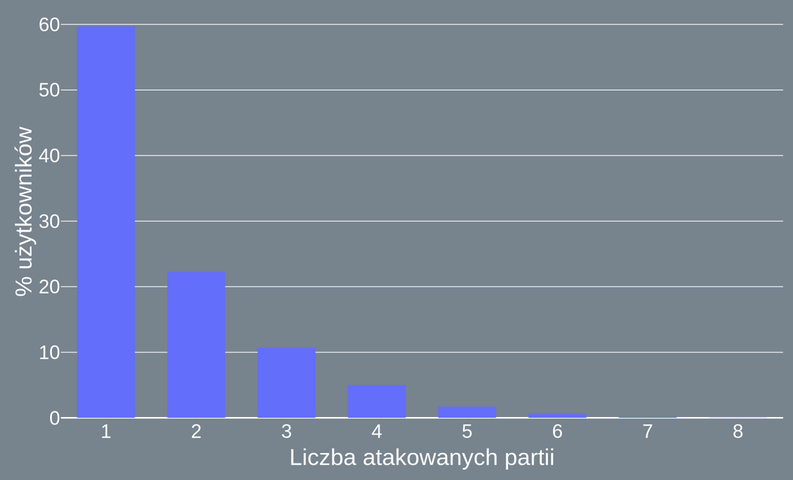
fastText · python · pandas · plotly · PyTorch

NAJWAŻNIEJSZE OBSERWACJE

KTÓRA PARTIA JEST NAJBARDZIEJ HEJTOWANA?



CZY HEJTERZY ATAKUJĄ TYLKO JEDEN CEL?



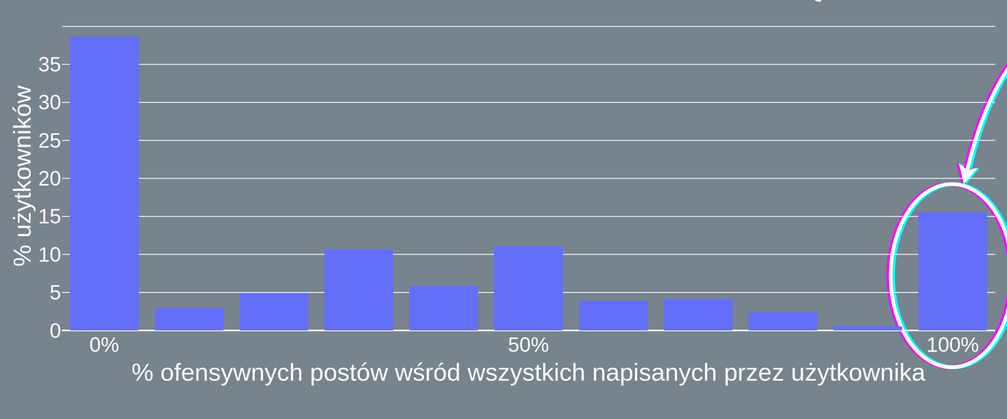
Zdecydowana większość atakuje od **1 do 3 partii** politycznych.
Ci którzy atakują więcej stanowią marginalną część.

Dodatkowo wśród użytkowników:

1/3 nigdy nie napisała hejtu

1/4 częściej pisze hejt niż cokolwiek innego

CZY HEJTERZY TYLKO HEJTUJĄ?



BOTY?

Znaleźliśmy ponad **6,000** użytkowników, którzy publikują wyłącznie treści będące mową nienawiści.
Stanowią oni około **15 %** wszystkich badanych.
Czy są to prawdziwi ludzie, czy może boty?