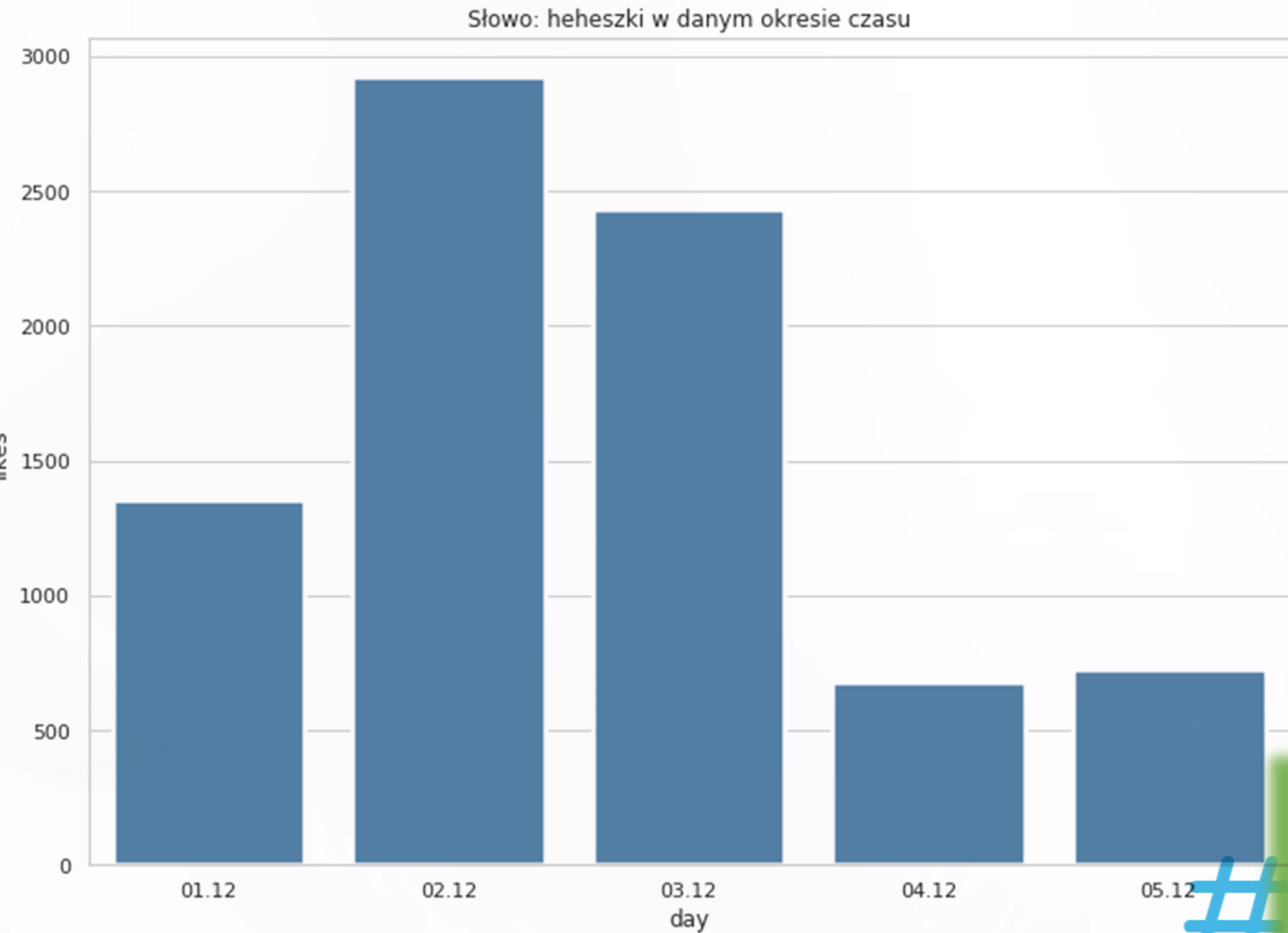




WONTED

WykOp TrEnds Detector

Najpopularniejsze
tematy wpisów

gdańsk
sprzedaż
europea
polski
interes
kobieta
heheszki
salon
hate
akcja
echo
informacja
speech
państwo
mireczkow
podrywajwykopem

Analiza trendów

Podsumowanie popularnych tematów w danym okresie
Wykres zmian popularnych tematów

Akwizycja danych

Wykorzystanie Wykop API
Selekcja danych
(tekst, hashtagi, reakcje)

Czy słowa kluczowe treści wpisów publikowanych na łamach **NAJWIĘKSZEGO** społecznościowego mogą wskazywać na bieżące trendy w polskim Internecie?

Przetwarzanie wstępne

Analiza morfo-syntaktyczna
Oczyszczanie tekstu
Wykrywanie nazw własnych

Budowa klasyfikatora

Maszyna Wektorów Nośnych
Las Drzew Losowych
Rekurencyjna sieć neuronowa LSTM

Inżynieria cech

Miary centralności grafu
TF-IDF
Części mowy
Nazwy własne

Konstrukcja grafu

Graf relacji pomiędzy lematami
Podobieństwo kosinusowe przez wektory zagłębień
Rozszerzanie grafu o relacje semantyczne ze Słownosieci



- Słowa kluczowe definiują trendy tylko częściowo
- Sukces metody detekcji słów kluczowych opartej o grafy relacji
 - Zaszumienie danych
 - Problemy z lematyzacją

CLEAN ALL THE
DATA



Piotr Zieliński

Bartłomiej Kurzeja

Mateusz Zimoch

