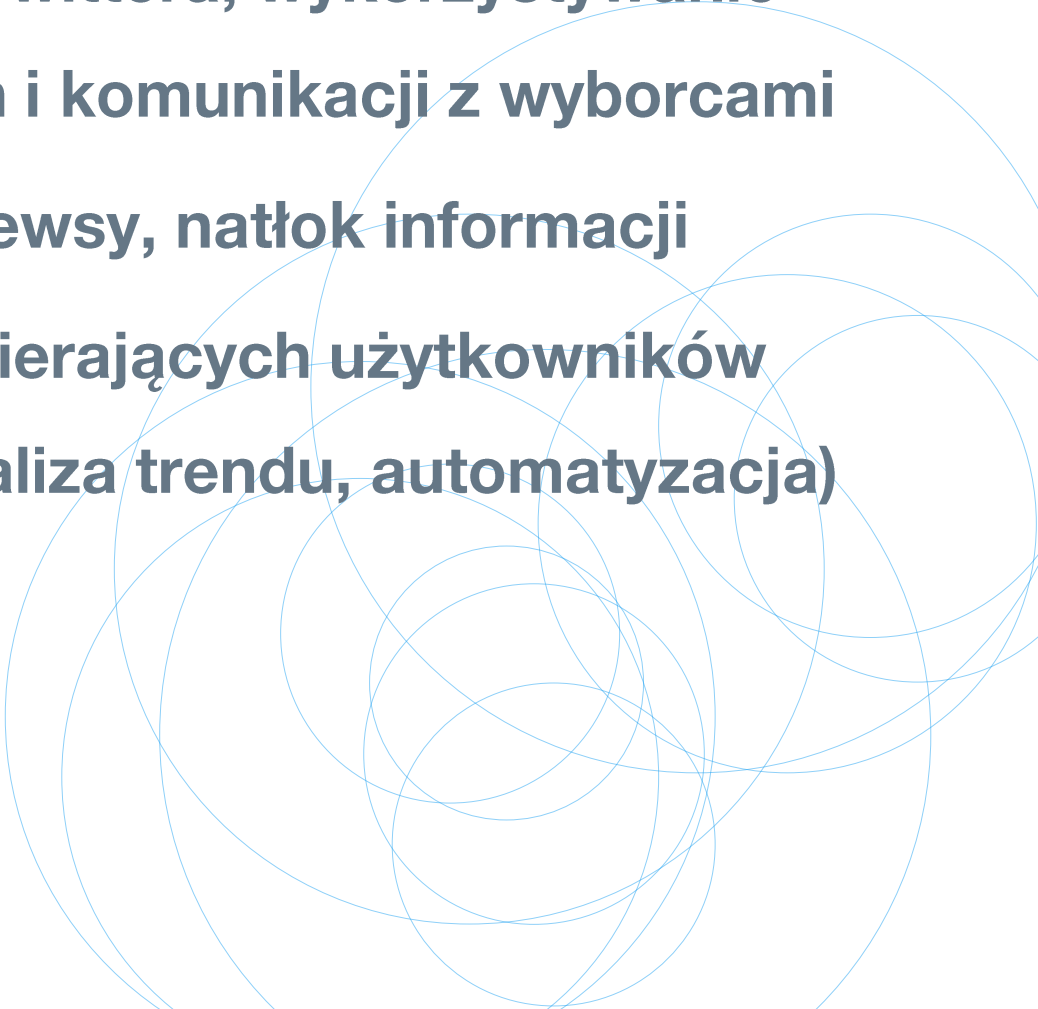


#wpispolityka

**Analiza wypowiedzi polskich
polityków w serwisie Twitter**

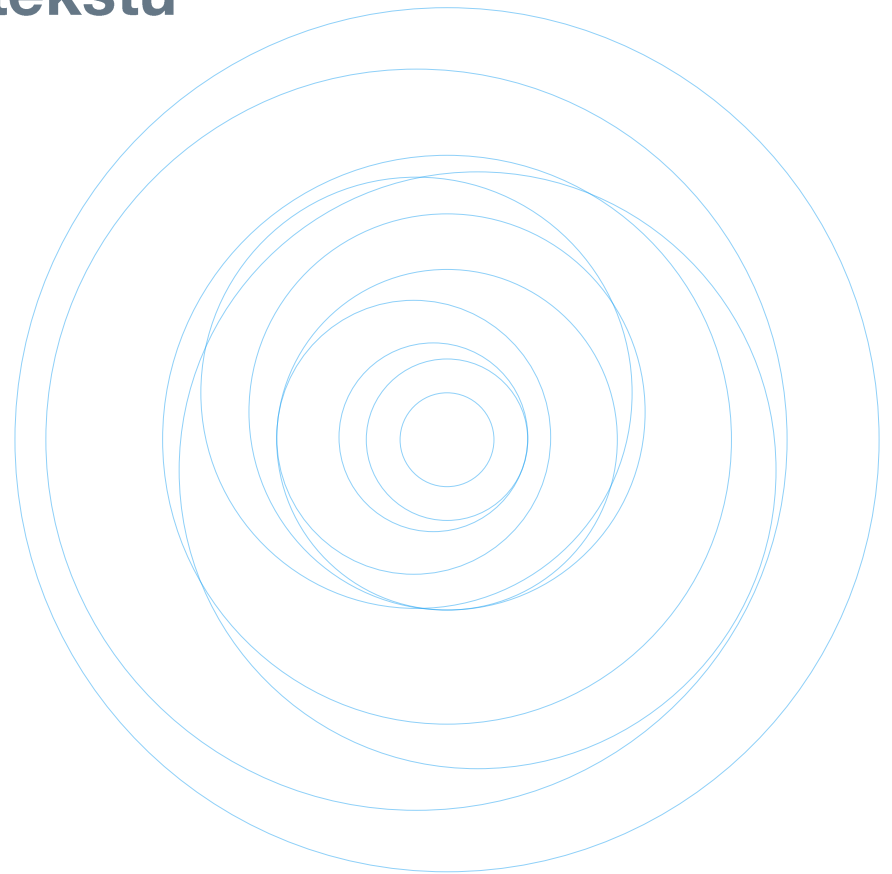


Kontekst i cel

- rosnąca popularność Twittera, wykorzystywanie serwisu w kampaniach i komunikacji z wyborcami
 - fałszywe konta, fake newsy, natłok informacji
 - potrzeba narzędzi wspierających użytkowników (weryfikacja treści, analiza trendu, automatyzacja)
- 

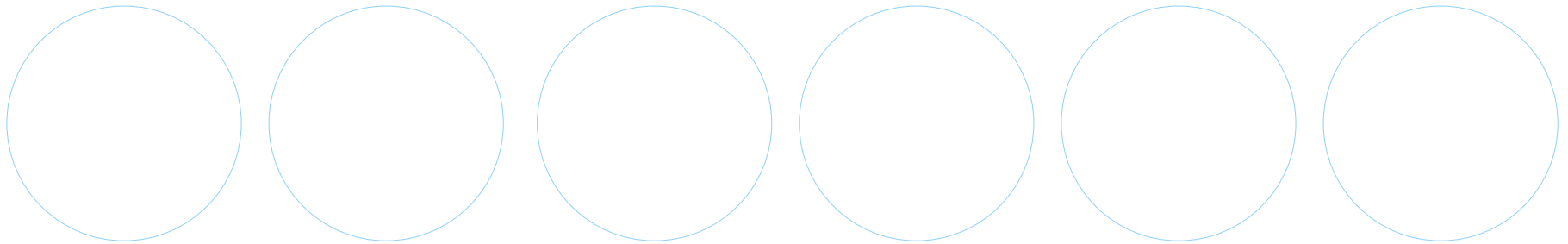
Cel

**stworzenie klasyfikatora przewidującego,
do której partii należy autor tekstu**



Plan

- określenie zakresu projektu
- pobranie danych (API, tweepy)
- eksploracja danych, wizualizacja wyników
- przygotowanie danych dla modelu
- przygotowanie i trenowanie modelu
- analiza wyników

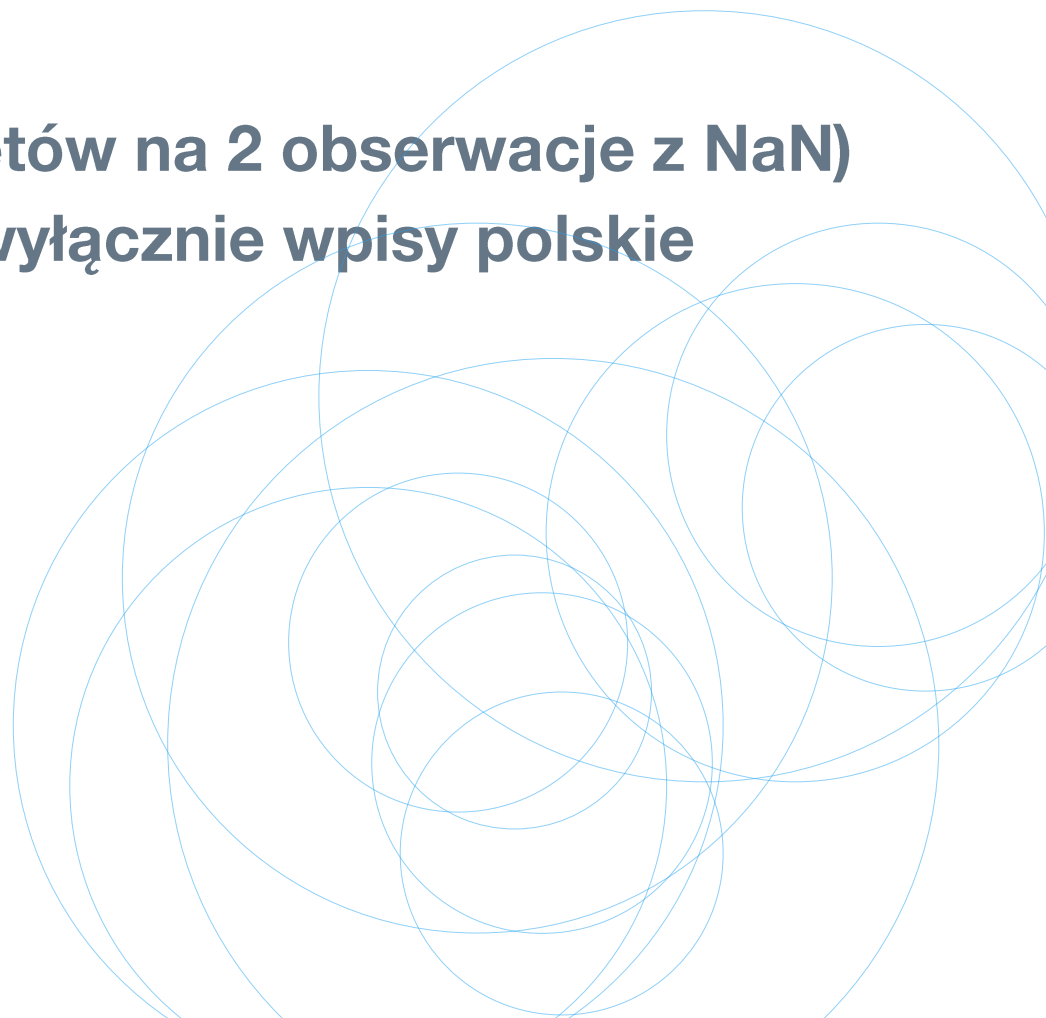


Pobrane dane

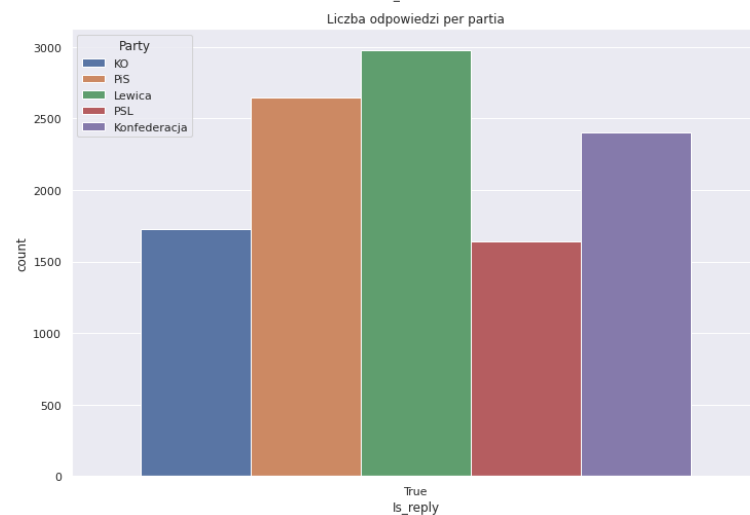
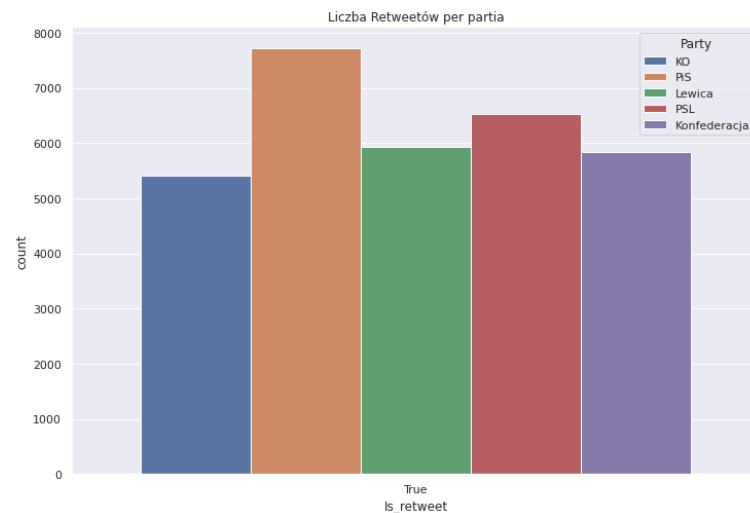
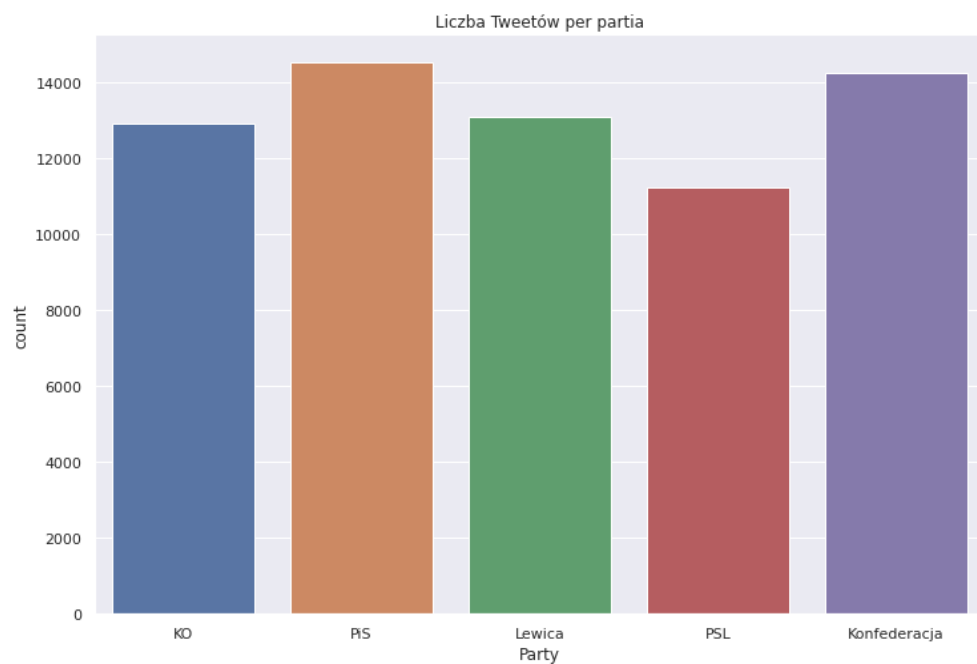
- 5 ugrupowań (PiS, KO, PSL, Lewica, Konfederacja)
- 26 kont polityków, 3 profile partyjne
- wpisy opublikowane po 1 stycznia 2020 r.
- 3000 wpisów per konto (lub wszystkie do 01.01.2021)
- 65954 wpisów:
 - 23161 Tweetów oryginalnych
 - 11396 odpowiedzi
 - 31397 Retweetów
- 14 cech

Eksploracja danych

- usunięcie 8 obserwacji
(błąd pobrania danych
rozbicie tekstu 4 Tweetów na 2 obserwacje z NaN)
- pod uwagę bierzemy wyłącznie wpisy polskie
96% procent wpisów

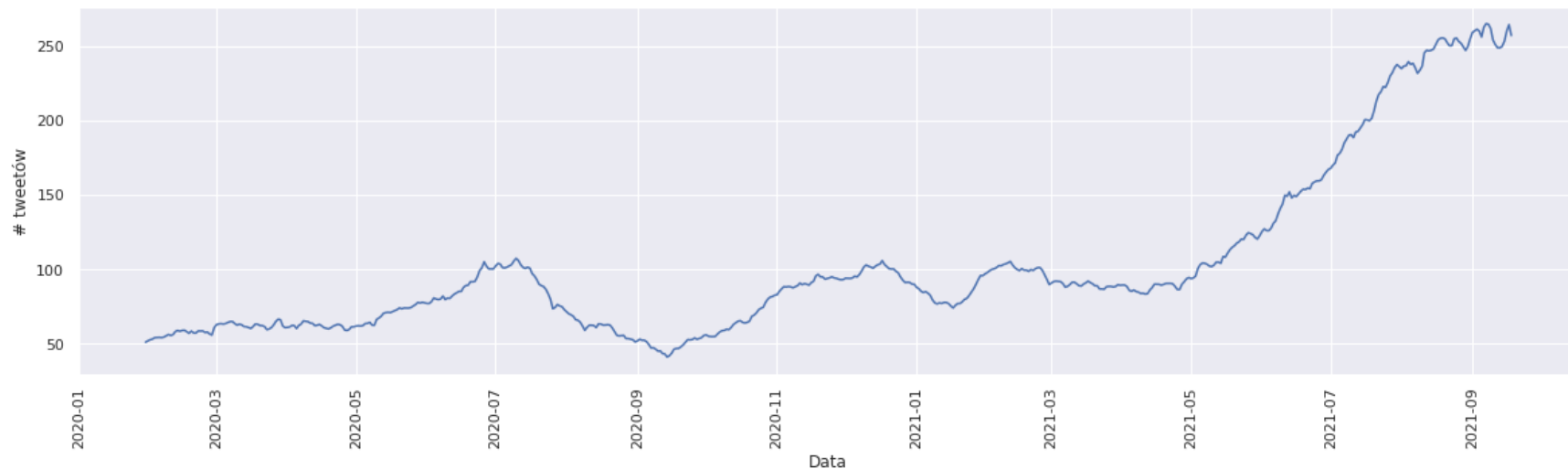


Rozkład danych

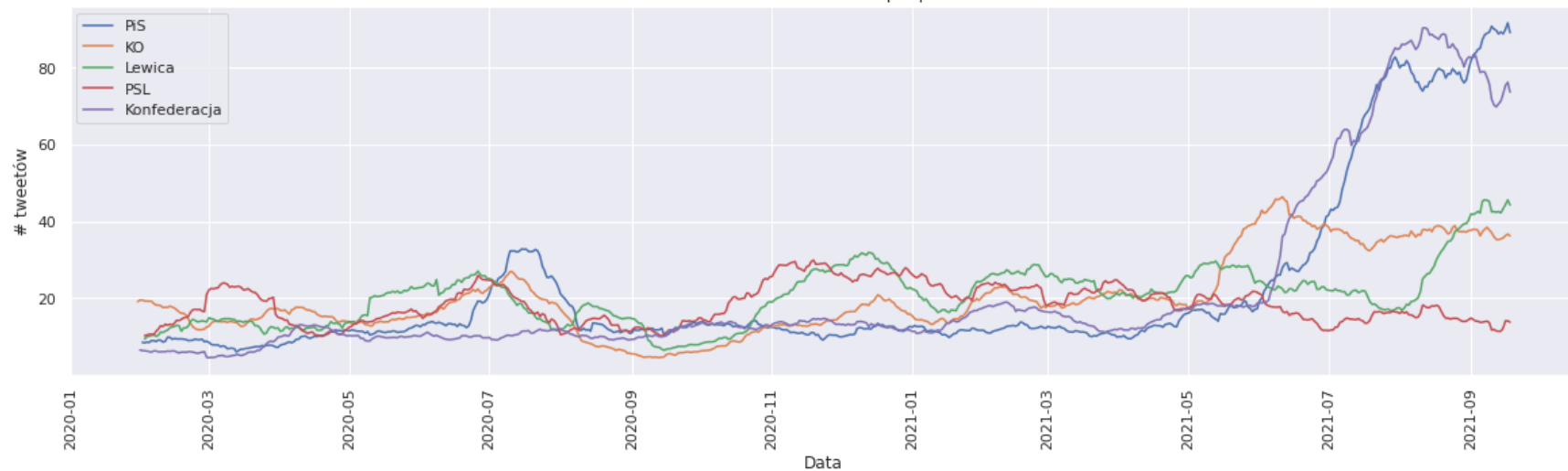


Aktywność w czasie

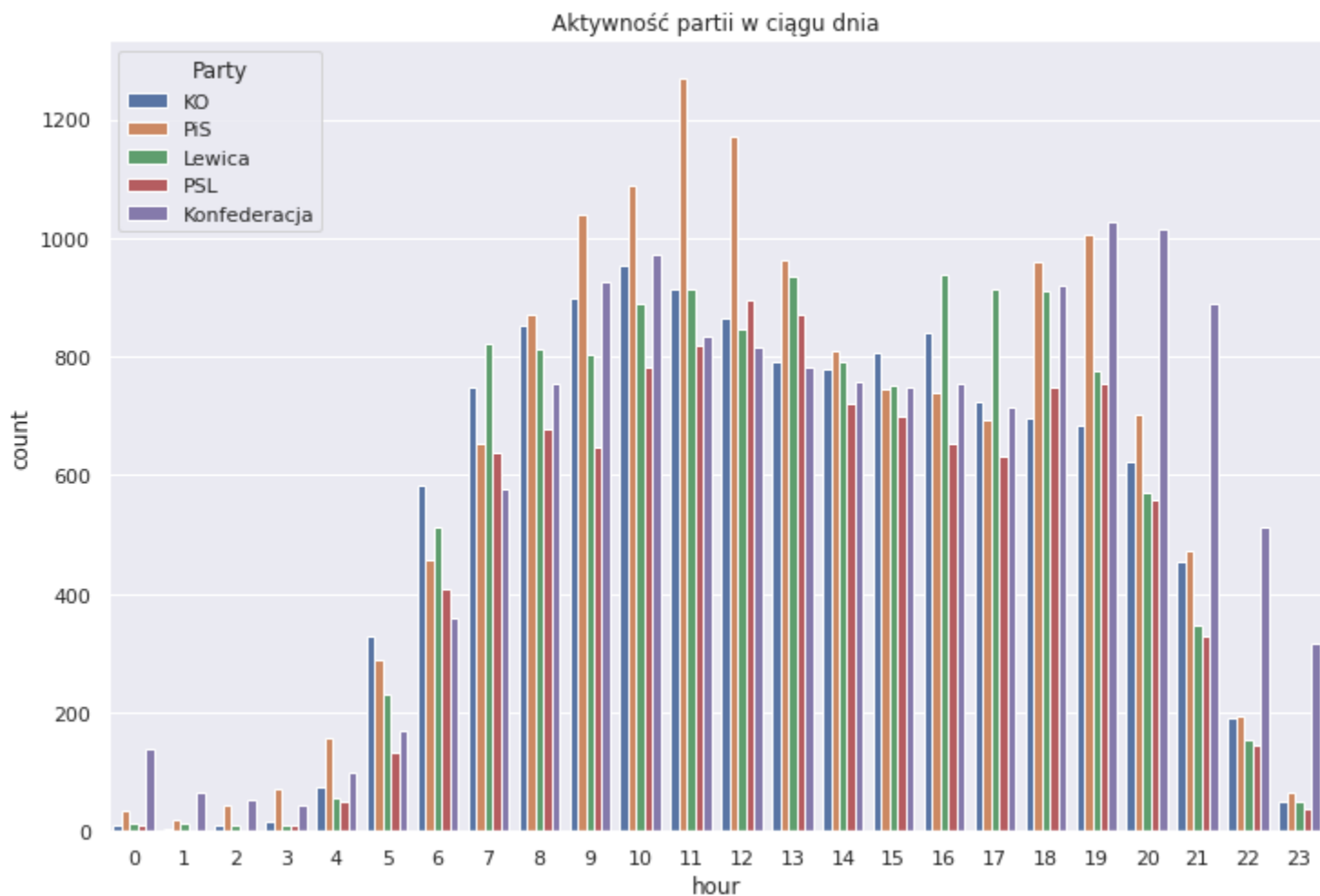
Liczba tweetów w czasie



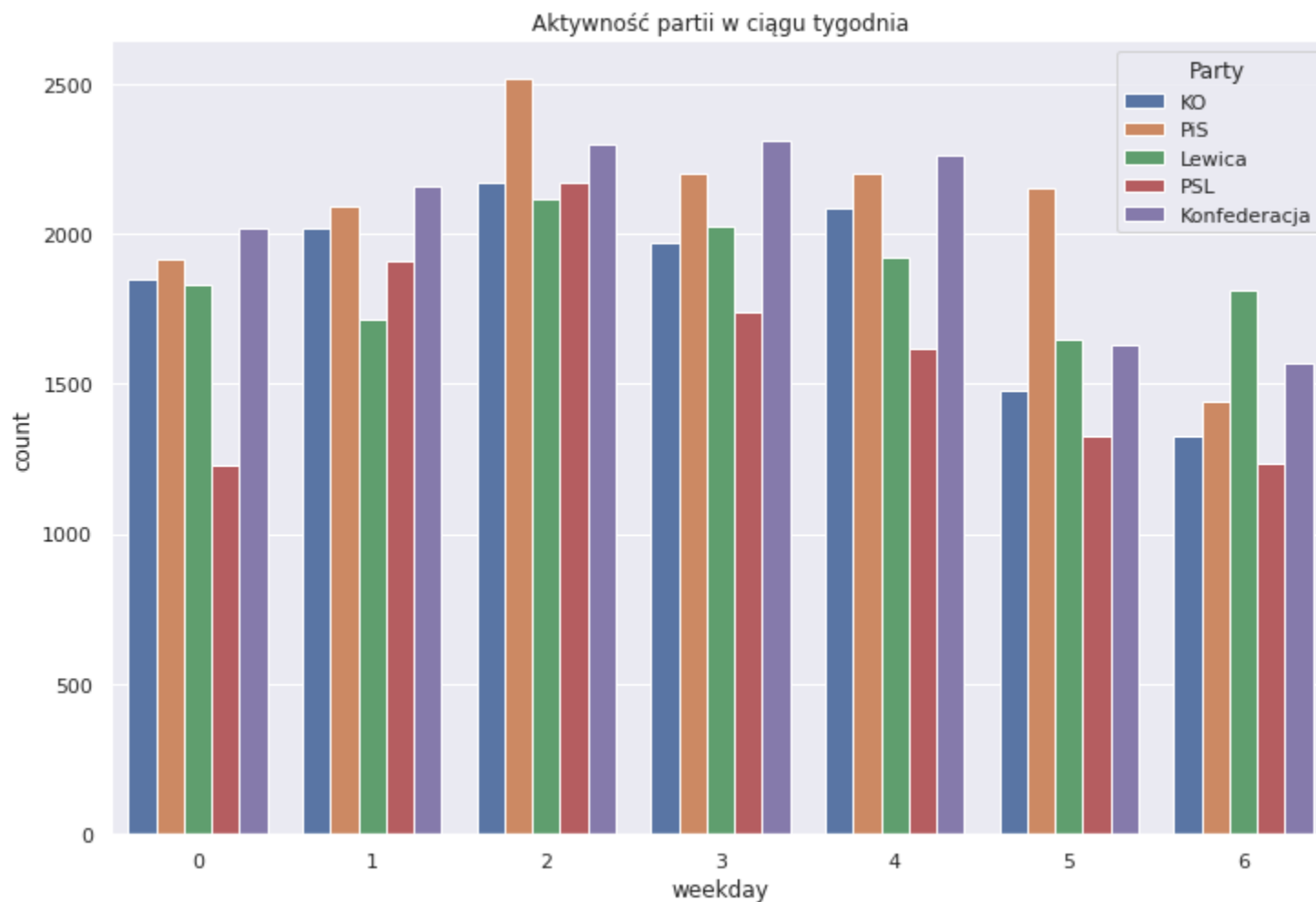
Liczba tweetów w czasie per partia



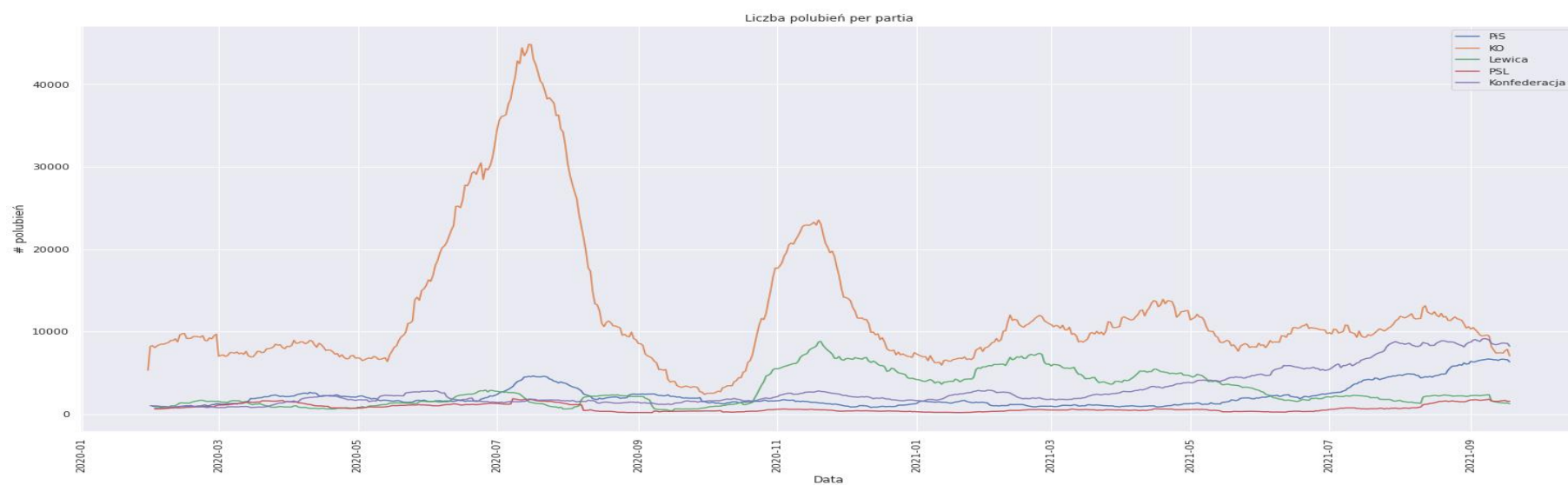
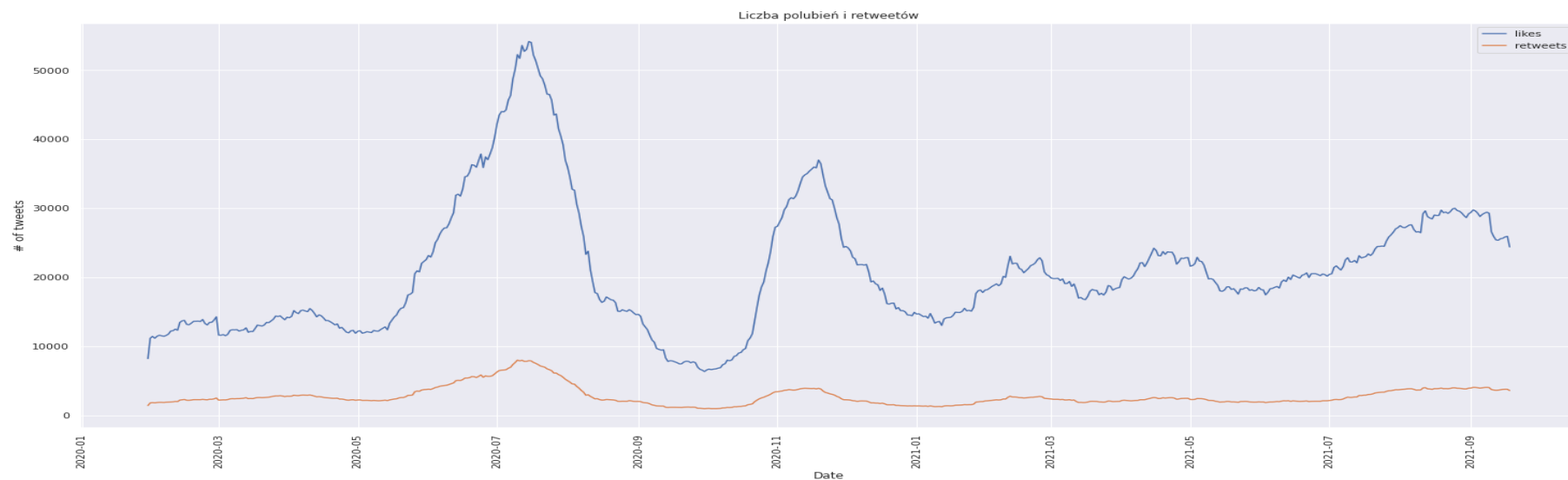
Aktywność w czasie



Aktywność w czasie

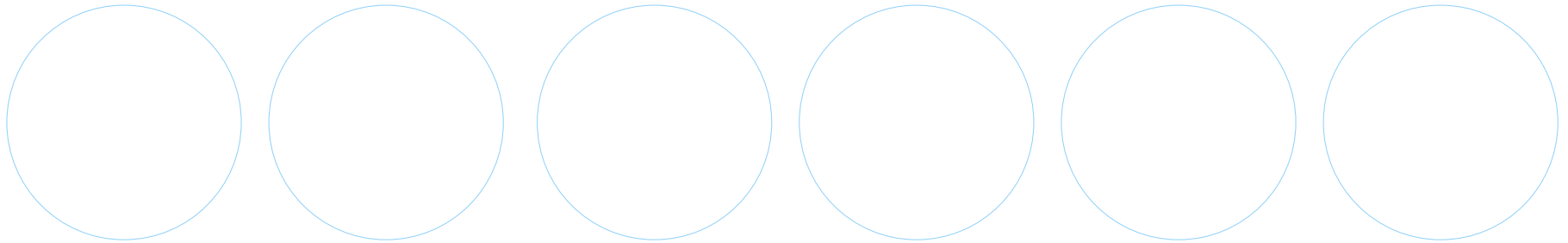


Polubienia, Retweety



Przygotowanie danych

- wczytanie danych (oryginalne wpisy polskie)
- podział na zbiór treningowy i testowy (data=01.08.21)
- przygotowanie korpusu tekstu i etykiet
- obróbka tekstu (niechciane znaki, numery, wielkie litery)
- lematyzacja (spacy)
- tokenizacja



Pierwszy model

Model: "sequential_2"

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, 40, 100)	7275900
gru_2 (GRU)	(None, 16)	5664
dense_2 (Dense)	(None, 5)	85
Total params: 7,281,649		
Trainable params: 7,281,649		
Non-trainable params: 0		
None		

	precision	recall	f1-score	support
KO	0.29	0.75	0.42	1726
Konfederacja	0.00	0.00	0.00	4370
Lewica	0.33	0.78	0.46	1877
PSL	0.36	0.57	0.44	642
PiS	0.76	0.36	0.49	2464
accuracy			0.36	11079
macro avg	0.35	0.49	0.36	11079
weighted avg	0.29	0.36	0.28	11079