

Analiza tekstowa wpisów na portalu Twitter dotyczących problematyki biedy i programów socjalnych

1. Wprowadzanie do projektu

Stygmatyzacja społeczna świadczeniobiorców jest jednym z istotnych czynników wpływających na efektywność polityki społecznej. Wiąże się bezpośrednio z problemem ograniczonego dostępu do świadczeń socjalnych i ich niewykorzystania przez osoby do tego uprawnione. Negatywne postrzeganie i stereotypy wiążące się z pomocą społeczną, przyczyniają się do niskich wskaźników korzystania z systemu - wiele osób, które rzeczywiście potrzebują pomocy, nie zwraca się po nią z powodu wstydu, lęku, strachu czy dumy. Jeżeli dana polityka skierowana jest wyłącznie do potrzebujących, to korzystanie z niej jest równoznaczne z przyznaniem się do nieradzenia sobie w życiu, co dla niektórych jest zbyt upokarzające. W krajach Unii europejskiej szacuję się, że zjawisko to może dotyczyć ponad 40% ludzi teoretycznie uprawnionych do pobierania różnego typu świadczeń. Rozwiązanie tego problemu ma kluczowe znaczenie z dwóch powodów:

- 1) Świadczenia nie spełniają swojego celu, jeśli nie docierają do osób, dla których są przeznaczone. Sprawia to również, że są kosztowo nieefektywne i środki na nie przeznaczone są marnotrawione.
- 2) Ludzie, którzy w ten sposób wypadają z systemu, są często tymi, którzy żyją w najgorszych warunkach. To właśnie na nich powinno nam zależeć najbardziej.

2. Cel Projektu i pytanie badawcze

Problemem, którym chciałbym się zająć w projekcie, jest stygmatyzacja społeczna osób otrzymujących świadczenia społeczne. Celem jest określenie, jakie jest nastawienie Polaków do programów socjalnych i pomocy społecznej oraz jak widzą osoby, które korzystają z tego typu rozwiązań.

Chcę spróbować dowiedzieć się, jaka jest skala negatywnego (lub pozytywnego) sentymentu do wyżej wymienionych zagadnień oraz jakie są najczęstsze określenia, które pojawiają się w dyskursie skoncentrowanym wokół nich.

Być może pozwoli to określić:

- Czy rzeczywiście mamy do czynienia ze stygmatyzacją osób pobierających zasiłki (lub będących w sytuacji życiowej, która ich do tego uprawnia)?
- Czy ogólny sentyment do programów socjalnych jest negatywny, czy pozytywny?
- Jakie określenia najczęściej pojawiają się w bieżącym dyskursie oraz w jakim kontekście się pojawiają.

3. Hipotezy badawcze

Postawione hipotezy brzmią następująco:

H1: sentyment do programów socjalnych i świadczeniobiorców jest zdecydowanie negatywny.

H2: dyskurs toczy się wokół bieżących wydarzeń politycznych (waloryzacja 500+, kwestie migracyjne).

4. Metodologia

- Początkowo miałem zamiar wykorzystać API Twittera w celu pobrania 1500 tweetów z ostatnich 7 dni. Jednak ten plan okazał się niemożliwy ze względu na zmiany w polityce platformy wobec web scrapingu.
- Próba wykorzystania metod pomijających API spotkała się z blokadą ze strony serwisu.
- W celu zebrania potrzebnych wpisów wykorzystałem usługi pozwalające na automatyczne zapisywanie pojawiających się tweetów w skoroszybie (Zapier i PhantomBuster).
- W przypadku Zapier użyłem zapytania: (ubóstwo OR subsydia OR świadczeniobiorcy OR bezrobotni OR biedni OR ubodzy OR świadczenia OR zasiłki OR 500plus OR babciowe OR 800plus OR socjal OR zasiłek) OR (polityka społeczna) OR (państwo opiekuńcze) OR (pomoc społeczna) OR (opieka społeczna) lang:pl. Pozwoliło mi to stworzyć bazę 1000 wiadomości - wśród nich znajdowały się również retweety oraz odpowiedzi (replies).
- W przypadku PhantomBuster użyłem zapytania: (ubóstwo OR subsydia OR świadczeniobiorcy OR bezrobotni OR biedni OR ubodzy OR świadczenia OR zasiłki OR 500plus OR babciowe OR 800plus OR socjal OR zasiłek) OR (polityka społeczna) OR (państwo opiekuńcze) OR (pomoc społeczna) OR (opieka społeczna) lang:pl -filter:replies. Rezultatem w tym przypadku był zbiór 111 wiadomości - nie zawierał on jednak odpowiedzi, a tweety się nie powtarzały. W mojej ocenie zbiór ten jest lepszej jakości, mimo mniejszych rozmiarów.
- Przeprowadziłem analizę na obydwu tych zbiorach danych i w obydwu przypadkach natrafiłem na ten sam problem: narzędzia służące do analizy sentymentu i tagowania części mowy działają bardzo słabo w przypadku języka polskiego (a przynajmniej ja nie potrafię uzyskać rezultatów, które byłyby użyteczne).
- W związku z tym zdecydowałem się przetłumaczyć zbiór danych pozyskanych za pomocą PhantomBuster na język angielski (korzystając z DeepL).
- Zdaję sobie sprawę, że nie jest to idealne rozwiązanie, ale okazało się najlepsze z tych, które były mi dostępne.
- Dokumentacje poprzednich prób dołączyłem do przesłanego rozwiązania.

Ostatecznie:

- Analiza została przeprowadzona na zbiorze 111 tweetów, pobranych na przestrzeni 4 dni za pomocą narzędzia PhantomBuster, używając zapytania (ubóstwo OR subsydia OR świadczeniobiorcy OR bezrobotni OR biedni OR ubodzy OR świadczenia OR zasiłki OR 500plus OR babciowe OR 800plus OR socjal OR zasiłek) OR (polityka społeczna) OR (państwo opiekuńcze) OR (pomoc społeczna) OR (opieka społeczna) lang:pl -filter:replies. Tweety zostały następnie przetłumaczone z języka polskiego na język angielski.

Narzędzia analizy danych:

- Język Python
- Jupyter Notebook

5. Analiza danych

Ładowanie potrzebnych bibliotek i zbioru danych

```
In [1]: # Pandas, nltk, spacy, regex, collections, matplotlib, wordcloud, textblob
import pandas as pd
```

```

pd.set_option('display.max_rows', 1000)
pd.set_option('display.max_columns', 100)
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
import spacy
import regex as re
from collections import Counter
import matplotlib.pyplot as plt
from wordcloud import WordCloud
from textblob import TextBlob

# Wczytanie danych z pliku Excel
df = pd.read_excel('ENG2.xlsx')

# Lista Stop Words
stop_words = set(stopwords.words('english'))

```

```

In [2]: # # # Pobieranie modelu języka angielskiego w Spacy
# !python -m spacy download en_core_web_md

# # Załadowanie modelu języka angielskiego
nlp = spacy.load('en_core_web_md')

```

Funkcja czyszcząca tweety

```

In [3]: def clean_tweet(tweet):

    tweet = re.sub(r"http\S+|www\S+|https\S+", '', tweet, flags=re.MULTILINE) # usuwa ws
    tweet = re.sub(r'@\w+|\#', '', tweet) # usuwa wszystkie wystąpienia oznaczeń użytko
    tweet = re.sub(r'[\W\s]', '', tweet) # usuwa wszelkie znaki, które nie są literami,
    tweet = re.sub(r'\bRT\b', '', tweet) # usuwa "RT" (Retweet) z tekstu tweeta

    # Tokenizacja tekstu na słowa
    tokens = word_tokenize(tweet)
    # Usunięcie znaków interpunkcyjnych i przekształcenie słów na małe litery
    words = [word.lower() for word in tokens if word.isalpha()]
    # Usunięcie stop words
    words = [word for word in words if word not in stop_words]
    # Lematyzacja słów
    lemmas = [token.lemma_ for token in nlp(' '.join(words))]
    # Połączenie lematów w jednozdaniowy tekst
    cleaned_tweet = ' '.join(lemmas)
    return cleaned_tweet

# Stworzenie nowej kolumny z wyczyszczonymi tweetami
df['CLEANED_TEXT'] = df['TEXT'].apply(clean_tweet)

# Tak dla przykładu wygląda pierwsze 30 tweetów
df.head(30)

```

```

Out[3]:

```

	TEXT	CLEANED_TEXT
0	{'Bill Gates wants everyone to experience mala...	bill gates want everyone experience malaria po...
1	{'These are poor people looking for their plac...	poor people look place earthfranceonfire
2	{'Militantly anti-immigrant Tusk on short vide...	militantly antiimmigrant tusk short video poor...
3	{'Shock☹️. Tetiana Kamraceva-Holodupceva, a "re...	shock tetiana kamracevaholodupceva refugee ukr...
4	{'A few brawls in de from the last few days:{'...	brawl last daysformer ambassador w melnyk demo...
5	{'And you what? Still poor?'} I don't even fee...	still poor do not even feel sorry billion fing...

6	{'In today's @gazeta_wyborcza you'll find my c...	today's you'll find commentary actual governmen...
7	{'Let's finally let these people in!'These are...	lets finally let people in these poor people ne...
8	{'800 plus. What you need to know #800plus htt...	plus need know
9	{'If Deputy Speaker of the Sejm, Ryszard Terle...	deputy speaker sejm ryszard terlecki associate...
10	Donald Tusk:{"These are poor people who are lo...	donald tusk these poor people look place earth ...
11	{'merkel and tusk invited the savages, - tusk ...	merkel tusk invite savage tusk growl would pun...
12	{'!!Paweł Zalewski of Poland 2050 against #800...	zalewski poland
13	{'Poor doctors, surgeons.... https://t.co/acEp...	poor doctor surgeon
14	{'President Macron, instead of taking away the...	president macron instead take away welfare tur...
15	{'Well poor people those bows for mitch were n...	well poor people bow mitch normally scary must...
16	{'PL Thank you for the meeting in #Olsztyn!} W...	thank meet olsztyn talk dignity revolution tak...
17	{'📺 The Council of Ministers today passed a bi...	council minister today pass bill increase benefit
18	Minister @LukaszSchreiber vs @styszka {'Minist...	minister vs minister vs take away social secur...
19	{'📺 Today's meeting of the Council of Minister...	today's meet council minister chair prime minis...
20	{'1) LICENSING SOCIETY FOR IMMIGRANTS - if an ...	license society immigrant immigrant work polan...
21	{'This is the one and only team in Poland, tha...	one team poland thanks swinoujście tunnel gazo...
22	{'The twelfth #Social Insurance Evening has ju...	twelfth social insurance evening begin topic s...
23	{'Bochenek, by not hiring employees in the MP'...	bochenek hire employee mps office explain over...
24	{'The German government has adopted a draft bu...	german government adopt draft budget provide s...
25	{'500 Plus boosts violence? Surprising stateme...	plus boost violence surprising statement po sp...
26	{'💬 Minister @MarlenaMalag: With the increase o...	minister increase benefit january parent careg...
27	On June 1, the new benefit period began, appli...	june new benefit period begin application subm...
28	{'Poor people. Those having a child. How much ...	poor people child much listen living room
29	{'Statutory poverty (annual CSO data): 13% f...	statutory poverty annual cso datum low child p...

Analiza Sentymentu

```
In [4]: # Analiza poziomu subiektywności tekstu
def getSubjectivity(tweet):
    return TextBlob(tweet).sentiment.subjectivity

# Analiza polaryzacji tekstu
def getPolarity(tweet):
    return TextBlob(tweet).sentiment.polarity

df['Subjectivity'] = df['CLEANED_TEXT'].apply(getSubjectivity)
df['Polarity'] = df['CLEANED_TEXT'].apply(getPolarity)

# Przykładowe 30 wierszy
df.head(30)
```

Out[4]:	TEXT	CLEANED_TEXT	Subjectivity	Polarity
0	{'Bill Gates wants everyone to experience mala...	bill gates want everyone experience malaria po...	0.600000	-0.400000

1	{'These are poor people looking for their plac...	poor people look place earthfranceonfire	0.600000	-0.400000
2	{'Militantly anti-immigrant Tusk on short vide...	militantly antiimmigrant tusk short video poor...	0.450000	-0.200000
3	{'Shock😞. Tetiana Kamraceva-Holodupceva, a "re...	shock tetiana kamracevaholodupceva refugee ukr...	0.500000	-0.025000
4	{'A few brawls in DE from the last few days:{'...	brawl last daysformer ambassador w melnyk demo...	0.354167	0.021875
5	{'And you what? Still poor?'} I don't even fee...	still poor do not even feel sorry billion fing...	0.800000	-0.450000
6	{'In today's @gazeta_wyborcza you'll find my c...	todays you ll find commentary actual governmen...	0.100000	0.000000
7	{'Let's finally let these people in!These are...	lets finally let people inthese poor people ne...	0.711905	-0.038095
8	{'800 plus. What you need to know #800plus htt...	plus need know	0.000000	0.000000
9	{'If Deputy Speaker of the Sejm, Ryszard Terle...	deputy speaker sejm ryszard terlecki associate...	0.066667	0.033333
10	Donald Tusk:{'These are poor people who are lo...	donald tuskthese poor people look place earth ...	0.675000	-0.625000
11	{'merkel and tusk invited the savages, - tusk ...	merkel tusk invite savage tusk growl would pun...	0.600000	-0.400000
12	{'!!Paweł Zalewski of Poland 2050 against #800...	zalewski poland	0.000000	0.000000
13	{'Poor doctors, surgeons.... https://t.co/acEp...	poor doctor surgeon	0.600000	-0.400000
14	{'President Macron, instead of taking away the...	president macron instead take away welfare tur...	0.225595	0.154762
15	{'Well poor people those bows for mitch were n...	well poor people bow mitch normally scary must...	0.653333	-0.260000
16	{'PL Thank you for the meeting in #Olsztyn!} W...	thank meet olsztyn talk dignity revolution tak...	1.000000	0.000000
17	{'📷 The Council of Ministers today passed a bi...	council minister today pass bill increase benefit	0.000000	0.000000
18	Minister @LukaszSchreiber vs @styszka {'Minist...	minister vs minister vs take away social secur...	0.111111	0.166667
19	{'📷 Today's meeting of the Council of Minister...	todays meet council minister chair prime minis...	0.000000	0.000000
20	{'1) LICENSING SOCIETY FOR IMMIGRANTS - if an ...	license society immigrant immigrant work polan...	0.200000	0.200000
21	{'This is the one and only team in Poland, tha...	one team poland thanks swinoujscie tunnel gazo...	0.483333	0.166667
22	{'The twelfth #Social Insurance Evening has ju...	twelfth social insurance evening begin topic s...	0.066667	0.033333
23	{'Bochenek, by not hiring employees in the MP'...	bochenek hire employee mps office explain over...	0.000000	0.000000
24	{'The German government has adopted a draft bu...	german government adopt draft budget provide s...	0.350000	0.266667

25	{500 Plus boosts violence? Surprising stateme...	plus boost violence surprising statement po sp...	0.500000	0.700000
26	{ Minister @MarlenaMalag: With the increase o...	minister increase benefit january parent careg...	0.000000	0.000000
27	On June 1, the new benefit period began, appli...	june new benefit period begin application subm...	0.262626	0.128788
28	{Poor people. Those having a child. How much ...	poor people child much listen living room	0.400000	-0.100000
29	{Statutory poverty (annual CSO data):} 13% f...	statutory poverty annual cso datum low child p...	0.335000	0.090000

W kolumnie 'Subjectivity' wyświetlany jest poziom subiektywności danego tweeta. Subiektywność określa ilość osobistych opinii i faktycznych informacji zawartych w tekście. Wysoka subiektywność oznacza, że tweet zawiera raczej opinie niż fakty.

- Przyjmuje wartości od 0 do 1 (0 - faktyczne informacje, 1 - subiektywna opinia).

W kolumnie 'Polarity' wyświetlana jest polaryzacja danego tekstu. Polaryzacja odnosi się do ogólnego sentymentu wyrażonego w danym tweecie.

- Przyjmuje wartości od -1 do 1 (-1 - bardzo negatywne, 0 - neutralne, 1 - bardzo pozytywne).

```
In [5]: # Dodanie kolumny 'Sentiment' określającej ogólny sentyment na podstawie wyniku analizy
# Wynik mniejszy niż 0 = negatywny
# Wynik równy 0 = neutralny
# Inaczej = pozytywny
def getSentiment(score):
    if score < 0:
        return "Negative"
    elif score == 0:
        return "Neutral"
    else:
        return "Positive"

df["Sentiment"] = df['Polarity'].apply(getSentiment)

# Przykładowe 30 wierszy
df.head(30)
```

	TEXT	CLEANED_TEXT	Subjectivity	Polarity	Sentiment
0	{Bill Gates wants everyone to experience mala...	bill gates want everyone experience malaria po...	0.600000	-0.400000	Negative
1	{These are poor people looking for their plac...	poor people look place earthfranceonfire	0.600000	-0.400000	Negative
2	{'Militantly anti-immigrant Tusk on short vide...	militantly antiimmigrant tusk short video poor...	0.450000	-0.200000	Negative
3	{'Shock 😡. Tetiana Kamraceva-Holodupceva, a "re...	shock tetiana kamracevaholodupceva refugee ukr...	0.500000	-0.025000	Negative
4	{'A few brawls in DE from the last few days:{'...	brawl last daysformer ambassador w melnyk demo...	0.354167	0.021875	Positive
5	{'And you what? Still poor?'} I don't even fee...	still poor do not even feel sorry billion fing...	0.800000	-0.450000	Negative
6	{'In today's @gazeta_wyborcza you'll find my c...	today's you ll find commentary actual governmen...	0.100000	0.000000	Neutral

7	{Let's finally let these people in!'These are...	lets finally let people inthese poor people ne...	0.711905	-0.038095	Negative
8	{'800 plus. What you need to know #800plus htt...	plus need know	0.000000	0.000000	Neutral
9	{If Deputy Speaker of the Sejm, Ryszard Terle...	deputy speaker sejm ryszard terlecki associate...	0.066667	0.033333	Positive
10	Donald Tusk:{"These are poor people who are lo...	donald tuskthese poor people look place earth ...	0.675000	-0.625000	Negative
11	{merkel and tusk invited the savages, - tusk ...	merkel tusk invite savage tusk growl would pun...	0.600000	-0.400000	Negative
12	{!!Paweł Zalewski of Poland 2050 against #800...	zalewski poland	0.000000	0.000000	Neutral
13	{'Poor doctors, surgeons.... https://t.co/acEp...	poor doctor surgeon	0.600000	-0.400000	Negative
14	{'President Macron, instead of taking away the...	president macron instead take away welfare tur...	0.225595	0.154762	Positive
15	{'Well poor people those bows for mitch were n...	well poor people bow mitch normally scary must...	0.653333	-0.260000	Negative
16	{PL Thank you for the meeting in #Olsztyn!} W...	thank meet olsztyn talk dignity revolution tak...	1.000000	0.000000	Neutral
17	{📷 The Council of Ministers today passed a bi...	council minister today pass bill increase benefit	0.000000	0.000000	Neutral
18	Minister @LukaszSchreiber vs @styszka {'Minist...	minister vs minister vs take away social secur...	0.111111	0.166667	Positive
19	{📷 Today's meeting of the Council of Minister...	todays meet council minister chair prime minis...	0.000000	0.000000	Neutral
20	{'1) LICENSING SOCIETY FOR IMMIGRANTS - if an ...	license society immigrant immigrant work polan...	0.200000	0.200000	Positive
21	{'This is the one and only team in Poland, tha...	one team poland thanks swinoujscie tunnel gazo...	0.483333	0.166667	Positive
22	{'The twelfth #Social Insurance Evening has ju...	twelfth social insurance evening begin topic s...	0.066667	0.033333	Positive
23	{'Bochenek, by not hiring employees in the MP'...	bochenek hire employee mps office explain over...	0.000000	0.000000	Neutral
24	{'The German government has adopted a draft bu...	german government adopt draft budget provide s...	0.350000	0.266667	Positive
25	{'500 Plus boosts violence? Surprising stateme...	plus boost violence surprising statement po sp...	0.500000	0.700000	Positive
26	{💬 Minister @MarlenaMalag: With the increase o...	minister increase benefit january parent careg...	0.000000	0.000000	Neutral
27	On June 1, the new benefit period began, appli...	june new benefit period begin application subm...	0.262626	0.128788	Positive
28	{'Poor people. Those having a child. How much ...	poor people child much listen living room	0.400000	-0.100000	Negative
29	{'Statutory poverty (annual CSO data): 13% f...	statutory poverty annual cso datum low child p...	0.335000	0.090000	Positive

```
In [6]: # Wyświetlenie wszystkich pozytywnych tweetów
j=1
sortedDF = df.sort_values(by=['Polarity'])
for i in range(0, sortedDF.shape[0]):
    if(sortedDF['Sentiment'][i] == 'Positive'):
        print(str(j) + ' ') +sortedDF['CLEANED_TEXT'][i])
        print()
        j=j+1
```

- 1) brawl last daysformer ambassador w melnyk demolish c masala military expert friendly recommend reading dostoevskyn avanti parental allowance will not rich income per yearwhe ther ketchup need keep fridge
- 2) deputy speaker sejm ryszard terlecki associates even asys social pay employment contr act something amiss
- 3) president macron instead take away welfare turd devastate france sensibly propose rig ht wants restrict social media nonlatin people you re year old click elephant yes solve problem n
- 4) minister vs minister vs take away social security noneu citizen want abolish ukraiania n citizen well maybe would worth ask polesand classic
- 5) license society immigrant immigrant work poland pay taxis should not get welfare conf ederation want give welfare citizenship work respect much
- 6) one team poland thanks swinoujscie tunnel gazoport vistula spit canal baltic pipe pkn orlen lot expressway investment soon central transportation port powerful army
- 7) twelfth social insurance evening begin topic social insurance benefit case sickness m aternity part three event leave series meeting cordially invitedemployersrp
- 8) german government adopt draft budget provide saving ministry except defense great agi tation cause plan low income ceiling entitlement parental benefit per year per family be nefit
- 9) plus boost violence surprising statement po spokesman read
- 10) june new benefit period begin application submit electronically deputy minister repu blicafternoon program family program program also first propose start next year benefit pln
- 11) statutory poverty annual cso datum low child poverty rate eu low unemployment eugt d ecrease monthonmonth inflation june high economic growth oecd
- 12) accelerate legislative work increase benefit pln today social policy family committe e unanimously approve government draft increase
- 13) legia leave austria clean slate interesting thing spar match goalkeeper virtually un employed know friendly match good defense key success
- 14) pole do not che pis po want immigration policy ie polish taxpayer fund social benefi t immigrant world happen come that s po pis government legislate
- 15) helpless mother child poor people seek happiness earth say illegal migrant healthy y oung guysfranceriot francehasfallen franceonfire emeute nanterre nahel
- 16) opposition party voter want reduction program large part want benefit eliminate alto gether meanwhile politician main formation contrary expectation electorate racing populi sm law justice party propose valorization
- 17) intense discussion immigration let remind article broad context cause today mass mig ration movement still remain relevant
- 18) intel invest poland we build intel new factory scandal poor pole fund factory rich a

merican

19) well year get knee perfect budget patient diarrhea stoma due short loop syndrome intestinal resection due rare tumor myxosarcoma welfare solidarity state unable provide advice nutritionist national health service

20) scandalous statement depreciate government action result family live decent condition show pathological situation justify fund pathology disgust

21) sasin take funeral benefit breakthrough do not wait good night

22) plus benefit certainly do job change financial situation family confirm statistical study also datum give eupoland among leader come decrease poverty marlenamalag

23) fulfil another announcement make prime minister jaroslaw kaczynski president program ul sejm working family program continuation consistent policy support polish family

24) pis spokesman sejm program increase fertility family come third subsequent child be not program would not birth rate large family certainly geopolitical situation influence decision

25) see kind new subculture culture fraudedfrauded fact people do not eat vegan since easy cheap they re poor try hard start strimingthey do not feed barf diet animal

26) raphael master stitch tell we program increase fertility family come third subsequent child do not give fact know woman three childrenso much increase mr

27) top kek xd local neighborhood groupkai information city renovating streetcar cross poor stand traffic jam selfdriver cyclist fault xd xd xd

28) glossary today voluntary social security corporate welfare fund future pension

29) first read citizen bill widow pension underway bill allow senior count high benefit lose spouse barely people law justice bench take care elderly

30) first read citizen widow pension project underway widow widower would entitled much high benefit present side law justice party emptiness concern senior senior citizen look like pis opposition support project

31) come surprise migration disturb wealthy great text

32) social security financial pyramid social insurer state guarantee payment one benefit system much jdgc pay relatively low premium security do not make money collect premium pay benefit

33) vice minister sejm since social policy polsce fundamentally change become modern policy fact introduce number universal program

34) sejm want fight poverty project deepen poverty among senior want eradicate poverty hunger among child increase money education every child school get hot mealn

35) sasin spectacular success do not chance mixed feeling hear take death benefit ask curiosity

36) valorization yes work people family policy must value labor force participation demographic crisis require systemic solution build strong economy work education electoral bidding

37) another step pass parliamentary committee social policy family adopt bill increase benefit zloty per child

38) supportive benefit people disability lose pioszachy ozn pis benefit okopress

39) people stigmatization single mother take care child home embarrass news race economic confiability get underway see nowhere europe child benefit tie work shock births

```
In [7]: # Wyświetlenie wszystkich negatywnych tweetów
j=1
sortedDF = df.sort_values(by=['Polarity'], ascending = False)
for i in range(0, sortedDF.shape[0]):
    if(sortedDF['Sentiment'][i] == 'Negative'):
        print(str(j) + ') '+sortedDF['CLEANED_TEXT'][i])
        print()
        j=j+1
```

- 1) bill gates want everyone experience malaria poor people
- 2) poor people look place earthfranceonfire
- 3) militantly antiimmigrant tusk short video poor people look place earth storm border p
art lukashenkos hybrid attackthis hypocrite probably think pole do not remember insult t
hink believe
- 4) shock tetiana kamracevaholodupceva refugee ukraine come poland benefit germany long s
trength push muslims line benefit poland different civilization year I m benefit culture
see
- 5) still poor do not even feel sorry billion fingertip prefer go work blamepoverty weal
th money gold
- 6) lets finally let people inthese poor people need helpif will not even noticelet tear
wall take right awayfrance
- 7) donald tuskthese poor people look place earth need make disgusting grim propaganda mi
grant people need help
- 8) merkel tusk invite savage tusk growl would punish nonacceptance poor citizen beat kic
k guest kraut
- 9) poor doctor surgeon
- 10) well poor people bow mitch normally scary must shit pant xd guy do not like fun turn
blind eye thing admit
- 11) poor people child much listen living room
- 12) poor people child much listen living room wall
- 13) poor man
- 14) germanyor ukparent disabled childneede constant carereceive support statehousingadap
tedmedical carebudgetallowance amount draftworke personcar depend disability
- 15) opposition club fully support program remain however family least one person work in
clude valorization let appreciate people hard work
- 16) donald tusk manage back seat poor symmetrist pis politician
- 17) attention sondaafter latest shocking statement folksdojnikin contradict himselfthere
high risk schizophreniaare favor put compulsory psychiatric care due high social harmple
ase participateamprtleftist do not sleep
- 18) valorization yes work people case family parent work collect social benefit hardwork
e people economically active value
- 19) drunk ukrainian woman neglect child tragedy almost occurredthis also show broad cont
ext daily basis see large number nonworke sit fashionable pub day drink alcoholbut work
get benefit
- 20) bark wrong tree purpose benefit fund law justice can not even die labor office line
job offer germany bambo even get poland still leave germany like polish unemployed

- 21) grandma watch show lefts convention header something welfare state e want give every one change channel country suppose minimally well what s training low pension bad high one
- 22) ey apply disability benefit disease think oh shit
- 23) minister warsaw program improve quality life polish family compare eu country low risk poverty currently work change program january government difficult
- 24) government vote today sejm government already know poland excessive deficit procedure edp opposition know government write explicitly multiannual state financial plan recommendation european commission recommend reduce spending dynamic
- 25) someone age remember say france elegance mess dirt poverty meter muck sick leftist government import black thug africa getting rid
- 26) pis spokesman invest polish child future young support polish family restore dignity since introduction program manage reduce extreme poverty polish family
- 27) submit amendment go economically active people also extend valorization single mother raise child want recognize people work hard activate take work
- 28) well impeacefulnextelection sausagesasin ill everything get death benefit increase election
- 29) poor japanese owe
- 30) analyze change donald tusk view among thing construction dam border admission illegal migrant plus benefit reconstruction unit polish army assessment putin russia leave choice enroll law justice party logical conclusion
- 31) poor also support confederation place awful american accept thing los angeles dream city average pole one walk drive straight garage home garage work do not see poor walk dystopia
- 32) poor refugee want escape war
- 33) chairman explain sejm plus meet demographic target idea link benefit employment trigger avalanche births ill admit logic jcc overwhelmed anyone understand would work
- 34) increase zloty apply economically active family least one person work also pole say meeting say say behind scene therefore vote favor solutionsejm
- 35) demography barrier pl development break demographic impasse say representative government previous term justify project impasse unfortunately break see few birth since world war iisejm
- 36) sejm last june jk convince we valorization impossible would proinflationary move must admit rarely agree chairman kaczynski case n
- 37) poor migrant wait impatiently folksdojczs next instructionsas matter factthis late selfpromotion classic break even campaignsoon poll show waiter typical losernevertheless wait wincyjfur dojczland naturliś
- 38) necessary appreciate active work population submit amendment valorization include work people condition approval demand support project work must pay
- 39) look poor ukrainian can not find one ill help village ostrówki volyn katyn lie every where chop bone skull pole poles kill axis trusty field lie bone polish child let ask son grandson sotenny klymchak
- 40) migrant unemployment european welfare provide high standard living work hard home that s welfare benefit need cut people get europe way

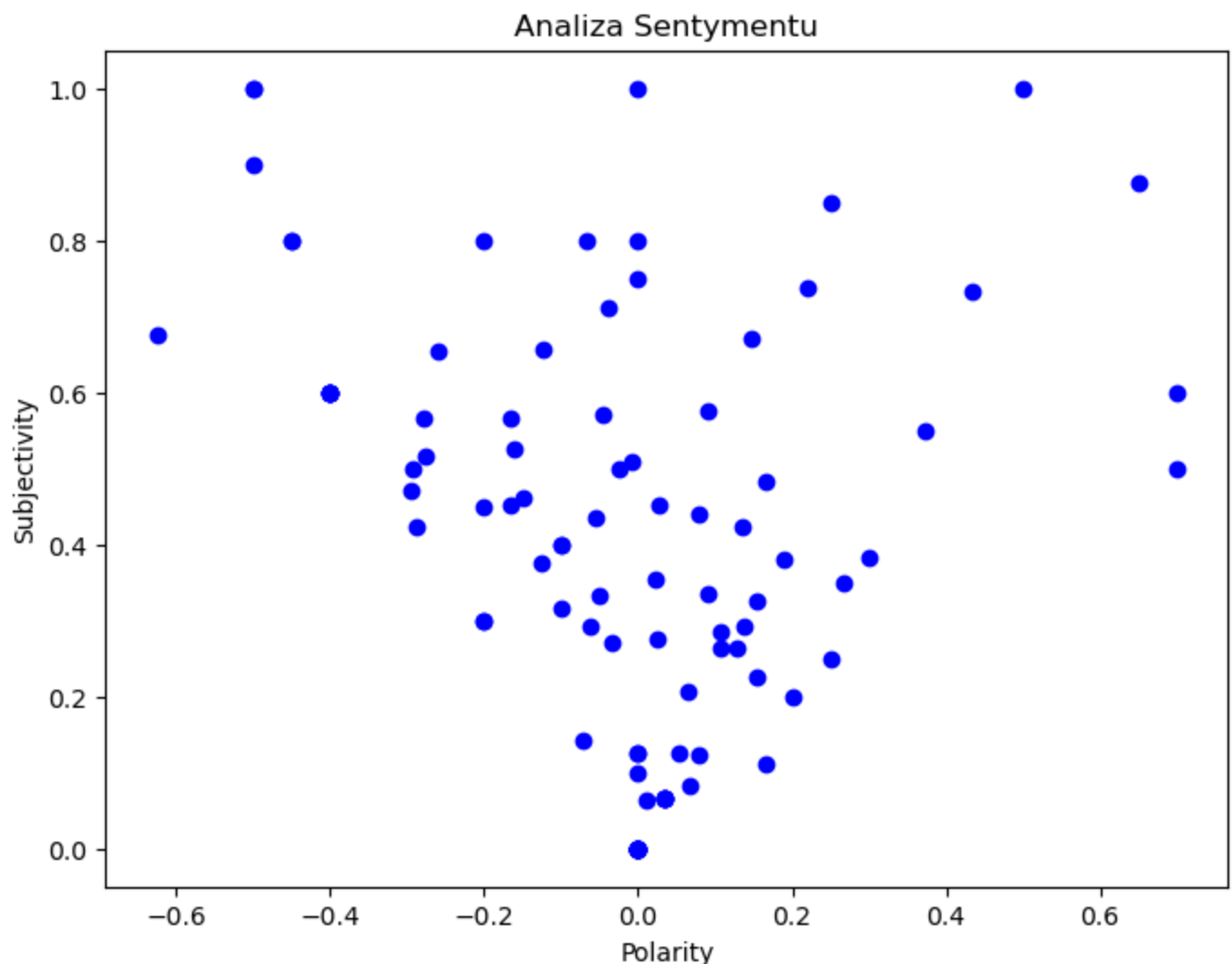
41) hear leave would lift sky pole power even limited benefit mop

39 tweetów zostało uznanych za wiadomości o sentymencie pozytywnym, a 41 za wiadomości o sentymencie negatywnym (neutralnych pozostaje 31).

Warto zwrócić uwagę na to, że te klasyfikacje często bywają nietrafione lub kontrowersyjne.

```
In [8]: # Wykres punktowy przedstawiający polaryzację/subiektywność
plt.figure(figsize=(8,6))
for i in range(0, df.shape[0]):
    plt.scatter(df['Polarity'][i], df['Subjectivity'][i], color = 'Blue')

plt.title('Analiza Sentymentu')
plt.xlabel('Polarity')
plt.ylabel('Subjectivity')
plt.show()
```



Z wykresu można odczytać, że:

- Ogólna polaryzacja jest bardzo nieznacznie negatywna.
- Poziom subiektywności jest nieznacznie bliższy 0 niż 1 - jest więc trochę więcej faktów niż czystych opinii.
- Czym wyższa subiektywność, tym polaryzacja jest wyraźniej bardziej 'rozstrzelona' - to oznacza, że opinie zawarte w tweetach są bardzo różne (niektóre są skrajnie negatywne, a niektóre skrajnie pozytywne).

- Czym niższa subiektywność (więcej faktów w danym tekście), tym polaryzacja jest średnio bliższa 0 - wyrażony sentyment jest więc bliższy neutralnemu.

```
In [9]: # % tweetów pozytywnych
ptweets = df[df.Sentiment == 'Positive']
ptweets = ptweets['CLEANED_TEXT']

round(ptweets.shape[0] / df.shape[0] *100 , 2)
```

Out[9]: 35.14

35,14 % tweetów wyraża sentyment pozytywny.

```
In [10]: # % tweetów negatywnych
ptweets = df[df.Sentiment == 'Negative']
ptweets = ptweets['CLEANED_TEXT']

round(ptweets.shape[0] / df.shape[0] *100 , 2)
```

Out[10]: 36.94

36,94% tweetów wyraża sentyment negatywny.

```
In [11]: # % tweetów neutralnych
ptweets = df[df.Sentiment == 'Neutral']
ptweets = ptweets['CLEANED_TEXT']

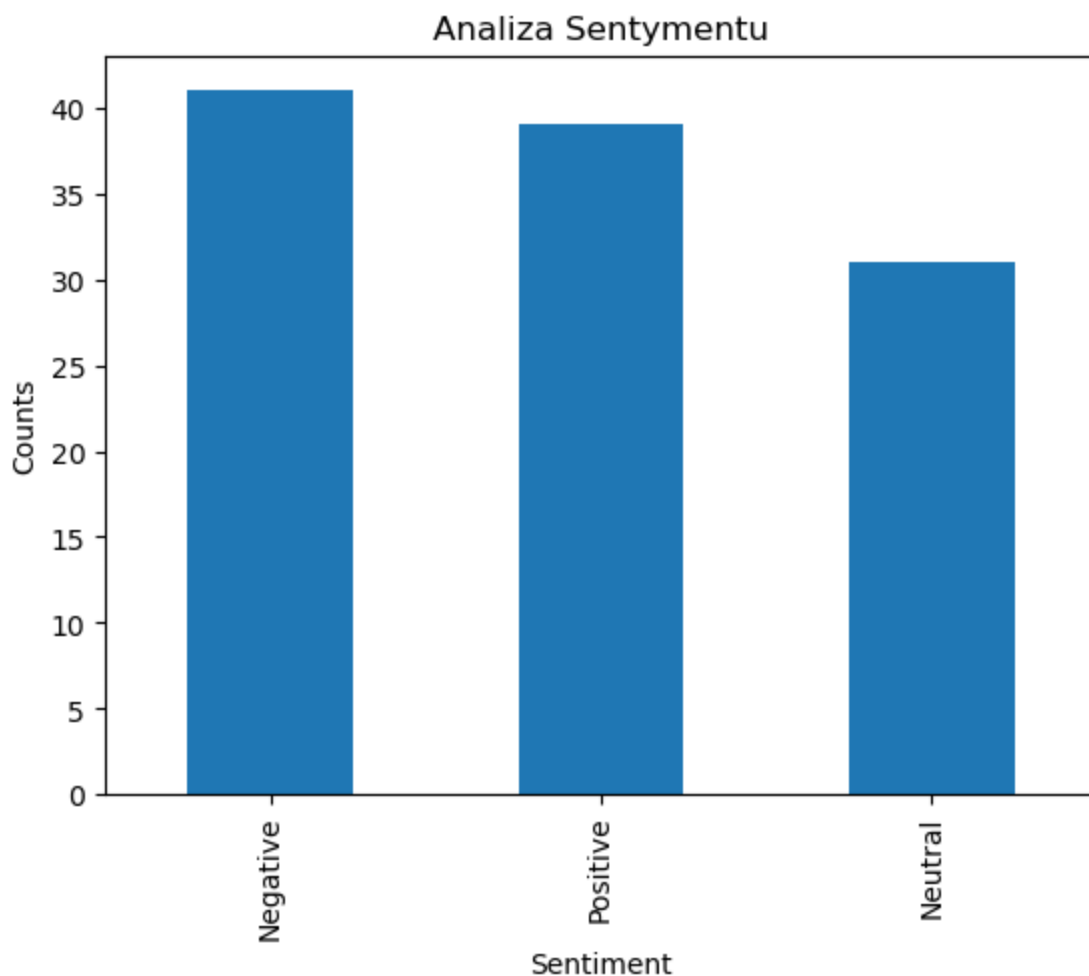
round(ptweets.shape[0] / df.shape[0] *100 , 2)
```

Out[11]: 27.93

27,93% tweetów wyraża sentyment neutralny.

```
In [12]: df['Sentiment'].value_counts()

plt.title("Analiza Sentymentu")
plt.xlabel('Sentiment')
plt.ylabel('Counts')
df['Sentiment'].value_counts().plot(kind='bar')
plt.show()
```



Wykres potwierdza poprzednie analizy:

- 39 (35,14%) pozytywnych tweetów
- 41 (36,94%) negatywnych tweetów
- 31 (27,93%) neutralnych tweetów.

```
In [13]: # Połączenie oczyszczonych tweetów w jednozdaniowy tekst
combined_text = ' '.join(df['CLEANED_TEXT'])

# Analiza sentymentu dla całego tekstu tweetów naraz
sentiment = TextBlob(combined_text).sentiment.polarity

# Wyświetlenie sentymentu
print('Sentiment:', sentiment)
```

Sentiment: -0.019478932178932186

Ogólny sentyment dla całego zbioru jest bardzo nieznacznie negatywny.

```
In [14]: # Poziom subiektywność dla całego tekstu tweetów naraz
subjectivity = TextBlob(combined_text).sentiment.subjectivity

# Wyświetlenie poziomu subiektywności
print('Subjectivity:', subjectivity)
```

Subjectivity: 0.44228388648388617

Poziom subiektywności jest nieznacznie bliższy 0 niż 1 - jest więc trochę więcej faktów niż czystych opinii.

Funkcja czyszcząca tweety i tagująca części mowy

```
In [15]: def clean_tweet(tweet):

    tweet = re.sub(r"http\S+|www\S+|https\S+", '', tweet, flags=re.MULTILINE) # usuwa ws
    tweet = re.sub(r'\@w+|\#', '', tweet) # usuwa wszystkie wystąpienia oznaczeń użytko
    tweet = re.sub(r'^\w\s', '', tweet) # usuwa wszelkie znaki, które nie są literami,
    tweet = re.sub(r'\bRT\b', '', tweet) # usuwa "RT" (Retweet) z tekstu tweeta

    # Tokenizacja tekstu na słowa
    tokens = word_tokenize(tweet)
    # Usunięcie znaków interpunkcyjnych i przekształcenie słów na małe litery
    words = [word.lower() for word in tokens if word.isalpha()]
    # Usunięcie stop words
    words = [word for word in words if word not in stop_words]

    # Lematyzacja + Analiza części mowy dla słów
    lemmas = []
    doc = nlp(' '.join(words))
    for token in doc:
        # Wybieramy tylko rzeczowniki (NOUN), czasowniki (VERB) i przymiotniki (ADJ)
        if token.pos_ in ['NOUN', 'VERB', 'ADJ']:
            lemmas.append(token.lemma_)

    # Połączenie lematów w jednozdaniowy tekst
    cleaned_tweet = ' '.join(lemmas)
    return cleaned_tweet

# Stworzenie nowej kolumny z wyczyszczonymi tweetami
df['CLEANED_TEXT'] = df['TEXT'].apply(clean_tweet)
```

Analiza Częstości

```
In [16]: # Tworzenie listy słów z oczyszczonych tweetów
all_words = ' '.join(df['CLEANED_TEXT']).split()

# Obliczanie częstości słów
word_freq = Counter(all_words)

# Wyświetlanie 100 najczęściej występujących słów
for word, frequency in word_freq.most_common(100):
    print(f'{word}: {frequency}')
```

```
benefit: 39
people: 25
poor: 24
work: 24
child: 23
family: 22
government: 17
increase: 17
program: 15
social: 14
want: 11
poverty: 11
polish: 11
get: 10
take: 9
welfare: 9
leave: 9
support: 9
pole: 8
law: 8
say: 8
year: 7
see: 7
```

make: 7
migrant: 7
high: 7
policy: 7
valorization: 7
look: 6
come: 6
bill: 6
know: 6
pay: 6
citizen: 6
low: 6
care: 6
pension: 6
project: 6
place: 5
need: 5
today: 5
fund: 5
opposition: 5
change: 5
widow: 5
tusk: 4
allowance: 4
demographic: 4
thing: 4
give: 4
much: 4
parent: 4
submit: 4
state: 4
raise: 4
collect: 4
active: 4
show: 4
justice: 4
labor: 4
situation: 4
financial: 4
senior: 4
zloty: 4
part: 3
think: 3
recommend: 3
go: 3
money: 3
let: 3
employment: 3
propose: 3
admit: 3
meet: 3
include: 3
education: 3
minister: 3
pass: 3
security: 3
ukrainian: 3
ask: 3
immigrant: 3
insurance: 3
begin: 3
adopt: 3
draft: 3
provide: 3
statement: 3
read: 3

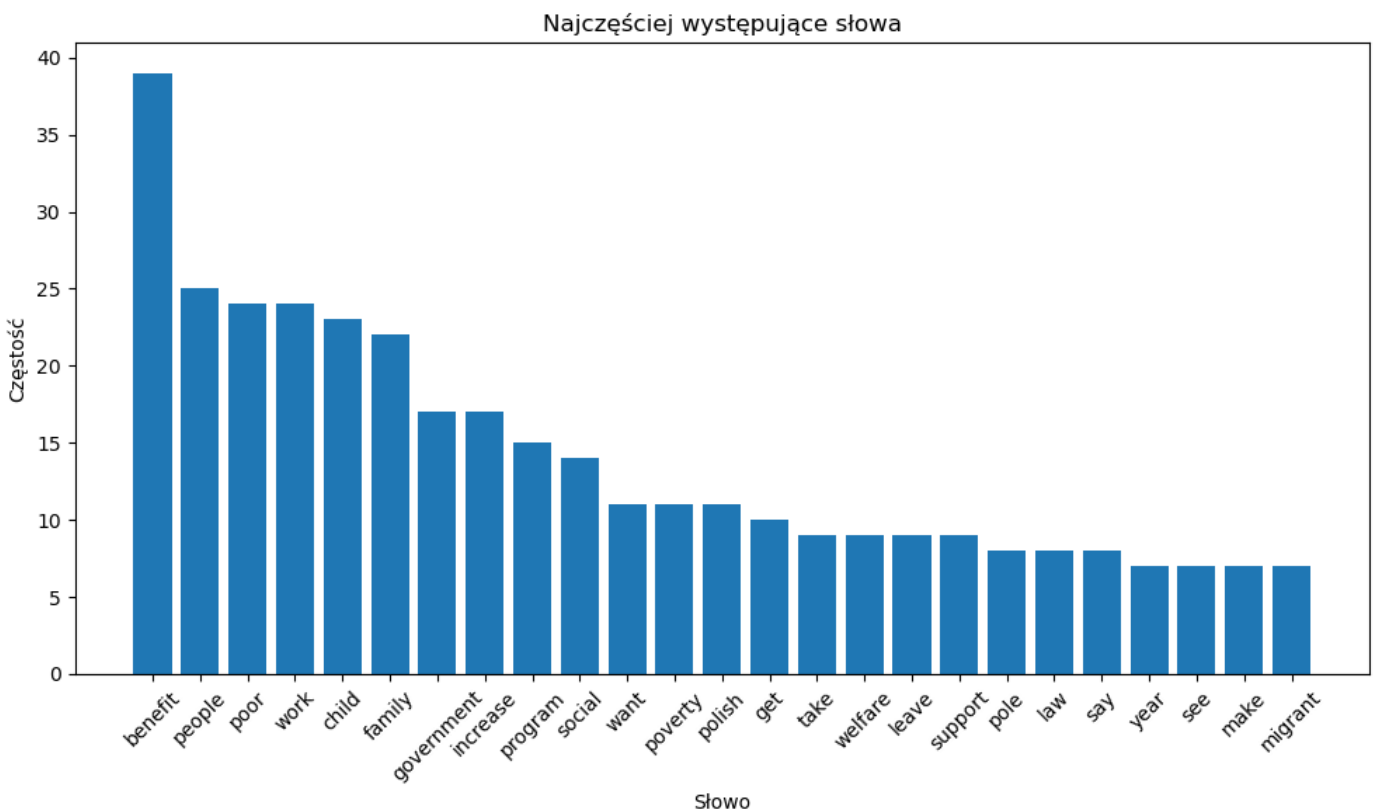

```
new: 3
start: 3
pln: 3
living: 3
datum: 3
rate: 3
disability: 3
amendment: 3
unemployed: 3
mother: 3
illegal: 3
```

```
In [17]: # Wybranie 25 najczęściej występujących słów
top_words = word_freq.most_common(25)

# Przygotowanie danych do wykresu
words = [word for word, _ in top_words]
frequencies = [freq for _, freq in top_words]

# Wygenerowanie wykresu słupkowego
plt.figure(figsize=(10, 6))
plt.bar(words, frequencies)
plt.xlabel('Słowo')
plt.ylabel('Częstość')
plt.title('Najczęściej występujące słowa')
plt.xticks(rotation=45)
plt.tight_layout()

# Wyświetlenie wykresu
plt.show()
```



Jeżeli pominiemy słowa, które w swoich polskich odpowiednikach były używane przy wyszukiwaniu, to najczęściej pojawiającymi się terminami w tym zbiorze są:

- people (ludzie)
- work (praca)
- child (dziecko)

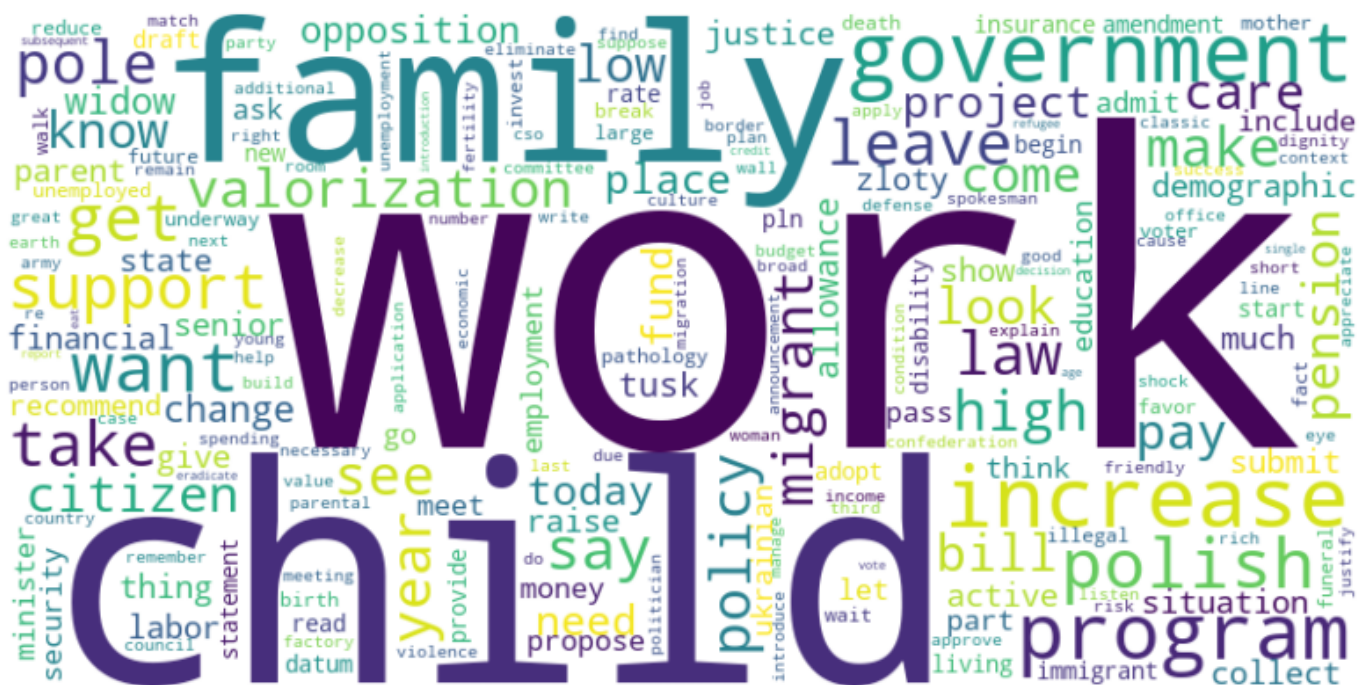
- family (rodzina)
- government (rząd)
- increase (zwiększyć)
- program (program)
- want (chcieć)
- polish (polski)
- get (dostać)
- take (wziąć)
- leave (odejść)
- support (wsparcie)
- pole (polak)
- law (prawo)
- say (powiedzieć)
- year (rok)
- see (widzieć)
- make (zrobić)
- migrant (migrant)

```
In [23]: # Lista słów do usunięcia
words_to_remove = ['benefit', 'poor', 'social', 'poverty', 'welfare', 'people']

# Usunięcie słów z word_freq
word_freq_cleaned = {word: freq for word, freq in word_freq.items() if word not in words_to_remove}

# Generowanie chmury słów z oczyszczonymi częstościami słów
wordcloud = WordCloud(width=800, height=400, background_color='white').generate_from_frequencies(word_freq_cleaned)

# Wyświetlanie chmury słów
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```



W celu poprawienia wizualizacji usunięte zostały słowa, które były używane przy wyszukiwaniu tweetów. Słowami, które wydają się być najbardziej istotne dla niniejszej analizy są:

- work
- child
- family
- program
- government
- increase
- valorization
- migrant

Moja analiza na podstawie powyższych słów:

- Najważniejszym elementem dyskursu jest waloryzacja (valorization) przez rząd (government) programu (program) 500+, a dokładniej zwiększenie (increase) kwoty tego świadczenia do 800 zł.
- Świadczenie jest przedstawiane przez rząd i jego zwolenników jako wsparcie dla rodzin (family), a w szczególności dla dzieci (child).
- Świadczenie ma być również środkiem przeciwdziałania niekorzystnym trendom demograficznym (demographic), poprzez zachęcanie ludzi do posiadania dzieci i zakładania rodzin.
- Kwestia demograficzna może być podnoszona również przez przeciwników polityki rządu, którzy wskazują, że przez cały czas trwania programu 500+, sytuacja demograficzna się nie poprawiła, a dzieci rodzi się mniej.
- Przeciwnicy tego rodzaju świadczeń często podnoszą również postulat, że wypłata zasiłku powinna być warunkowana pracą (work) lub że ludzie powinni sami zapracować na wszystko, co mają.
- Kolejnym tematem może być też kwestia wypłaty świadczeń socjalnych migrantom (migrant) i uchodźcą, którzy przybywają do Polski.
- Migranci i uchodźcy są też prawdopodobnie utożsamiani z biedą i ubóstwem.

Częstość występowania słów w zależności od części mowy

NN - rzeczownik w liczbie pojedynczej

JJ - przymiotnik

NNS - rzeczownik w liczbie mnogiej

VB - czasownik

RB - przysłówki

```
In [19]: # Podział listy częstości słów na podstawie części mowy
pos_freq = {}
for word, freq in word_freq.items():
    pos = nltk.pos_tag([word])[0][1] # Analiza części mowy dla słowa
    if freq >= 3: # Warunek częstotliwości >= 3
        if pos in pos_freq:
            pos_freq[pos][word] = freq
        else:
            pos_freq[pos] = {word: freq}

# Wyświetlanie częstości słów dla każdej części mowy o częstotliwości >= 3
for pos, freq in pos_freq.items():
    print(f'--- {pos} ---')
    sorted_freq = sorted(freq.items(), key=lambda x: x[1], reverse=True) # Sortowanie w
    for word, frequency in sorted_freq:
        print(f'{word}: {frequency}')
    print('\n')
```

--- NN ---

benefit: 39
work: 24
child: 23
family: 22
government: 17
increase: 17
program: 15
want: 11
poverty: 11
welfare: 9
support: 9
pole: 8
law: 8
year: 7
migrant: 7
policy: 7
valorization: 7
look: 6
bill: 6
pay: 6
citizen: 6
care: 6
pension: 6
project: 6
place: 5
need: 5
today: 5
fund: 5
opposition: 5
change: 5
widow: 5
tusk: 4
allowance: 4
thing: 4
parent: 4
submit: 4
state: 4
raise: 4
collect: 4
show: 4
justice: 4
labor: 4
situation: 4
zloty: 4
part: 3
think: 3
recommend: 3
money: 3
employment: 3
propose: 3
admit: 3
meet: 3
include: 3
education: 3
minister: 3
pass: 3
security: 3
ask: 3
immigrant: 3
insurance: 3
begin: 3
adopt: 3
draft: 3
provide: 3
statement: 3
read: 3

start: 3
pln: 3
living: 3
datum: 3
rate: 3
disability: 3
amendment: 3
mother: 3
voter: 3
favor: 3
invest: 3
pathology: 3
wait: 3
future: 3
birth: 3
fact: 3
death: 3
walk: 3
break: 3

--- JJ ---
poor: 24
social: 14
polish: 11
high: 7
low: 6
demographic: 4
much: 4
active: 4
financial: 4
senior: 4
ukrainian: 3
new: 3
unemployed: 3
illegal: 3
large: 3
funeral: 3

--- NNS ---
people: 25

--- VB ---
get: 10
take: 9
leave: 9
say: 8
see: 7
make: 7
come: 6
know: 6
give: 4
go: 3
let: 3
reduce: 3

--- RB ---
underway: 3

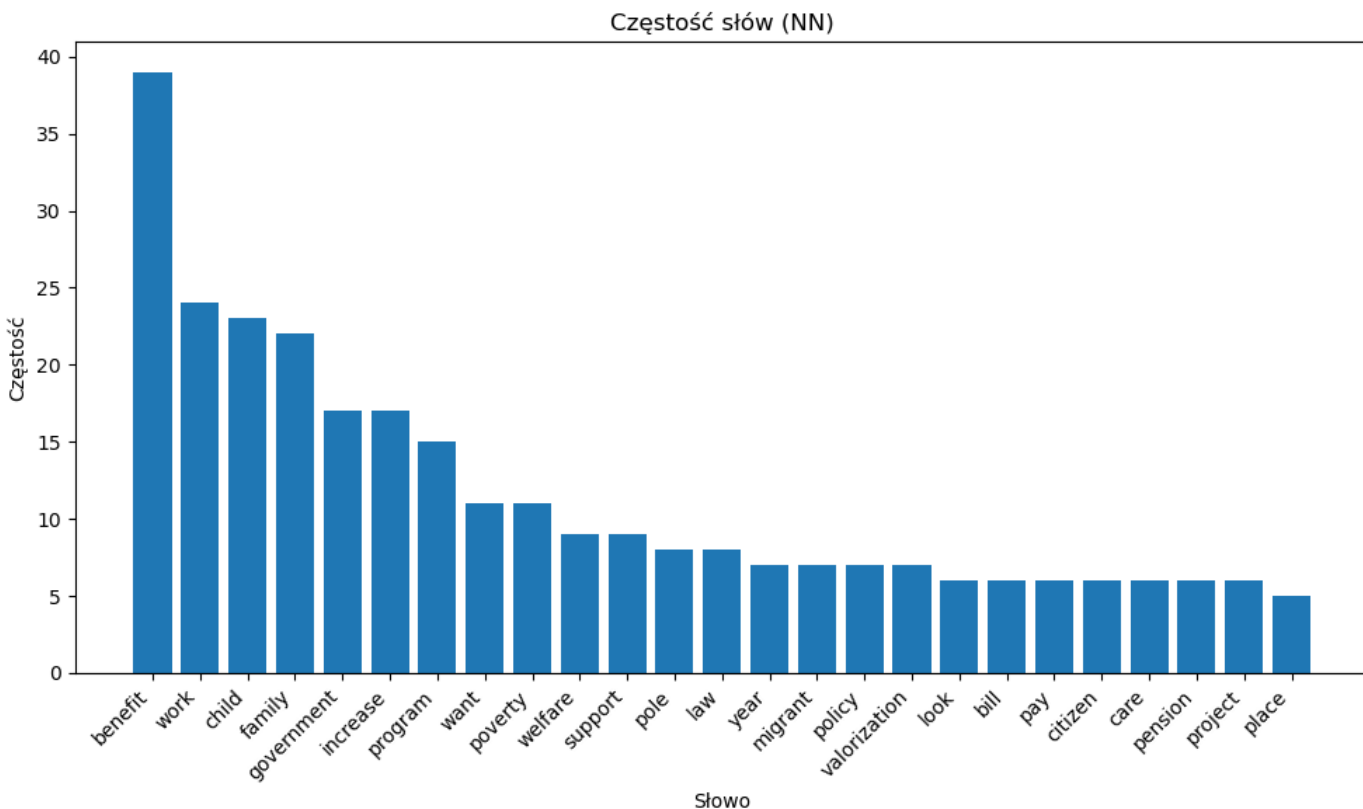
```

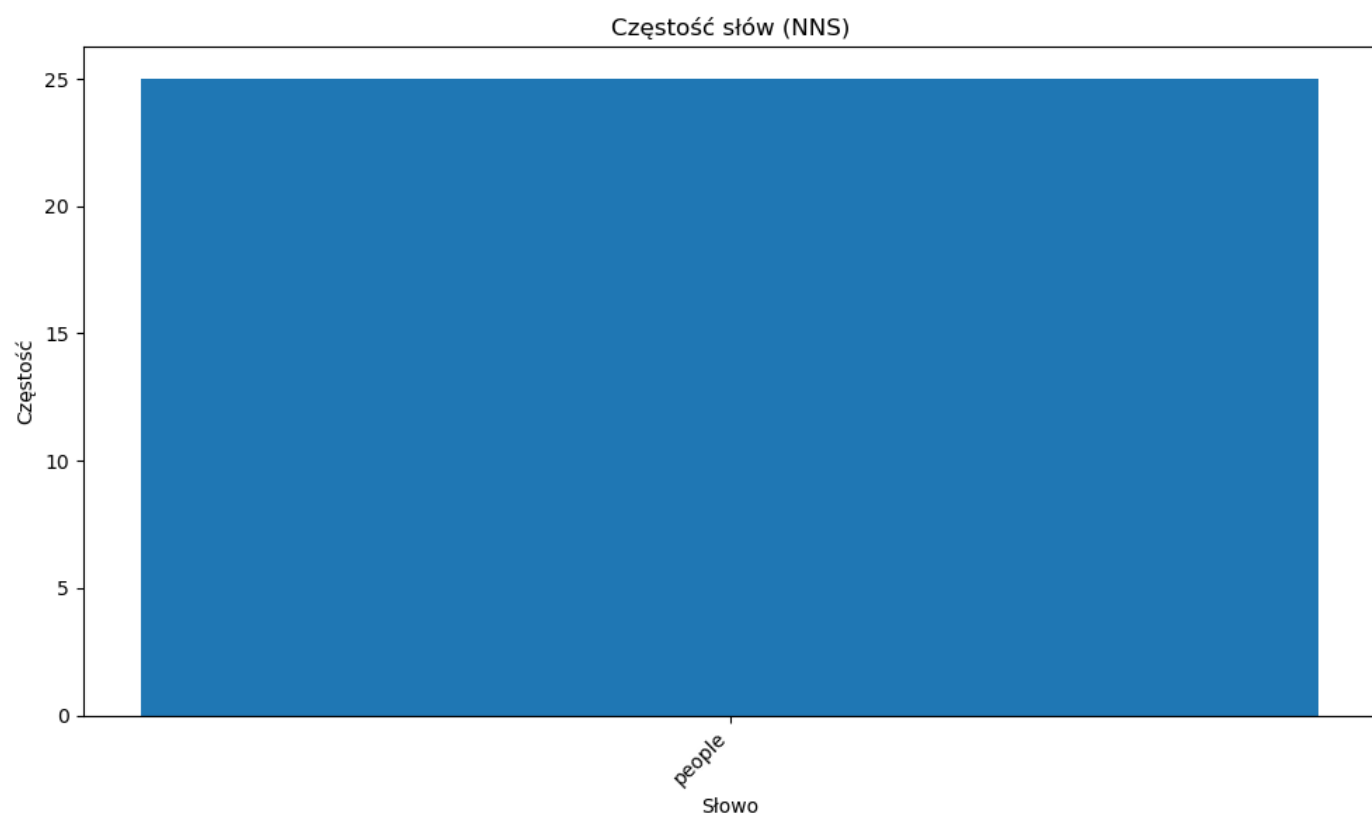
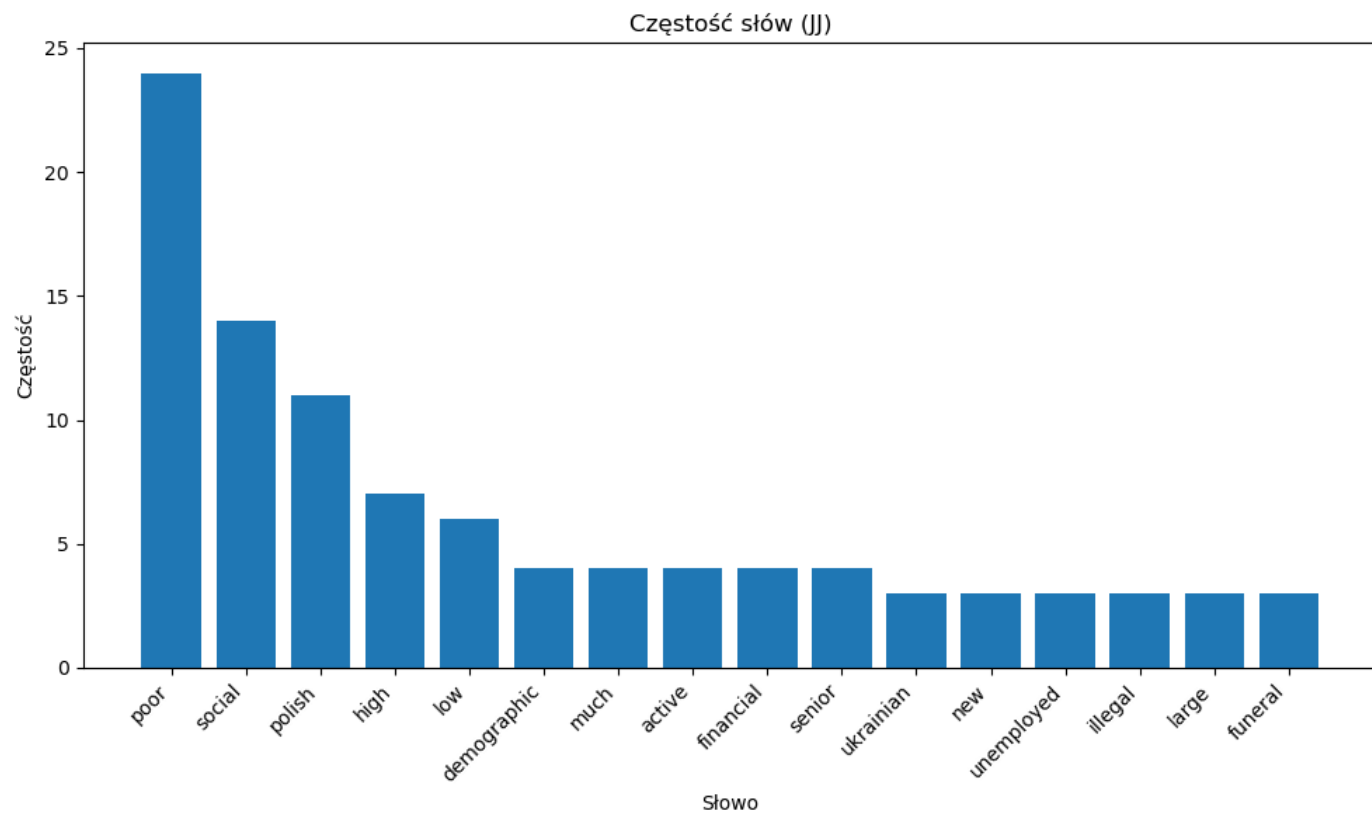
for pos, freq in pos_freq.items():
    sorted_freq = sorted(freq.items(), key=lambda x: x[1], reverse=True) # Sortowanie w
    words = [word for word, _ in sorted_freq[:25]] # Wybór najczęstszych 25 słów
    frequencies = [frequency for _, frequency in sorted_freq[:25]]

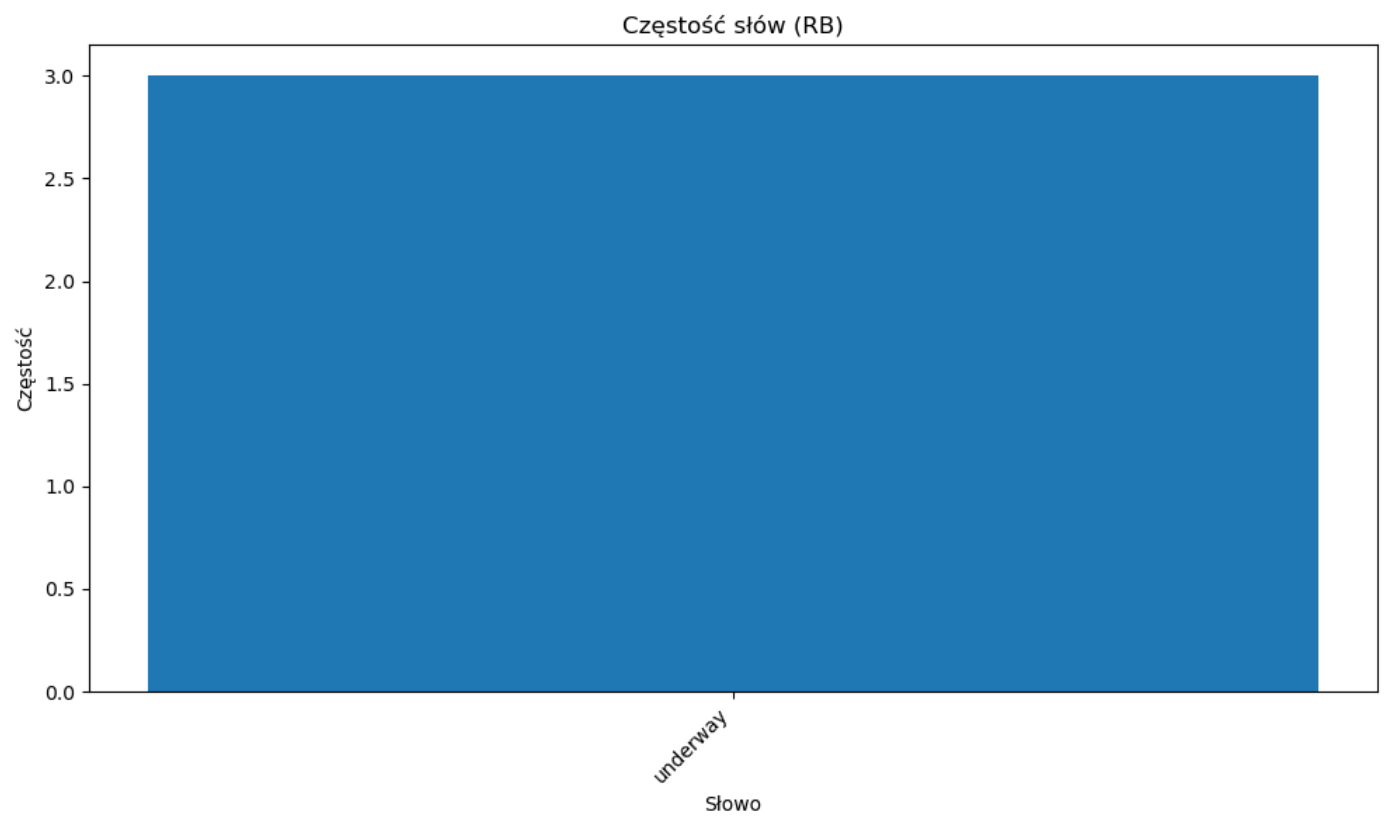
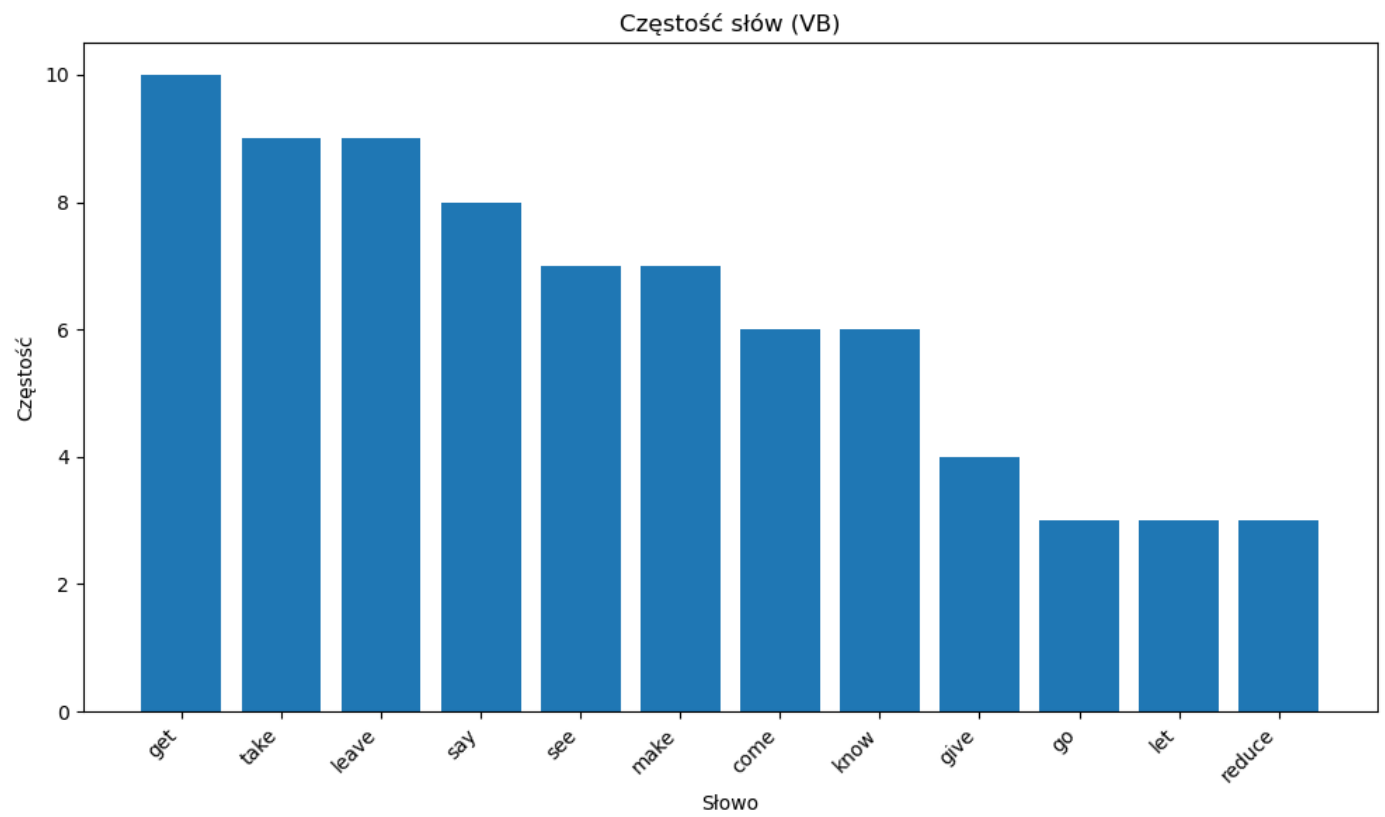
# Wygenerowanie wykresu słupkowego dla danej części mowy
plt.figure(figsize=(10, 6))
plt.bar(words, frequencies)
plt.xlabel('Słowo')
plt.ylabel('Częstość')
plt.title(f'Częstość słów ({pos})')
plt.xticks(rotation=45, ha='right') # Skrócenie etykiet i obrót o 45 stopni
plt.tight_layout()

# Wyświetlenie wykresu
plt.show()

```







Najczęstsze słowa:

rzeczownik w liczbie pojedynczej

- work
- child
- family
- government
- increase
- program

- support
- migrant
- valorization

przymiotnik

- demographic
- senior (senior)
- ukrainian (ukraiński)
- unemployed (bezrobotny)
- illegal (nielegalny)
- funeral (pogrzebowy)

rzeczownik w liczbie mnogiej

- nic znaczącego

czasownik

- nic znaczącego

przysłówek

- nic znaczącego

Do analizy najwięcej wnoszą rzeczowniki w liczbie pojedynczej oraz przymiotniki.

Wymienione tu przymiotniki można zinterpretować w następujący sposób:

- Zwracają uwagę na grupy społeczne, które kojarzą się z pobieraniem świadczeń socjalnych: starsi, ukraińcy, bezrobotni, nielegalni (migranci).
- Kwestia biedy, świadczeń socjalnych i pomocy społecznej jest znowu powiązana z zagadnieniem migracji (ukrainian, illegal).
- Przymiotnik 'funeral' odnosi się najprawdopodobniej do obietnicy podwyższenia zasiłku pogrzebowego przez rząd.

6. Podsumowanie

Na podstawie przeprowadzonej analizy sentymentu można stwierdzić, że opinie na temat osób pobierających zasiłki oraz programów socjalnych są bardzo zróżnicowane (jest bardzo podobna ilość negatywnych ocen jak i pozytywnych). Warto również zauważyć, że znacząca (choć najmniejsza) część tweetów jest neutralna i skupia się na przekazaniu informacji (są to często komunikaty rządowe). Z wyników badania wyłania się raczej obraz społeczeństwa spolaryzowanego pod względem opinii, niż społeczeństwa, które ma negatywny obraz pomocy społecznej oraz które stygmatyzuje świadczeniobiorców. Hipotezę H1, że sentyment do programów socjalnych i świadczeniobiorców jest zdecydowanie negatywny, należy więc odrzucić.

Prawdziwa okazała się natomiast hipoteza H2, że dyskurs toczy się wokół bieżących wydarzeń politycznych (waloryzacja 500+, kwestie migracyjne). Polityka społeczna za sprawą narracji rządowej kojarzy się przede wszystkim ze wspieraniem rodzin, dzieci i poprawianiem sytuacji demograficznej. Z drugiej strony dla wielu wiąże się również z dawaniem ludziom pieniędzy, na które nie zapracowali. Ważnym elementem jest również

spór wokół wypłacania świadczeń socjalnym migrantom i uchodźcą oraz o to, czy są rzeczywiście ludźmi biednymi i potrzebującymi pomocy.