



Horyzonty Polityki
2024, Vol. 15, N° 52



BARBARA KIJEWSKA

<http://orcid.org/0000-0002-5334-0928>

Uniwersytet Gdański

barbara.kijewska@ug.edu.pl

KATARZYNA RACA

<http://orcid.org/0000-0003-1760-1673>

Uniwersytet Gdański

katarzyna.raca@ug.edu.pl

DOI: 10.35765/HP.2572

Identyfikacja i analiza przemocy wobec polskich polityków w postach na Twitterze

Streszczenie

CEL NAUKOWY: Ustalenie występowania i charakterystyka przemocy wobec polityków na platformie społecznościowej Twitter (obecnie X), z uwzględnieniem różnic płci.

PROBLEM I METODY BADAWCZE: Badania nad przemocą polityczną dotyczą głównie fizycznych napaści, skupiając się na konfliktach zbrojnych lub zmianach reżimów. Powszechny dostęp do Internetu zapoczątkował nowy wymiar przemocy objawiającej się nienawistnymi i agresywnymi działaniami często skierowanymi wobec osób publicznych. Do realizacji badania wykorzystane zostały narzędzia *text mining* oraz sztucznej inteligencji.

PROCES WYWODU: W pierwszej części artykułu omówiono problem przemocy wobec polityków, następnie przedstawiono narzędzia zbierania i opracowania danych wraz z uzasadnieniem wyboru modelu analizy. W dalszej części artykułu dokonano analizy zebranego materiału.

WYNIKI ANALIZY NAUKOWEJ: Badanie nie potwierdziło obecności przemocowych treści w postach oznaczonych polityków, ale ujawniło, że prawie jedna czwarta tweetów zawierała pejoratywne treści dyskredytujące i dewaluujące polityków. Chociaż częstotliwość negatywnych komentarzy była podobna zarówno w przypadku polityków płci żeńskiej (24%), jak i męskiej (23%), istniały różnice w ich charakterze. Negatywne uwagi dotyczyły przede wszystkim zdolności intelektualnych polityków, niezależnie od płci.

Sugerowane cytowanie: Kijewska, B., i Raca, K. (2024). Identyfikacja i analiza przemocy wobec polskich polityków w postach na Twitterze. *Horyzonty Polityki*, 15(52), 145–165. DOI: 10.35765/HP.2572.

WNIOSKI, INNOWACJE, REKOMENDACJE: Badanie potwierdza różnice płciowe w pejoratywnych komentarzach związanych z tradycyjnie przypisanymi rolami płciowymi. Obelgi kierowane pod adresem kobiet odnosiły się do ich życia prywatnego, seksualizacji i uprzedmiotowienia, podczas gdy negatywne uwagi pod adresem mężczyzn dotyczyły intencji i sprawczości polityków. Badania potwierdzają brutalizację języka w sferze publicznej, a omawiane wypowiedzi balansują na granicy dopuszczalnej krytyki i zniewagi lub zniewagi. Wyniki badania mogą posłużyć do oceny występujących w mediach społecznościowych emocji i opinii społeczeństwa dotyczących konkretnych osób w partiach rządzących.

SŁOWA KLUCZOWE:

politycy, Polska, Twitter, *text mining*, przemoc polityczna

Abstract

IDENTIFICATION AND ANALYSIS OF VIOLENCE
AGAINST POLISH POLITICIANS IN TWITTER POSTS

RESEARCH OBJECTIVE: To determine the incidence and characteristics of violence against politicians on the social media platform Twitter (now X), considering gender differences.

THE RESEARCH PROBLEM AND METHODS: Research on political violence is mainly concerned with physical assaults and focuses on armed conflict or regime change. Widespread access to the Internet has ushered in a new dimension of violence, manifested in hateful and aggressive actions often directed at public figures. Text mining and artificial intelligence methods were used in this study.

THE PROCESS OF ARGUMENTATION: In the first part of the article, the problem of violence against politicians is discussed, followed by a presentation of the data collection and processing methods, together with the rationale for the choice of the analysis model. The remainder of this article analyzes the collected material and conclusions.

RESEARCH RESULTS: The study did not confirm the presence of violent content in the posts of tagged politicians but revealed that almost a quarter of the tweets contained pejorative content disparaging and devaluing politicians. Although the frequency of negative comments was similar for female (24%) and male (23%) politicians, there were differences in their nature. Negative comments focused on the intellectual capacity of politicians regardless of gender.

CONCLUSIONS, INNOVATIONS, AND RECOMMENDATIONS: This study confirms gender differences in the pejorative comments associated

with traditionally assigned gender roles. Insults directed at women referred to their private lives, sexualization, and objectification, while negative comments directed at men referred to politicians' intentions and agency. The research confirms the brutalization of language in the public sphere, with the utterances in question balancing between acceptable criticism and defamation or insult. The results of the study can be used to assess the emotions and opinions of the public on social media regarding specific individuals in the ruling parties.

KEYWORDS:

politicians, Poland, Twitter, text mining, political violence

WPROWADZENIE

Badania nad przemocą wobec kobiet w polityce są stosunkowo nowym obszarem badań politologicznych, choć samo zjawisko (*violence against women in politics*, VAWIP) istnieje od tysiącleci, wpływając na kobiety wchodzące w przestrzeń publiczną. Badania potwierdzają, że kobiety w polityce stają się obiektem przemocy ze względu na swoją płć i stereotypową rolę, co ma poważne konsekwencje dla demokracji i równości płci (Bardall, Bjarnegård & Piscopo, 2020; Håkansson, 2021; IPU, 2018; Piscopo, 2016; Restrepo Sanín, 2022). W literaturze przedmiotu VAWIP definiuje się jako akty lub groźby przemocy, które skutkują fizyczną, psychologiczną i ekonomiczną krzywdą lub cierpieniem kobiet zaangażowanych w politykę, w tym kandydatek ubiegających się o urząd, ponieważ ich płć postrzegana jest jako zagrożenie dla hegemonicznych norm politycznych (Kuperberg, 2021). Warto zauważyć, że rozumienie przemocy nie ogranicza się do jej fizycznych przejawów, choć słowa takie jak „molestowanie”, „zastraszanie”, „nadużywanie” i „dyskryminacja” mogą być używane w połączeniu z „przecmą”, by zwrócić uwagę na niefizyczne ataki wobec politycznej aktywności kobiet. Badacze używają terminu „przemoc” jako pojęcia zbiorczego, które obejmuje wszystkie formy przymusu. Zdaniem (Krook & Sanín, 2020, s. 744) przemoc semiotyczna „jest popełniana poprzez poniżające obrazy i seksistowski język”, którego celem jest przedstawienie kobiet jako niekompetentnych i niewidzialnych. Przemoc semiotyczna szczególnie często badana jest w Internecie, mimo że jako zjawisko nie ogranicza się tylko do niego (Bardall, Bjarnegård & Piscopo, 2020).

Publiczny i performatywny charakter nadużyć w sieci, choć często skierowany wobec konkretnych polityków, wpływa na poglądy szerszej publiczności i ma na celu delegitimizację kobiet jako aktorów politycznych. Akty przemocy semiotycznej czynią kobiety niekompetentnymi poprzez podkreślanie rzekomej niemożności pogodzenia roli kobiety i przywódczyni. Sprawcy tej formy przemocy uprzedmiotawiają kobiety, podkreślają ich wygląd lub seksualność, opisują je jako złe matki lub osoby, które zawiodły, bo wybrały ambicje polityczne ponad rodzinę. Zatem przemoc semiotyczna wobec kobiet w mediach odnosi się do szkodliwych przedstawień i przekazów, które mogą utrzymywać nierówność płci, wzmacniać stereotypy i przyczyniać się do dyskryminacji i przemocy wobec kobiet. Obejmuje to przedstawianie kobiet jako nadmiernie emocjonalnych, uległych, zależnych od mężczyzn lub skupionych wyłącznie na swoim wyglądzie, uprzedmiotawianie, w tym redukcja ich do wyglądu fizycznego, oraz seksualizację polegającą na ocenie ich atrakcyjności seksualnej. Mężczyźni politycy również mogą być wysmiewani lub wyszydzani na podstawie postrzeganych słabości, w tym wyglądu, zdolności fizycznych, intelektualnych lub ekspresji emocjonalnej, jednak z badań wynika, że kobiety są dwa razy bardziej narażone na seksistowskie nadużycia w Internecie niż mężczyźni i częściej postrzegają przemoc online jako poważny problem w swoim życiu (Duggan, 2017; IPU, 2016, 2018, 2021). Chociaż nie można zaprzeczyć, że politycy płci męskiej również są często atakowani lub obrażani, badania wskazują, że jest to zjawisko zróżnicowane pod względem płci i występuje niezależnie od poziomu rozwoju demokracji, równości płci czy kultury politycznej (Bardall, Bjarnegård & Piscopo, 2020; Druciarek & Niżyńska, 2019; Holm, 2020; Krook & Sanín, 2020, 2016; Sanín, 2022).

Celem niniejszego badania jest ustalenie występowania i charakterystyka przemocy na platformie społecznościowej Twitter. W związku z tym przyjęto trzy założenia: 1) na portalu Twitter występuje przemoc semiotyczna definiowana jako wypowiedzi tekstowe zawierające groźby, uprzedmiotawiające, dehumanizujące oraz mające charakter seksistowski, w tym molestowanie względem polityków; 2) na portalu Twitter znajdują się treści o wydźwięku pejoratywnym, to jest takie, które naruszają zasady szacunku, a więc wypowiedzi obraźliwe, jak przekleństwa i wyzwiska, bądź dyskryminujące;

3) niezależnie od charakteru treści wypowiedzi różnią się ze względu na płeć polityka.

ZBIERANIE I PRZYGOTOWANIE DANYCH

Aby odpowiedzieć na postawione założenia, za pośrednictwem API Twittera zgromadzono 5 046 027 tweetów z okresu 13.10.2019–28.03.2022 (897 dni) z oznaczeniami polskich polityków. Posty były kwalifikowane do badania, gdy występowało w nich imię i nazwisko polityka lub nazwa jego profilu na portalu, jeśli taki istniał (np. Mateusz Morawiecki lub @MorawieckiM). W doborze polityków do analizy kierowano się trzema kryteriami – afiliacją polityczną, pełnioną funkcją i zajmowanym szczeblem władzy (krajowym/samorządowym). Zbiór polityków, o których wypowiedziano się na TT, stanowiło ostatecznie 81 osób, co zostało przedstawione w poniższej tabeli.

Tabela 1. Zbiór polityków, którzy występowali w wiadomościach tekstowych na Twitterze w podziale ze względu na partię oraz płeć

	Koalicja Obywatelska	Prawo i Sprawiedliwość	Inne partie lub komitety	Suma
Kobiety	13	9	15	37
Mężczyźni	13	15	16	44
Suma	26	24	31	81

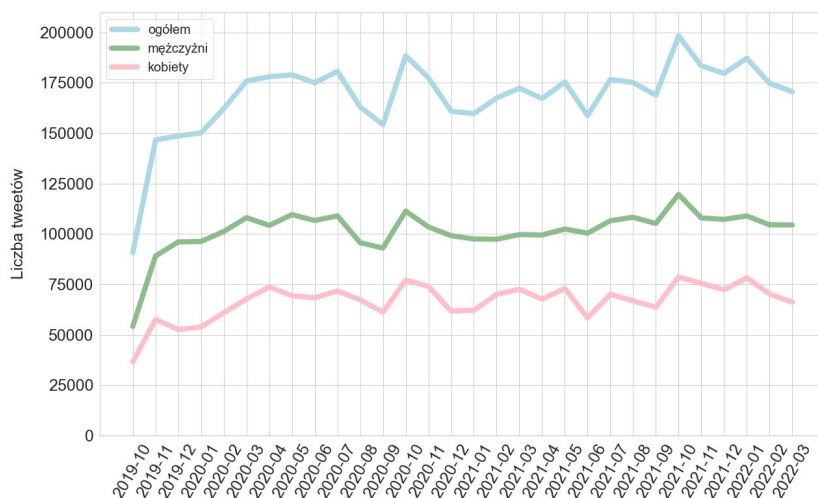
Źródło: opracowanie własne.

Ze względu na ograniczenia API TT z każdego dnia zapisano maksymalnie 1000 tweetów przypadających na jednego polityka. Oznacza to, że poprzez stworzony algorytm możliwe było zgromadzenie 81 000 tweetów dziennie. Mediana liczby dziennych tweetów w uzyskanej bazie danych wyniosła 3649 postów, co jest wartością 22 razy niższą od oczekiwanej przez badaczy. Oznacza to, że zebrane dane odzwierciedlają faktyczną strukturę tweetów odnoszących się do polskich polityków na portalu społecznościowym Twitter.

Gromadzenie danych rozpoczęto w dniu, w którym odbyły się w Polsce wybory parlamentarne w 2019 roku. Przez kilka lat kadencji partii Prawa i Sprawiedliwości doszło do wielu zdarzeń wywołujących emocje, co przekładało się na aktywność użytkowników na

portalu społecznościowym Twitter (wyk. 1). Największa miesięczna liczba postów zebrana została w październiku 2021 roku. Daty z najwyższą dzienną liczbą tweetów odnoszą się do wydarzeń takich jak strajk kobiet (26.10.2020), kryzys na granicy polsko-białoruskiej (10.11.2021), rezygnacja ekspertów z udziału w pracach Rady Medycznej (14.01.2022) czy debata prezydencka pomiędzy Andrzejem Dudą a Rafałem Trzaskowskim (03.07.2020).

Wykres 1. Liczba zebranych postów w analizowanym okresie z uwzględnieniem płci polityków



Źródło: opracowanie własne.

Na podstawie zebranych danych powstały podstawowe statystyki dotyczące najczęściej występujących słów. Wykonanie takiej analizy wymaga uprzedniego przygotowania tekstu (Allahyari et al., 2017). W ramach tego procesu podjęto następujące działania:

1. sprowadzenie słów do małych liter,
2. usunięcie stron internetowych oraz oznaczeń użytkowników,
3. usunięcie znaków interpunkcyjnych,
4. usunięcie znaków graficznych (takich jak emoji, emotikony),
5. usunięcie liczb,
6. usunięcie słów o nieistotnym znaczeniu (np. zaimków, przyimków),
7. tokenizacja, czyli podział tekstu na słowa.

Liczba oraz rodzaj przeprowadzanych operacji w ramach przygotowywania danych tekstowych zależy od celu analizy, ale również od języka. W związku z tym, że zebrane do badania posty napisano w języku polskim, pominięty został etap sprowadzania słów do ich rdzeni (*stemming*) lub podstawowych form (lematyzacja). Ujednolicenie form wyrazów w polszczyźnie sprawia trudności, gdyż niektóre słowa posiadają więcej niż jedną formę podstawową. Na przykład leksem *celi* może oznaczać pomieszczenie lub osiągnięcie rezultatu (zależy to od kontekstu).

Tabela 2. Najczęściej występujące słowa w analizowanej bazie danych

Pojedyncze słowa	Pary słów	Trójki słów
pis	jarosław kaczyński	ha ha ha
panie	andrzej duda	prezes jarosław kaczyński
chyba	panie pośle	pis jarosław kaczyński
kaczyński	panie ministrze	prezydent andrzej duda
proszę	mariusz kamiński	tujestpolska tujestpolska tujestpolska

Źródło: opracowanie własne.

Po przygotowaniu danych wykonane zostały podstawowe statystyki występowania wyrazów. Wśród najczęściej wykorzystywanych pojawiają się wyrażenia odnoszące się do partii Prawo i Sprawiedliwość, która wygrała wybory parlamentarne w 2019 roku. W parach i trójkach słów widoczne są odwołania do prezesa partii oraz prezydenta Rzeczypospolitej Polskiej. W rankingu najpopularniejszych zwrotów i wyrażeń występuje m.in. *proszę* czy *panie ministrze*, co wskazuje na uprzejmość użytkowników Twittera względem polityków.

METODY WYKRYWANIA PRZEMOCY SEMIOTYCZNEJ

Upowszechnienie Internetu przyniosło zmiany w sposobie komunikacji społeczeństwa. Całkowicie zniwelowało barierę geograficzną oraz czasową. Przeciętny polski użytkownik spędza w sieci średnio 6 godzin i 17 minuty dziennie, z czego prawie 2 godziny poświęca na *social media* (We Are Social & Meltwater, 2024). Możliwość uzyskania

szybkiej informacji, jak również chęć utrzymywania bliskich kontaktów ze swoimi znajomymi czy wyrażania własnych opinii to jedne z elementów przyciągających społeczność do sieci (Brandtzæg & Heim, 2009). Innym argumentem w tej kwestii jest możliwość pozostania anonimowym. Anonimowość na portalach społecznościowych wpływa na większą śmiałość w wyrażaniu własnego zdania lub znieważaniu innych ze względu na brak późniejszych konsekwencji (Zhang & Kizilcec, 2014). Sprzyja to powstawaniu wiadomości o charakterze pejoratywnym, w szczególności wobec osób publicznych (Alrehili, 2019). Ich wzrastająca liczba w mediach społecznościowych przyczyniła się do powstania badań naukowych identyfikujących i omawiających to zjawisko (Al-Hassan, Al-Dossari, 2019).

W literaturze naukowej występują propozycje modeli pozwalających na wykrywanie obraźliwych zwrotów; są one wdrażane w różnych obszarach oraz językach. Na przykład w Hiszpańskim Krajowym Biurze Przeciwko Przystępstwu z Nienawiści zostało wdrożone narzędzie HaterNet. Pozwala ono na identyfikację i monitoring mowy nienawiści w języku hiszpańskim na Twitterze (Pereira-Kohatsu et al., 2019). Próby utworzenia modelu koncentrującego się na wykrywaniu mowy nienawiści wobec kobiet w postach na Twitterze podejmowano również, uwzględniając odmienne aspekty kulturowe (Ahluwalia et al., 2018). Jedno z badań naukowych dotyczyło wykrywania cybernienawiści w postach na Twitterze w języku tureckim poprzez wyszukiwanie hashtagów odnoszących się do wyboru odzieży przez kobiety (Sahi, Kilic, Saglam, 2018). Inne badania skupiają się zaś na mizoginicznych i seksistowskich zwrotach (Frenda et al., 2019) bądź orientacji seksualnej i tożsamości płciowej (Arcila-Calderón et al., 2021). Naukowcy analizujący przemoc kierowaną wobec wpływowych kobiet na Twitterze w Indiach zidentyfikowali trzy rodzaje takich komunikatów: lekceważące obelgi, poniżanie etniczno-religijne oraz molestowanie seksualne (Kumar, Gruzd & Mai, 2021).

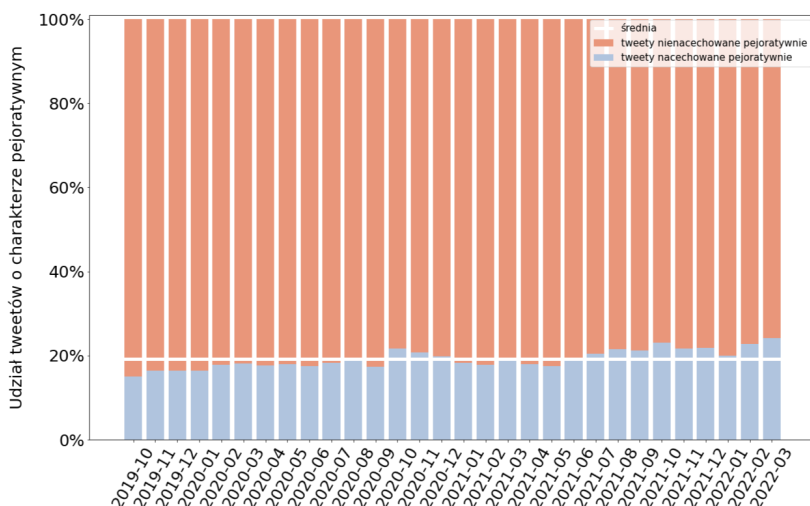
W badaniu użyty został model językowy typu Transformer, który pozwala na rozwiązywanie problemów związanych z przetwarzaniem języka naturalnego takich jak: klasyfikacja zdań lub słów znajdujących się w zdaniu, tworzenie streszczenia, wskazanie odpowiedzi na pytanie, utworzenie tłumaczenia. W zależności od typu zadania wyróżniamy trzy wersje tego modelu: enkoder, dekode-

oraz enkoder-dekoder (Vaswani et al., 2017). Modele tylko z koderem są dobre w przypadku zadań wymagających zrozumienia danych wejściowych, takich jak klasyfikacja zdań do odpowiednich klas, co wykonano w niniejszym artykule. Ogólny, wstępnie wytrenowany model pozwala na statystyczne rozumienie języka naturalnego, a nie praktyczne zastosowanie. Dopiero po dostrojeniu, czyli przy użyciu adnotacji przypisanych przez człowieka, jest on w stanie wykonać zadanie takie jak klasyfikacja (Casola, Lauriola & Lavelli, 2022). Kluczowym elementem w takich modelach jest język. Jeśli analizowane dane są w języku polskim, nieskuteczne byłoby zastosowanie modelu przystosowanego do języka angielskiego. Wynika to z istotnych różnic w systemach gramatycznych tych języków. Z tego względu w niniejszym badaniu uwzględnione zostały polskie odpowiedniki modeli typu enkoder. Wybór podjęto na podstawie Kompleksowej Listy Ewaluacji Językowych (KLEJ), czyli polskiej listy 10 zadań sprawdzających jakość wytrenowanych modeli typu Transformer. Wśród nich znajduje się zadanie CBD – Cyberbullying Detection, którego celem jest wskazanie wiadomości na Twitterze zawierających treści szkodliwe. Uwzględniając najlepszy wynik otrzymany w tym zadaniu oraz moc obliczeniową komputera używanego przez badaczy, wybrano model HerBERT(base). Zastosowano w nim oryginalną architekturę modelu BERT oraz tokenizer Byte-Pair Encoding (Mroczkowski, et al., 2021). Parametry wybranego modelu nie zostały zmienione, lecz przystosowano go do wykrywania w wiadomościach tekstowych przemocy na podstawie: udostępnionego przez KLEJ niezbalansowanego zbioru uczącego (10 041 tweetów) z etykietami (szkodliwe – wulgarne, obsceniczne / nieszkodliwe informacje) oraz samodzielnie etykietowanej próbki własnych danych pobranych w ramach badania (6 000 tweetów). Wymienione powyżej zbiory danych zostały połączone oraz zbalansowane, a następnie wdrożone do ogólnego modelu HerBERT(base). Później model ten został sprawdzony na zbiorze testowym obejmującym 1000 tweetów udostępnionych przez KLEJ ($F1 = 0,65$) oraz własny zbiór 2000 tweetów ($F1 = 0,45$). Uzyskane oceny modelu zbliżone były do wskazanych dla modelu HerBERT(base) w zadaniu CBD umieszczonym w benchmark publikowanym w ramach KLEJ. W związku z tym przygotowany model wdrożony został do pobranych danych odnoszących się do polskich polityków.

IDENTYFIKACJA PRZEMOCY SEMIOTYCZNEJ WOBEC POLITYKÓW

Na podstawie analizy statystycznej par słów w tweetach zidentyfikowanych przez przyjęty model nie zaobserwowano treści przemocowych (przejawów przemocy semiotycznej). Pierwsza hipoteza (H1) o brzmieniu: „Na portalu Twitter występuje przemoc semiotyczna definiowana jako wypowiedzi tekstowe zawierające groźby, dehumanizujące oraz mające charakter seksistowski, w tym molestowanie względem polityków” nie została potwierdzona. W ramach kolejnego etapu przeanalizowano zbiór wygenerowanych przez model par słów. Zastosowany model zidentyfikował 24% tweetów o charakterze pejoratywnym z całego okresu badania. Stanowiły one wypowiedzi obraźliwe, jak przekleństwa i wyzwiska, bądź dyskryminujące. Udział takich wiadomości odchyłał się w niewielkim stopniu w czasie (por. wyk. 2), a jego najwyższa wartość odnotowana została w marcu 2022 roku.

Wykres 2. Udział tweetów o charakterze pejoratywnym w badanym okresie



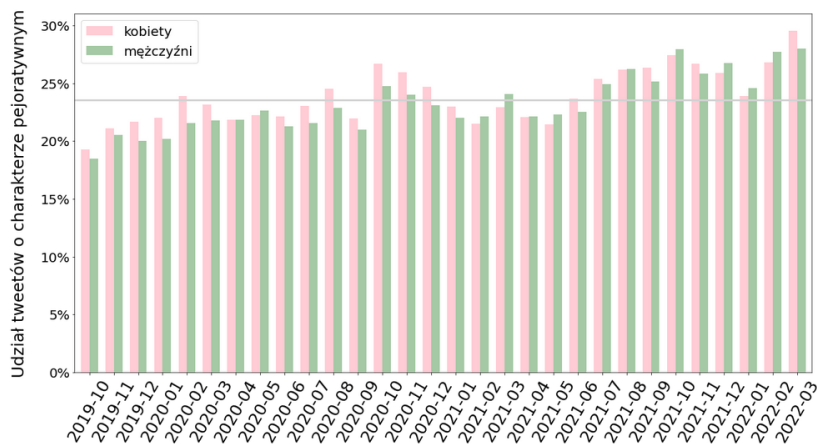
Źródło: opracowanie własne.

Najczęściej pojawiające się pary słów w analizowanym zbiorze danych to: *jarosław kaczyński* (12587), *ha ha* (1 696), *kamieni kupa* (1526),

daj spokój (1426). Natomiast wśród najpopularniejszych trójek słów znalazły się następujące zestawienia: *ha ha ha* (776), *dupa kamieni kupa* (733), *prezes jarosław kaczyński* (438), *duda tchórz duda* (394). Wśród najczęściej występujących wyrażen nie ma leksemów o charakterze pejoratywnym ze względu na to, że model ocenia charakter wiadomości na podstawie treści całego komunikatu. Rzadko zdarza się, żeby komentarz składał się z samych pejoratywnie nacechowanych wyrazów. Prezes partii Prawo i Sprawiedliwość (*jarosław kaczyński*) może występować tak często z uwagi na sposób pobierania oraz przygotowania danych tekstowych. Znajduje się on w niewielkiej grupie posłów nieposiadających konta na Twitterze, co sprawia, że jego posty były wyszukiwane głównie poprzez jego imię i nazwisko zawarte w wiadomościach tekstowych. Przy tworzeniu statystyk usunięte zostały odnośniki do profili polityków na portalu społecznościowym (np. @ZiobroPL), natomiast imiona i nazwiska pozostały.

Wiele spośród wyodrębnionych sformułowań odnosi się do sytuacji politycznych, które miały miejsce kilka(naście) lat wcześniej. Na przykład *dupa kamieni kupa* nawiązuje do wypowiedzi polityka określającej stan państwa ujawniony z nielegalnego podsłuchu rozmów polityków (tzw. afera taśmowa) w partii rządzącej w 2014 r.

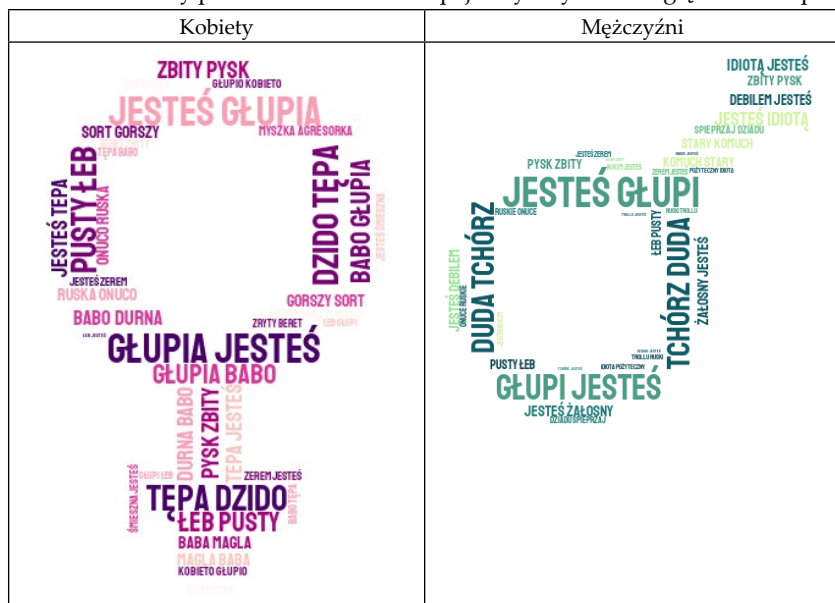
Wykres 3. Udział tweetów o charakterze pejoratywnym w czasie i z uwzględnieniem płci



Źródło: opracowanie własne.

W kolejnym kroku sprawdzono różnicę co do liczby wiadomości o charakterze pejoratywnym, uwzględniając płeć polityków. Zarówno w przypadku kobiet, jak i mężczyzn występują miesiące z większym udziałem takich tweetów (por. wyk. 3). Jednak w całym badanym okresie dla kobiet ten udział jest bliższy 24%, natomiast dla mężczyzn wynosi on 23% wszystkich wiadomości tekstowych w tych grupach. Istnieje niewielka różnica w udziale ze względu na płeć.

Tabela 3. Chmury par słów o charakterze pejoratywnym z uwzględnieniem płci



Źródło: opracowanie własne.

W ramach wyodrębnionych najczęściej występujących par słów z uwzględnieniem płci można zauważyć inwektywy odnoszące się głównie do oceny intelektu polityka. Wskazują na to takie sformułowania jak: *jesteś głupi* (970), *jesteś idiotą* (681), *jesteś debilem* (523), *jesteś głupia* (589), *tepa dzido* (448), *pusty łeb* (420). W związku z tym w niniejszym badaniu uwzględniono poszerzone badanie kategorii pejoratywnych mogących występować w parach słów dla tweetów 4 mężczyzn oraz 4 kobiet o najwyższym udziale pejoratywnych treści na Twitterze.

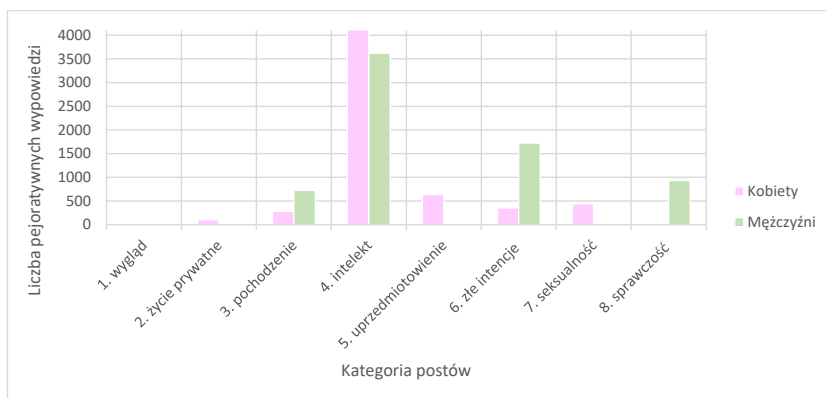
Tabela 4. Kategorie postów wśród mężczyzn i kobiet o najwyższym udziale tweetów pejoratywnych

	Kategorie postów	Liczba kobiet doświadczających pejoratywnych określeń	Liczba mężczyzn doświadczających pejoratywnych określeń
1	wygląd/ciało/ubiór (stereotypowa rola K)	4 z 4	2 z 4
2	życie prywatne/uwagi o życiu bliskich/życie seksualne/orientacja seksualna (stereotypowa rola K)	1 z 4	żaden z 4
3	pochodzenie/zawód/miejsce zamieszkania/etniczność	żadna z 4	3 z 4
4	intelekt/możliwości poznawcze/doświadczenie	4 z 4	4 z 4
5	uprzedmiotowienie/dehumanizacja	4 z 4	4 z 4
6	opisujące złe intencje/merkantylność/cynizm/pazerność/karierowiczostwo/ cechy charakteru, kultury osobistej	4 z 4	4 z 4
7	czynności seksualne/przekleństwa i wyzwiska związane z seksualnością	2 z 4	2 z 4
8	sprawczość – słabość, niedojrzałość (stereotypowa rola M)	żadna z 4	4 z 4

Źródło: opracowanie własne.

W badaniu zidentyfikowane zostały różnice w występowaniu kategorii pejoratywnych tweetów ze względu na płeć dla wybranych polityków (tab. 4). Na podstawie uzyskanych wyników postanowiono obliczyć dodatkowy wskaźnik pozwalający na porównanie natężenia, z jakim występują poszczególne kategorie wiadomości pejoratywnych z uwzględnieniem płci (wyk. 4). W tym celu z najczęściej występujących par słów wyodrębniono sformułowania w pejoratywnych kategoriach: oddzielnie dla kobiet oraz mężczyzn. Przyjęte zostało założenie, że najczęściej występujące pary słów to takie, które występują co najmniej 100 razy w analizowanych tweetach. Następnie zliczono ich występowanie, a uzyskane wyniki zamieszczone zostały na wykresie 4.

Wykres 4. Liczebność pejoratywnych wypowiedzi (par słów) z uwzględnieniem wyodrębnionych kategorii z podziałem na kobiety i mężczyzn



Źródło: opracowanie własne.

Według przeprowadzonej analizy pejoratywne określenia odnoszące się do życia prywatnego (kat. 2, np. *wyrodna matka, matka niepełnosprawnego*), uprzedmiotowienia (kat. 5, np. *żywa gotówka, dno dna*) oraz seksualności (kat. 7, np. *zrób loda, pokaż cycki*) częściej dotyczą kobiet. Mężczyźni bardziej narażeni są na opinie obraźliwe względem pochodzenia (kat. 3, np. *ruska onuca, żydokomunista*), złych intencji (kat. 6, np. *klamca oszust, złodziej oszust*) oraz sprawczości (kat. 8, np. *maminsynkiem, donosicielem; jesteś nieudacznikiem*). Najpowszechniej występująca pejoratywna kategoria odnosiła się do możliwości intelektualnych zarówno kobiet, jak i mężczyzn, przy czym w przypadku kobiet występowała częściej (M – 3621 i K – 4152). Niektóre z wyodrębnionych kategorii postów występują dla obu płci, ale ich częstość występowania jest inna (por. wyk. 3). Pejoratywne określenia odnoszące się do wyglądu nie występują w przypadku obu płci. Oznacza to, że użytkownicy Twittera rzadko w swoich opiniach odnoszą się do wyglądu polityka, gdyż sformułowanie z tej kategorii nie znalazło się w najczęściej występujących parach słów.

DYSKUSJA I WNIOSKI

Przeprowadzone badania nie potwierdziły występowania treści o charakterze przemocowym w postach, w których oznaczeni byli politycy. Ukazały jednak duży udział (prawie jedna czwarta) tweetów z treścią nacechowaną pejoratywnie, których celem jest dyskredytacja i dewaluacja polityka. Choć pejoratywne wypowiedzi kierowane są z podobną częstotliwością zarówno wobec polityczek (24%), jak i polityków (23%), można zauważyć różnice w ich charakterze. Negatywne uwagi w największym stopniu dotyczyły możliwości intelektualnych polityków niezależnie od płci. Badanie potwierdza występowanie różnic w pejoratywnych komentarzach ze względu na płeć adresata. Różnice te wiążą się z tradycyjnie przypisywanymi rolami płciowymi, zgodnie z którymi kobieta odpowiada za życie prywatne, w tym relacje rodzinne czy opiekę nad bliskimi, i jest oceniana pod kątem seksualności. Zatem obelgi, jakimi określano polityczki, odwoływały się do oceny ich życia prywatnego, wynikały z seksualizowania i uprzedmiotowienia mającego wypuklić rzekomą niemożność pogodzenia roli kobiety i przywódczyni. Do mężczyzn kierowano zaś negatywne uwagi odnoszące się do ich intencji oraz sprawczości. Przypisywanie złych intencji rozumianych jako notoryczne kłamanie, karierowiczostwo i wiarołomstwo kontrastuje z modelem mężczyzny jako męża stanu. Zgodnie z tradycyjną rolą społeczną mężczyźni powinni być silni, stabilni emocjonalnie i decyzyjni, natomiast komentarze, których byli adresatami, odnosiły się do ich słabości i labilności.

Badania potwierdzają brutalizację języka w przestrzeni publicznej (Annusewicz, 2017; Balczyńska-Kosman, 2013; Burzyński, 2012). Choć swoboda debaty politycznej stanowi podstawę funkcjonowania demokratycznego społeczeństwa, nie sposób nie zauważyć, że omawiane wypowiedzi leżą na granicy między dopuszczalną krytyką a zniesławieniem czy zniewagą.

Badania z zakresu wykrywania przemocy w social mediach powinny być rozwijane w różnych regionach świata. Różnice kulturowe i językowe sprawiają, że poszczególne elementy publikowanych wiadomości tekstowych w różnych miejscach mogą mieć inne znaczenie. Na przykład w Nigerii mowa nienawiści wyrażana jest przez specjalne symbole liczbowe (Mullah & Zainon, 2021). Aspekt

ten dotyczy również innych obszarów badawczych. W polskim dyskursie politycznym charakterystycznym hasłem jest osiem gwiazdek symbolizujących osiem liter wyrażających w wulgarny sposób niechęć wobec partii sprawującej władzę w latach 2015–2023.

Krytyczne opinie wobec polityków na platformach społecznościowych to jeden z możliwych sposobów na dostrzeżenie siły niezadowolonych obywateli z powodu wprowadzanych przez rząd zmian. W szczególności w czasie przedwyborczych kampanii można go wykorzystać do wprowadzenia zmian wizerunkowych partii czy przygotowania własnych odpowiedzi na niewygodne pytania wyborców. Pejoratywne wypowiedzi na temat działań podejmowanych przez rządzących występowały od zawsze i mogą one prowadzić do pozytywnych zmian. Podobnym zagadnieniem, mogącym jednak wywoływać o wiele więcej negatywnych skutków, jest krytykowanie w ten sposób polityków, a nie ich działań. Zdarza się również, że negatywne nastawienie wobec osoby publicznej przybiera groźniejszą postać nienawiści. Z przeprowadzonego badania nie wynika, aby takie treści na portalu społecznościowym Twitter występowały w postach z oznaczonymi kontami polityków. Autorki badania przypuszczają, że nienawistne wypowiedzi mogą być kierowane wobec konkretnych polityków pod postami na ich profilach. W takich wpisach oznaczenia są niepotrzebne, gdyż użytkownikom automatycznie wyświetlają się odpowiedzi. W badaniu przeprowadzonym w ten sposób dla polityków w Stanach Zjednoczonych zaobserwowano znaczącą różnicę w traktowaniu polityków w social mediach w zależności od ich przynależności partyjnej, płci czy pochodzenia etnicznego (Solovev & Pröllochs, 2022).

Wykonana analiza ma pewne ograniczenia związane z bazą danych treningowych, która została uwzględniona. Po pierwsze dane uwzględnione w badaniu pochodziły z Twittera, jednak większość z nich nie miała charakteru przemocowego, a sam zbiór danych nie był wystarczający. Autorkom badania zależało na wykorzystaniu już istniejącej bazy danych, aby powiększyć przygotowany zbiór danych w celu uczenia algorytmu na jak największej liczbie przykładów. Wybrane przez autorki podejście cechują ograniczenia związane z różnym rozumieniem pojęcia mowy nienawiści przez autorów wykorzystanego zbioru KLEJ. Ponadto autorki miały na celu zbadanie występowania przemocy werbalnej wobec polityków, czego

nie uwzględniał drugi zbiór danych. Kolejne ograniczenie związane ze zgromadzonymi danymi stanowi fakt, że nienawistne posty oraz konta użytkowników je tworzących są usuwane przez administratorów portalu. W związku z tym samodzielne zbieranie danych z portalu społecznościowego Twitter i etykietowanie ich zawartości pod względem występowania przemocy nie jest dobrym sposobem na utworzenie zbioru, który miałby uczyć algorytm wykrywania treści przemocowych wobec polityków. W celu udoskonalenia tego etapu badania można by było skierować do moderatorów portalu społecznościowego prośbę o udostępnienie usuwanych treści lub listy polityków, którzy otrzymują takie komentarze. Innym podejściem pozwalającym na rozwiązanie omawianego problemu byłoby wykorzystanie portalu społecznościowego, na którym nie usuwa się treści przemocowych.

Kolejnym ograniczeniem badania są możliwości obliczeniowe zastosowanego modelu. Bez odpowiedniego sprzętu komputerowego obliczenia za pomocą modelu HerBERT byłyby wykonywane przez kilka dni dla analizowanego przykładu. Jest to spowodowane liczbą parametrów, które są analizowane podczas procesu uczenia się. W związku z ograniczeniami sprzętowymi autorek wybrany został model o mniejszej złożoności. Kolejne badania mogłyby uwzględniać kilka innych modeli oraz baz danych w celu porównania uzyskanych wyników. Warto by było wykonać porównanie pomiędzy metodami słownikowymi oraz statystycznymi.

Pomimo braku rezultatu w postaci modelu pozwalającego na wykrywanie treści przemocowych niniejszy artykuł wypełnia lukę badawczą w obszarze polskojęzycznego dyskursu politycznego w social mediach. Wciąż brakuje badań wdrażających statystyczne modele pozwalające na identyfikację negatywnych zachowań widocznych w treściach publikowanych w sieci. Dotyczy to nie tylko nienawistnych postów, ale również niewiarygodnych informacji mogących skutkować niekorzystną zmianą wizerunku polityków. Zastosowany model pozwolił na wyodrębnienie postów zawierających pejoratywne sformułowania i znacząco ograniczył zbiór danych do 20%. Analiza wszystkich pobranych wiadomości pod względem występowania przemocy na przykład za pomocą stworzonych przez autorki reguł (choćby poprzez wyszukiwanie konkretnych fraz) mogłaby być bardziej czasochłonna i wcale niekoniecznie bardziej dokładna.

Wyniki badania oraz uzyskany model klasyfikacji mogłyby posłużyć do oceny występujących w mediach społecznościowych emocji i opinii społeczeństwa dotyczących konkretnych osób w partiach rządzących. Należy jednak pamiętać, że Twitter to jedna z wielu platform społecznościowych. Obywatele wyrażają swoje przekonania również na portalach informacyjnych czy w ankietach sondażowych. W związku z tym warto rozszerzyć badania w tej kwestii o inne kanały komunikacji.

BIBLIOGRAFIA

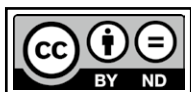
- Ahluwalia, R., Soni, H., Callow, E., Nascimento, A., & De Cock, M. (2018). Detecting Hate Speech Against Women in English Tweets. *CEUR Workshop Proceedings*, 2263. DOI: 10.4000/BOOKS.AACCADEMIA.4698.
- Al-Hassan, A., & Al-Dossari, H. (2019). Detection of Hate Speech in Social Networks: a Survey on Multilingual Corpus. *Conference: 6th International Conference on Computer Science and Information Technology*, 83–100. DOI: 10.5121/csit.2019.90208.
- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E.D., Gutierrez, J.B., & Kochut, K. (2017). A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. *Proceedings of KDD Bigdas, Halifax, Canada*. DOI: 10.48550/arXiv.1707.02919.
- Alrehili, A. (2019). Automatic hate speech detection on social media: A brief survey. *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA, 2019-November*. DOI: 10.1109/AICCSA47632.2019.9035228.
- Annusewicz, O. (2017). Twitter jako przestrzeń autoprezentacji politycznej. Zarządzanie wizerunkiem polityka w mediach społecznościowych. *Studia Politologiczne, Studia i Analizy*, 45, 91–112.
- Arcila-Calderón, C., Amores, J.J., Sánchez-Holgado, P., Blanco-Herero, D., Fersini, E., Arsovski, S., & Hansen, D.L. (2021). Using Shallow and Deep Learning to Automatically Detect Hate Motivated by Gender and Sexual Orientation on Twitter in Spanish. *Multimodal Technologies and Interaction 2021*, 5(10), 63. DOI: 10.3390/MTI5100063.
- Balczyńska-Kosman, A. (2013). Język dyskursu publicznego w polskim systemie politycznym. *Środkowoeuropejskie Studia Polityczne*, 2, 143. DOI: 10.14746/ssp.2013.2.08.
- Bardall, G., Bjarnegård, E., & Piscopo, J.M. (2020). How is Political Violence Gendered? Disentangling Motives, Forms, and Impacts. *Political Studies*, 68(4), 916–935. DOI: 10.1177/0032321719881812.

- Brandtzæg, P.B., & Heim, J. (2009). Why people use social networking sites. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5621 LNCS, 143–152. DOI: 10.1007/978-3-642-02774-1_16/COVER.
- Burzyński, R. (2012). *Metafory jako narzędzie poznania polityki i oddziaływania politycznego*. Praca doktorska napisana pod kierunkiem prof. dr hab. Mirosława Karwata. Pozyskano z: <https://repozytorium.uw.edu.pl/bitstreams/1cb5d8af-2d07-4fd5-a457-e3961b7c6c35/download> (dostęp: 18.09.2024).
- Casola, S., Lauriola, I., & Lavelli, A. (2022). Pre-trained transformers: an empirical comparison. *Machine Learning with Applications*, 9, 100334. DOI: 10.1016/J.MLWA.2022.100334.
- Druciarek, M., & Niżyńska, A. (2019). „To się stało już tak przezroczyste, że o tym zapominam”. *Przemoc wobec kobiet na polskiej scenie politycznej*. Pozyskano z: <https://www.isp.org.pl/pl/publikacje/to-sie-stalo-juz-tak-przezroczyste-ze-o-tym-zapominam-przemoc-wobec-kobiet-na-polskiej-scenie-politycznej> (dostęp: 18.09.2024)
- Duggan, M. (2017). *Online Harassment 2017*. Pozyskano z: <https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/> (dostęp: 18.09.2024).
- Frenda, S., Ghanem, B., Montes-Y-Gómez, M., & Rosso, P. (2019). Online Hate Speech against Women: Automatic Identification of Misogyny and Sexism on Twitter. *Journal of Intelligent & Fuzzy Systems*, 36(5), 4743–4752. DOI: 10.3233/JIFS-179023.
- Håkansson, S. (2021). Do women pay a higher price for power? Gender bias in political violence in Sweden. *Journal of Politics*, 83(2), 515–521. DOI: 10.1086/709838.
- Holm, M. (2020). Violence against women in politics: Emerging perspectives, new challenges. *European Journal of Politics and Gender*, 3(2), 295–297. DOI: 10.1332/251510820X15808264127039.
- IPU (2016). *Sexism, harassment and violence against women parliamentarians*. Issues Brief. Pozyskano z: <https://www.ipu.org/resources/publications/issue-briefs/2021-11/sexism-harassment-and-violence-against-women-in-parliaments-in-africa> (dostęp: 18.09.2024).
- IPU (2018). *Sexism, harassment and violence against women in parliaments in Europe*. Issues Brief. Pozyskano z: <https://www.ipu.org/resources/publications/issue-briefs/2018-10/sexism-harassment-and-violence-against-women-in-parliaments-in-europe> (dostęp: 18.09.2024).
- IPU (2021). *Sexism, harassment and violence against women parliamentarians in Africa*. Issues Brief. Pozyskano z: <https://www.ipu.org/resources/publications/issue-briefs/2021-11/sexism-harassment-and-violence-against-women-in-parliaments-in-africa> (dostęp: 18.09.2024).

- KLEJ [Kompleksowa Lista Ewaluacji Językowych] (b.d.). Pozyskano z: <https://klejbenchmark.com/leaderboard/> (dostęp: 18.09.2024).
- Krook, M.L., & Sanín, J.R. (2020). The Cost of Doing Politics? Analyzing Violence and Harassment against Female Politicians. *Perspectives on Politics*, 18(3), 740–755. DOI: 10.1017/S1537592719001397.
- Krook, M.L., & Sanín, J.R. (2016). Violence against women in politics: A defense of the concept. *Política y Gobierno*, 24(1), 16–39.
- Kuperberg, R. (2021). Incongruous and illegitimate Antisemitic and Islamophobic semiotic violence against women in politics in the United Kingdom. *Journal of Language Aggression and Conflict*, 9(1), 100–126. DOI: 10.1075/jlac.00055.kup.
- Mroczkowski, R., Rybak, P., Wróblewska, A., Gawlik, I. (2021). *HerBERT: Efficiently Pretrained Transformer-based Language Model for Polish*. In *Proceedings of the 8th BSNLP Workshop on Balto-Slavic Natural Language Processing*, 1–10. DOI: 10.48550/arXiv.2105.01735.
- Mullah, N.S., & Zainon, W.M.N.W. (2021). Advances in Machine Learning Algorithms for Hate Speech Detection in Social Media: A Review. *IEEE Access*, 9, 88364–88376. DOI: 10.1109/ACCESS.2021.3089515.
- Pereira-Kohatsu, J.C., Quijano-Sánchez, L., Liberatore, F., & Camacho-Collados, M. (2019). Detecting and Monitoring Hate Speech in Twitter. *Sensors* 2019, 19(21), 4654. DOI: 10.3390/S19214654.
- Piscopo, J.M. (2016). State capacity, criminal justice, and political rights: Rethinking violence against women in politics. *Política y Gobierno*, 23(2), 437–458.
- Restrepo Sanín, J. (2022). Criminalizing Violence against Women in Politics: Innovation, Diffusion, and Transformation. *Politics and Gender*, 18(1), 1–32. DOI: 10.1017/S1743923X20000173.
- Sahi, H., Kilic, Y., & Saglam, R.B. (2018). Automated Detection of Hate Speech towards Woman on Twitter. *UBMK 2018 – 3rd International Conference on Computer Science and Engineering*, 533–536. DOI: 10.1109/UBMK.2018.8566304.
- Sanín, J.R. (2022). Violence against women in politics as an unintended consequence of democratization. *International Feminist Journal of Politics*, 24(1), 16–39. DOI: 10.1080/14616742.2021.2014343.
- Solovey, K., & Pröllochs, N. (2022). Hate Speech in the Political Discourse on Social Media: Disparities Across Parties, Gender, and Ethnicity. *WWW 2022 – Proceedings of the ACM Web Conference 2022*, 3656–3661. DOI: 10.1145/3485447.3512261.
- Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, 30. DOI: 10.48550/arXiv.1706.03762.

We Are Social, & Meltwater (2024). Digital 2024 Global Overview Report. W *Digital 2024 Global Overview Report*. Pozyskano z: <https://data-reportal.com/reports/digital-2024-global-overview-report> (dostęp: 18.09.2024).

Copyright and License



This article is published under the terms of the Creative Commons Attribution – NoDerivs (CC BY- ND 4.0) License <http://creativecommons.org/licenses/by-nd/4.0/>