

PRAWDA CZY SPISEK

Czy można rozpoznać teorie spiskowe na Twitterze ?



Weronika Jakubowska,
Magdalena Kociotek,
Dominik Kurowski

Opiekun grupy: dr hab. inż. Tomasz Kajdanowicz



ILLUMINACI



SZCZEPHENIA



CZARNE
HELIKOPTERY



WTC 9/11



NOWY PORZĄDEK
ŚWIATA



PROGRAMOWANIE
PREDYKCYJNE



CHEMTRAILS



REPTILIANIE



ZAMACH NA
KENNEDIEGO

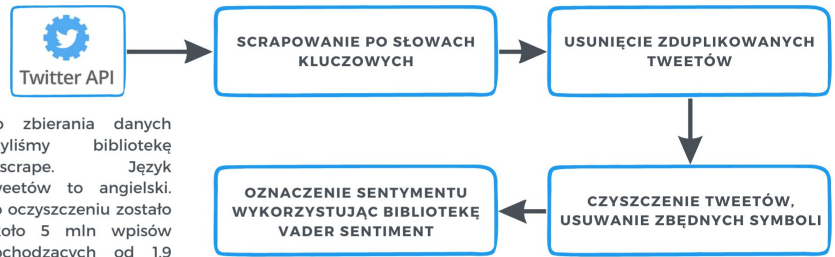
1. OPIS PROBLEMU

Teorie spiskowe są popularne od wielu lat. Ostatnio w trakcie pandemii koronawirusa byliśmy świadkami powstania kolejnych opinii i szerzenia dezinformacji. Każdy z nas bardzo często słyszy, że w szpitalach leżą statyści, a COVID-19 został wymyślony przez Billa Gatesa. Nadmiar takich informacji, powoduje problem z odróżnieniem prawdy od teorii spiskowych.

W ramach pracy postawiliśmy następującą hipotezę:

Teorie spiskowe na Twitterze są charakterystyczne pod względem treści i odbioru.

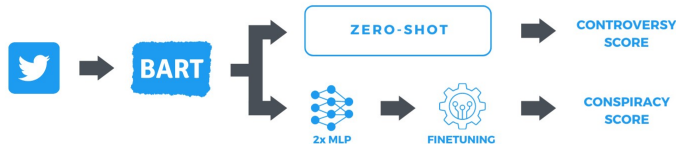
2. ZBIERANIE DANYCH



Do zbierania danych użyliśmy bibliotekę snsrape. Język tweetów to angielski. Po oczyszczeniu zostało około 5 mln wpisów pochodzących od 1,9 mln użytkowników. (po równo dla każdego słowa kluczowego).

3. BUDOWANIE MODELU

- Wykorzystano pre-trenowany model językowy BART (facebook/bart-large-mnli)
- Do klasyfikacji dodano dwie warstwy w pełni połączone
- Odmroźono warstwy modelu językowego i douczono całość na anotowanym zbiorze Hoaxes and Hidden agendas: A Twitter Conspiracy Theory Dataset [1] oraz treści opisów teorii spiskowych z Wikipedii.
- Jako teorie spiskowe uznano tweety dla których conspiracy score był większy równy **0.75**
- Jako kontrowersyjne uznawano tweety dla których controversy score był większy równy **0.75**



[1] WWW 22: Companion Proceedings of the Web Conference 2022 Pages 876-880 https://doi.org/10.1145/3487553.3524665

4. WYNIKI

- Osiągnięto dokładność wynoszącą **78,2%**
- Model radzi sobie lepiej z rozpoznawaniem informacji prawdziwych niż spisku

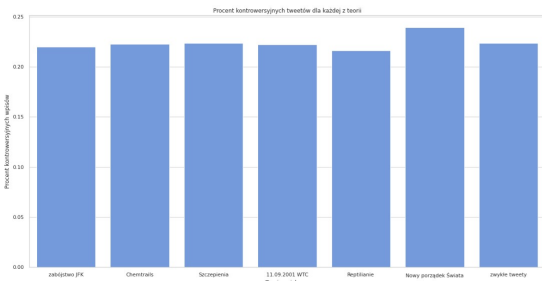
	PRECISION	RECALL	F1-SCORE
PRAWDA	0.8800	0.8148	0.8462
SPISEK	0.5726	0.6907	0.6262

MACIERZ POMYŁEK

	PRAWDA	SPISEK
PRAWDA	220	50
SPISEK	30	67

5. ANALIZY

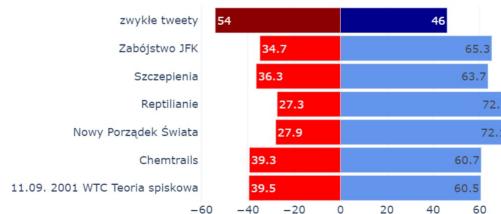
KONTROWERSYJNOŚĆ



Teorie spiskowe nie różnią się znacznie od zwykłych tweetów pod względem kontrowersyjności.

WYDŹWIĘK

negatywne vs pozytywne (% tweetów)



Teorie spiskowe od zwykłych tweetów różnią się tym, że przeważa dla nich pozytywny wydzwięk.

POPULARNOŚĆ

Teorie dzielą się na sezonowe (np. teoria o zamachu na World Trade Center) i o stałym zaangażowaniu użytkowników (np. Reptilianie).

AKTYWNOŚĆ UŻYTKOWNIKÓW



REAKCJE



NOWY PORZĄDEK
ŚWIATA

- największa liczba polubień
- największa liczba retweetów



CHEMTRAILS

- najmniejsza liczba polubień, retweetów i odpowiedzi



REPTILIANIE

- największa liczba odpowiedzi na tweet



SZCZEPHENIA

- wśród wpisów niekontrowersyjnych mają największą liczbę odpowiedzi na tweet

