

# Analiza porównawcza klasycznych kwestionariuszy i metryk pozyskanych z mediów społecznościowych na przykładzie wojny w Ukrainie

## 1. Wprowadzenie

W przeglądzie systematycznym przeprowadzonym przez Reveihac i współpracowników (2022) badającej relacje między danymi pochodzącymi z mediów społecznościowych (DMS) a klasycznymi kwestionariuszami (KK) stwierdzono, że badania mające na celu przewidywanie zachowań (np. zachowań wyborczych) i innych zjawisk (np. giełdy) na podstawie DMS nie osiągnęły jak dotąd konsensusu [1]. W odpowiedzi autorzy podkreślili znaczenie analizy relacji pomiędzy danymi kwestionariuszowymi a tymi zbieranymi z internetu. Dokładne zbadanie tej relacji może zwiększyć nasze zrozumienie, dlaczego niektóre badania pokazują związki między DMS a KK lub rzeczywistymi zachowaniami ludzi, a niektóre nie. Zakres tego problemu nie pozwala jednak na próbę jego rozwiązania za pomocą jednego badania. Obecne badanie jest pewnym krokiem w drodze do tego rozwiązania.

Nieliczna literatura, która próbowała porównać te dwa źródła danych, zajmowała się głównie przewidywaniem wyników wyborów [2], będących rodzajem KK, z wykorzystaniem analizy sentymentu, która posługując się osiągnięciami z zakresu przetwarzania języka naturalnego (*natural language processing*, NLP) stosuje różnorodne metody analizowania tekstu w celu określenia stanu emocjonalnego jego autora i wpływu, jaki ten tekst może wywrzeć na emocje innych osób [3]. Wydaje się, że sukces w tym względzie był bardzo ograniczony, ponieważ DMS wykazywały zdolność predykcyjną tylko w niektórych przypadkach [1].

Decyzja o wyborze wyników wyborów jako przedmiotu zainteresowania w badaniach nad ekwiwalencją wyników pozyskanych poprzez KK i DMS wydaje się być w dotychczasowej literaturze podyktowany głównie z kwestią mocy statystycznej. Bardzo trudno jest zebrać dane pochodzące z KK i charakteryzujące się wystarczającą ziarnistością, tj. wysokim stopniem szczegółowości (np. w odniesieniu do czasu zbierania i przechowywania danych – dane zbierane codziennie są bardziej ziarniste niż te, które zbierane są raz w tygodniu), by dokonać zadowalającego porównania danych kwestionariuszowych z DMS [1]. Chociaż moc statystyczna jest rzeczywiście bardzo ważnym zagadnieniem, szczególnie istotnym z punktu widzenia planowania badania i analizy jego wyników, ograniczanie badań nad zależnością między danymi wyborczymi i DMS nie wyczerpują tematu zależności między danymi pochodzącymi z KK i DMS. Stąd też konieczne jest również rozważenie szerszego zakresu tematów i metryk NLP. Te zagadnienia motywują obecną pracę.

---

<sup>1</sup> hplisiecki@sd.psych.pan.pl, IP PAN; Uniwersytet SWPS.

<sup>2</sup> flakus.marysia@gmail.com, IFIS PAN.

<sup>3</sup> piotr.koc@yahoo.com, IFIS PAN.

<sup>4</sup> artur.pokropek@gmail.com, IFIS PAN.

Nasze badanie jest, zgodnie z naszą najlepszą wiedzą, pierwszą pre-rejestrowaną [4] próbą porównania wyników KK i DMS. Jednocześnie, porównanie to dotyczy danych niezwiązanych z preferencjami wyborczymi, a fenomenem rzadziej analizowanym w tym kontekście: lękiem przed wojną. W badaniu tym staramy się potwierdzić słusność dwóch metod NLP poprzez porównanie ich rezultatów uzyskanych na podstawie reprezentatywnej próbki tekstów i komentarzy z polskojęzycznego Twittera z wynikami klasycznej ankiety przeprowadzonej wielokrotnie na przestrzeni kilku miesięcy na reprezentatywnej grupie Polaków. Lęk przed wojną mierzony był w sytuacji narastającego niepokoju wojennego odczuwanego przez Polaków w odpowiedzi na wojnę w Ukrainie w okresie od marca do października 2022 roku.

## **2. Metody**

### **2.1. Obecne badanie**

Badanie Twittera miało na celu potwierdzenie słusności kilku metod przetwarzania języka naturalnego, poprzez porównanie ich wyników uzyskanych na reprezentatywnej próbie Twittera z wynikami klasycznej ankiety. Interesującym nas konstruktom był niepokój wojenny odczuwany przez Polaków w odpowiedzi na wojnę w Ukrainie w okresie od marca do sierpnia 2022 roku (wtedy zbierane były dane kwestionariuszowe).

### **2.2. Ogólne przesłanki**

Porównywanie metryk NLP pochodzących z Twittera z wynikami klasycznych ankiet jest problematyczne ze względu na skrajne różnice w naturze tych pomiarów. Na przykład, podczas gdy metryka NLP może liczyć liczbę pozytywnych słów w zdaniach, które zawierają słowo "wojna", aby sprawdzić, jak pisarze postrzegają ten konstrukt, klasyczna ankieta będzie wyraźnie zadawać pytania o koncepcję wojny.

Może to skutkować różnicami zarówno jakościowymi, jak i ilościowymi. W szczególności, metryki NLP mogą nie uchwycić niuansów i kontekstu postrzegania konstruktów przez pisarza, ponieważ często opierają się wyłącznie na liczeniu konkretnych słów. Z drugiej strony, klasyczna ankieta skutkuje nienaturalnymi danymi, przy czym uczestnicy często są proszeni o refleksję nad własnym zachowaniem. Dodatkowo, podczas gdy metryki NLP opierają się na dużej ilości danych, próbka może nie być reprezentatywna dla populacji i może być stronnicza. Z drugiej strony, badania ankietowe pozwalają na bardziej kontrolowaną próbę i mogą zwiększyć reprezentatywność wyników. Dlatego też, porównując te dwa rodzaje miar, ważne jest, aby rozważyć ograniczenia i potencjalne źródła stronniczości każdej z metod oraz zastosować podejście wielometodowe, aby uzyskać bardziej kompleksowe zrozumienie badanego konstruktów [1].

Aby zminimalizować te różnice, zdecydowaliśmy się najpierw zmapować relację pomiędzy dwoma typami metryk, poprzez regresję ich na siebie przy użyciu pierwszych 7 fal badania. W tym momencie nastąpiła wstępna rejestracja naszego badania. Następnie przeprowadziliśmy kolejne 2 fale ankiety, aby sprawdzić, czy zauważony związek się utrzyma.

### **2.3. Ankieta**

Aby zmierzyć zmiany w lęku przed wojną w czasie, skonstruowaliśmy skalę opartą na trzech pozycjach. Każda z pozycji jest przedstawiana uczestnikom poprzez zdanie: „W jakim stopniu zgadzasz się z następującymi stwierdzeniami?”. Lista pozycji znajduje się poniżej:

- Wojna między Rosją a Ukrainą może przerodzić się w wojnę światową.
- Wojna między Rosją a Ukrainą może przerodzić się w konflikt nuklearny.
- Wojna między Rosją a Ukrainą może wkrótce rozprzestrzenić się na Polskę.

Uczestnicy proszeni są o użycie pięciopunktowej skali Likerta przy udzielaniu odpowiedzi na te pozycje (1 – zdecydowanie nie, 5 – zdecydowanie tak).

## **2.4. Populacja**

Próba składała się z ponad 1000 uczestników. Uczestnicy mieli od 18 do 96 lat. Na każdym etapie zrekrutowaliśmy niezależną grupę uczestników z wykorzystaniem quota sampling, uwzględniając następujące cechy populacji: płeć, wiek, miejsce zamieszkania. Badanie przeprowadzono między 8 marca a 29 sierpnia 2022 roku, z dziewięcioma punktami pomiarowymi: etapy 1-5 w pięciu kolejnych tygodniach, natomiast etapy 6-9 co pięć tygodni

## **2.5. Analizy z zakresu przetwarzania języka naturalnego (NLP)**

Używając specyficznych słów kluczowych związanych z wojną stworzyliśmy podkorpus zawierający 1955118 tweetów związanych bezpośrednio z wojną w Ukrainie [5]. Używamy tego pod korpusu do analizy publicznego postrzegania wojny i poziomu niepokoju związanego z wojną.

### **2.5.1. Latent Semantic Scaling (LSS)**

LSS jest metodą, która została wykorzystana do stworzenia leksykonów sentymentów do oceny pozytywności tekstów [6]. Zazwyczaj leksykony tworzone są w długim, żmudnym procesie składającym się z dwóch etapów. Pierwszy etap obejmuje wybór wszystkich słów istotnych dla danego wymiaru zainteresowania, np. pozytywny vs. negatywny. Drugi etap wymaga anotacji tych słów w zależności od tego, gdzie znajdują się w tym wymiarze. Oba etapy wymagają zwykle zaangażowania wielu osób i wielu godzin powtarzalnej pracy. LSS omija ten męczący proces, przypisując w sposób automatyczny oceny wymiarowe wszystkim słowom w interesującym nas korpusie. Wystarczy stworzyć dwa relatywnie krótkie leksykony słów, które określają istotę rozróżnienia pomiędzy dwoma biegunami interesującego nas wymiaru (np. pozytywny: „dobry”, „wspaniały”, „wielki”; negatywny: „zły”, „okropny”, „najgorszy”). Słowa te są następnie reprezentowane są numerycznie za pomocą modeli word embeddings, które potrafią oszacować podobieństwo pomiędzy różnymi słowami. Produktem końcowym jest duży leksykon, który pomaga zmierzyć różnice pomiędzy różnymi tekstami wybranymi z korpusu w odniesieniu do interesującego nas konstruktów. Na przykład, w opisanym powyżej przypadku, interesującym nas konstruktami było rozróżnienie między dobrem a złem.

Do analizy niepokoju społecznego związanego z wojną w Ukrainie stworzyliśmy dwa leksykony, pozytywny i negatywny (patrz tab. 1).

Następnie użyliśmy modelu Polish Glove [7] do obliczenia metryki lęku wojennego dla każdego unikalnego słowa w zlematyzowanym korpusie, tworząc leksykon lęku wojennego. Powstały leksykon miał akceptowalną ważność twarzy, klasyfikując słowa takie jak „śmiertelny”, „ostrzeżenie” i „zabijanie” jako związane z lękiem wojennym, a słowa takie jak „elegancja”, „akceptacja” i „xdd” jako niezwiązane z nim.

Tabela 1. Leksykony niepokoju wojennego

Pozytywny Leksykon (związany z niepokojem)	Negatywny leksykon (związany z bezpieczeństwem)
niebezpieczeństwo, strach, tragedia, zgroza, lęk, obawa, atak, zagrożenie, czołg, rakietą, karabin, zabić, niewola, strzał, ludobójstwo, śmierć, ofiara, ryzyko, alarm, niepokój, niestabilność, niepewność, trwoga, panika, konfrontacja, walka, agresja, starcie, dysharmonia, inwazja, napięcie, horror, zniszczyć, bomba, wróg, przeciwnik	bezpieczeństwo, pokój, rozejm, spokój, harmonia, pogodzić, porozumienie, równowaga, normalność, stabilność, ład, stabilizacja, zgoda, ochrona, nadzieja, consensus, współpraca, uspokoić, luz, rozluźnić, stoicyzm, opanowanie, równowaga, sielankowość, sielanka, odprężenie, porządek, bez troska, komfort, cierpliwość

Źródło: [opracowanie własne].

### 2.5.2. Afektywne normy słów

Afektywne normy słów to duże leksykony słów tworzone przez psychologów i wykorzystywane do prowadzenia badań nad emocjami [8]. Były one wykorzystane do określenia ładunku emocjonalnego różnych tekstów [9, 10]. W ramach leksykonu używanego w obecnym badaniu [8] zgromadzono wiele różnych wymiarów psychologicznych. W przypadku naszego badania za najbardziej kluczowe uznaliśmy wymiary walencji (czyli stopnia, w jakim dane słowo jest pozytywne) oraz pobudzenia (czyli stopnia, w jakim dane słowo jest energetyzujące). Zgodnie z istniejącą literaturą przyjęliśmy, że stan lęku charakteryzuje się niskim poziomem walencji i wysokim poziomem pobudzenia [11, 12].

Aby określić ładunek emocjonalny zgromadzonych Tweetów, zastosowaliśmy bazę norm słownych obejmującą 4900 polskich słów [8]. Używając tych norm, oceniliśmy każde słowo w każdym Tweecie, jeśli było ono dostępne w bazie norm słownych. Następnie wyniki zostały uśrednione dwukrotnie: (1) na poziomie Tweetów i (2) na poziomie dni. Drugie uśrednienie zaowocowało skonstruowaniem poziomów lęku dla każdego dnia, podczas którego zbierano Tweety.

### 2.5.3. Frakcja Tweetów związanych z wojną

Aby oszacować popularność tematu wojny wśród użytkowników Twittera, stworzyliśmy czwartą metrykę, dzieląc liczbę Tweetów odnoszących się do wojny każdego dnia przez liczbę wszystkich Tweetów opublikowanych w tych dniach.

### 2.5.4. Tworzenie miar zbiorczych

Ponieważ dane z badania były zbierane przez okres kilku dni, podczas każdej instancji badania, uśredniliśmy liczbę Tweetów zebranych w tych okresach, aby stworzyć zagregowane wskaźniki NLP.

## 2.6. Skala lęku przed wojną

Aby sprawdzić właściwości psychometryczne skali lęku wojennego, zastosowaliśmy confirmacyjną analizę czynnikową, używając estymatora dla wskaźników porządkowych, tj. WLSMV, na próbie zbiorczej. Standaryzowane ładunki czynnikowe dla poszczególnych pozycji wahały się od 0,874 do 0,933. Rzetelność skal była wysoka, na co wskazuje alfa Cronbacha – 0,90.

Aby zapewnić możliwość porównywania średnich latentnych opartych na trzech pozycjach dla każdej okazji pomiarowej, testujemy niezmienniczość pomiarową skali.

Aby porównać średnie latentne, musimy upewnić się, że poziom inwariantności skali jest uzasadniony. Szczegółową procedurę testowania inwariantności pomiaru ze wskaźnikami porządkowymi, którą zastosowaliśmy, można znaleźć w pracach [13, 14]. Do przeprowadzenia testów wykorzystaliśmy pakiet semTools [15]. Do porównania przybliżonych indeksów dopasowania użyliśmy progów zaproponowanych przez Rutkowskiego i innych [16]. Wyniki pokazały, że pomiar jest niezmienniczy, niezależnie od tego, jaką metodę testowania zastosujemy: skalowany test różnic chi-square czy przybliżone indeksy dopasowania. Środki latentne zostały obliczone dla rozwiązania skalarnego, gdzie średnia i wariancja zmiennej latentnej zostały ustawione na 0 i 1, odpowiednio, dla pierwszej okazji pomiarowej.

## 2.7. Transformacja Miar NLP

Metryki NLP wzbogacamy dodatkową metryką zagregowaną w postaci głównego komponentu analizy PCA przeprowadzonych na pozostałych metrykach. Aby połączyć informacje ze wskaźników NLP i ilości Tweetów stosujemy procedurę ważenia. W pierwszym kroku wykorzystujemy informacje o wynikach badania, frakcji Tweetów oraz różnych wskaźnikach NLP zagregowanych do 7 punktów zbierania danych z badania. Szukamy wartości wagi, która minimalizuje błąd średniokwadratowy predykcji lęku przed wojną, mierzonego na podstawie danych ankietowych z wykorzystaniem wskaźnika NLP oraz informacji o frakcji Tweetów o wojnie:

$$survey_{war\ anxiety} = Waga * \exp(frakcja) * nlp_{indicator} + e$$

Gdzie *survey war anxiety* to wynik kwestionariusza, *Waga* to współczynnik regresji modelu bez stałej na 7 punktach danych. *frakcja* to proporcja ilości Tweetów związanych z wojną do wszystkich Tweetów opublikowanych na Polskim Twitterze danego dnia, *nlp\_indicator* to metryka NLP, a *e* to punkt przecięcia regresji. Otrzymana waga to po prostu oszacowany współczynnik z modelu OLS zastosowany do 4 wskaźników NLP.

Tabela 2. Wagi i RMSE dla poszczególnych metryk NLP

Metryka NLP	Waga	RMSE
LSS	2.42054	.12978
Pobudzenie	-.0217751	.22097
Walencja	1.974041	.2072
PCA	4.360449	.17502

Źródło: [opracowanie własne].

W drugim kroku wzięliśmy oszacowane WAGI i zważyliśmy wyniki nlp mnożąc je przez  $WAGE * \exp(frakcja)$ , czyli  $nlp * WAGA * \exp(frakcja)$ , aby uzyskać połączone wyniki. W przyszłym zbieraniu danych będziemy używać wag oszacowanych w tym kroku.

### 3. Wyniki

Testy  $t$  nie wykazały żadnej istotnej różnicy pomiędzy trendami kwestionariusza a metrykami NLP. Dokładne wyniki były następujące: dla LSS  $t = 1,46$ ,  $p = 0,18$ ; dla walencji  $t = 2,93$ ,  $p = 0,019$ ; dla pobudzenia  $t = 0,1$ ,  $p = 0,92$ ; dla PCA  $t = 0,09$ ,  $p = 0,93$ . Po kontroli wielokrotnych porównań za pomocą poprawki Bonferroniego żaden z czterech testów nie przekroczył progu istotności.

W kolejnej analizie porównujemy kierunek i nachylenie liniowych trendów obserwowalnych za pomocą estymacji wykorzystujących DMS i porównujemy je z trendów z KK. Dla każdej z miary policzyliśmy model regresji OLS, w którym zmienna zależna to wskaźnik określający lęk przed wojną a zmienną niezależną jest liczba tygodni (arbitralnie zero oznacza pierwszy tydzień pomiaru). W tabeli XXX przedstawiamy współczynniki kierunkowe (WK) charakteryzujące nachylenie linii trendu, ich błąd standardowy oraz test istotności statystycznej, gdzie hipotezą zerową jest brak trendu ( $WK = 0$ ). Dodatkowo podajemy standaryzowany WK oraz  $R^2$ . Współczynniki kierunkowe modeli opartych na DMS porównywane są z wynikami dla KK. Hipoteza zerowa dla testu porównania zakłada, iż WK (a zatem trendy) dla różnych modeli są sobie równe.

Tabela 3. Porównanie nachylenia trendów (współczynników kierunkowych regresji) między modelami opartymi na danych z twittera i danych kwestionariuszowych

Miara	Wsp. Kierunkowy (WK; beta)	Błąd Standardowy	$P >  t $	Standaryzowany WK	$R^2$	Porównanie WK z twittera vs WK z badań sondażowych	
						$\chi^2(1)$	$P > \chi^2$
LSS	,023	,008	,006	,491	,364	14,71	0,0001
Pobudzenie	,0003	,0000	,000	,677	0,440	61,35	0,0000
Walencja	-,003	,001	,026	-,4056	0,165	31,93	0,0000
PCA	-,037	,007	,000	-,689	0,475	6,67	0,0098
Badanie sondażowe	-,012	,002	,001	-,889	0,7908	—	—

Źródło: [opracowanie własne].

Wyniki LSS i Pobudzenia wskazują niewielki dodatni trend, podczas gdy Walencja, PCA oraz wyniki oparte na Badania sondażowych negatywny (sugerują spadek zaniepokojenia wojną). Porównanie współczynników dla DMS i KK jednoznacznie wskazuje, iż dają one różne konkluzje (możemy odrzucić hipotezę o równym nachyleniu linii trendów). Każdy z trendów opartych o DMS jest istotnie różny od trendu opartego na badaniach sondażowych.

## 4. Dyskusja

Pomimo nieistotnej różnicy w teście  $t$  pomiędzy wynikami metryki i kwestionariusza, twierdzenie o braku różnicy między tymi dwoma źródłami danych jest nieuzasadnione. Wynika to zarówno z niskiej mocy statystycznej badania, jak i istotnych różnic pomiędzy trendami sondażowymi

Wynik ten uwypukla problemy związane z prowadzeniem badań porównawczych pomiędzy nowymi metrykami big data, a starymi klasycznymi miarami. Podczas gdy ile te pierwsze są często obszerne i często przekładają się na istotne wyniki ze względu na swoją wysoką moc statystyczną, te drugie wymagają dużych zasobów do przeprowadzenia. Z tego powodu niezwykle trudno jest porównywać te dwa źródła danych, gdyż moc badania może nie być wystarczająca do wykrycia ewentualnych różnic. W obecnym badaniu jednak pomimo niskiej mocy statystycznej, osiągnięta różnica w trendach jest jednoznaczna.

Wykorzystanie metryk big data, takich jak NLP, w badaniach z zakresu nauk społecznych ma potencjał oferowania cennego wglądu w opinie i zachowania publiczne. Jednak brak badań walidacyjnych dla tych metryk utrudnia ocenę ich wiarygodności i ważności. Jest to szczególnie problematyczne w przypadku metryk NLP, które często opierają się wyłącznie na liczeniu konkretnych słów i mogą nie uchwycić niuansów i kontekstu postrzegania danego konstruktów przez pisarza.

Pomimo istotnych różnic w dwóch zaobserwowanych źródłach danych, możliwości ekstrapolacji wniosków z obecnego badania są ograniczone. Nie tylko wybór konkretnych metod, i narzędzi, ale też specyficzna populacja i język mogły wpłynąć na osiągnięte wyniki. Zadanie walidacji metryk NLP i ustalenia ich porównywalności z klasycznymi miarami ankietowymi jest więc wyzwaniem, które wymaga wspólnego wysiłku dużej liczby naukowców badających wiele różnych metryk na przekroju różnych kultur i języków. Kluczowe jest, aby badacze starannie rozważali ograniczenia i potencjalne źródła stronniczości przy wykorzystywaniu danych z Twittera i metryk NLP w swoich badaniach.

Podsumowując, obecne badanie wykazuje różnice w wynikach metryk LSS [16] i analizy sentymentu metodą słownikową przy użyciu konkretnej bazy danych [4], i wynikach klasycznych kwestionariuszy, na temat lęków populacji Polskiej w stosunku do wojny w Ukrainie. Wyniki rzucają cień niepewności na rzetelność metryk NLP do analizowania postaw ogółu społeczeństwa Polskiego na podstawie danych Twitterowych.

Badania finansowane z kosztów grantu Narodowego Centrum Nauki (NCN) „Laboratorium badawcze cyfrowych nauk społecznych” (SONATA BIS-10, Nr UMO-020/38/E/HS6/00302).

## Literatura

1. Reveilhac M., Steinmetz S., Morselli D., *A systematic literature review of how and whether social media data can complement traditional survey data to study public opinion*, Multimedia Tools and Applications, 81(7), 2022, s. 10107-10142.
2. O'Connor B., Balasubramanian R., Routledge B., Smith N., *From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series*, Proceedings of the International AAAI Conference on Web and Social Media, 4(1), Article 1, 2010.
3. Medhat W., Hassan A., Korashy H., *Sentiment analysis algorithms and applications: A survey*, Ain Shams Engineering Journal, 5(4), 2014, s. 1093-1113.

4. <anonymized>, *Can war anxiety levels be predicted from Twitter data? – The Ukraine War case*, <https://osf.io/>
5. *Twitter API Documentation*, [data dostępu: 01.07.2022].
6. Watanabe K., *Latent Semantic Scaling: A Semisupervised Text Analysis Technique for New Domains and Languages*, *Communication Methods and Measures*, 15(2), 2022, s. 81-102.
7. Dadas S., *Polish NLP resources (1.0)*, <https://github.com/sdadas/polish-nlp-resources> [data dostępu: 01.07.2022]
8. Imbir K.K., *Affective Norms for 4900 Polish Words Reload (ANPW\_R): Assessments for Valence, Arousal, Dominance, Origin, Significance, Concreteness, Imageability and, Age of Acquisition*, *Frontiers in Psychology*, 7, 2016, s. 1081.
9. Hills T.T., Proto E., Sgroi D., Seresinhe C.I., *Historical analysis of national subjective wellbeing using millions of digitized books*, *Nature Human Behaviour*, 3(12), 2019, s. 1271-1275.
10. Oscar N., Fox P.A., Croucher R., Wernick R., Keune J., Hooker K., *Machine learning, sentiment analysis, and tweets: An examination of Alzheimer's disease stigma on Twitter*, *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 72(5), 2017, s. 742-751.
11. Jackson J.C., Watts J., List J.-M., Puryear C., Drabble R., Lindquist K.A., *From Text to Thought: How Analyzing Language Can Advance Psychological Science*, *Perspectives on Psychological Science*, 2021.
12. Russell J.A., *Core affect and the psychological construction of emotion*, *Psychological Review*, 110(1), 2003, s. 145-172.
13. Rutkowski L., Svetina D., *Measurement Invariance in International Surveys: Categorical Indicators and Fit Measure Performance*, *Applied Measurement in Education*, 30(1), 3027, s. 39-51.
14. Wu H., Estabrook R., *Identification of Confirmatory Factor Analysis Models of Different Levels of Invariance for Ordered Categorical Outcomes*, *Psychometrika*, 81(4), 2016, s. 1014-1045.
15. Jorgensen T.D., Pornprasertmanit S., Schoemann A.M., Rosseel Y., Miller P., Quick C., Garnier-Villareal M., Selig J., Boulton A., Preacher K., *Package 'semTools.'*, Website: <<https://cran.r-project.org/web/packages/SemTools/SemTools> [data dostępu: 01.07.2022]
16. Rutkowski L., Liaw Y.L., Svetina D., Rutkowski D., *Multistage testing in heterogeneous populations: Some design and implementation considerations*, *Applied Psychological Measurement*, 46(6), 2022, s. 494-508.

## **Analiza porównawcza klasycznych kwestionariuszy i metryk pozyskanych z mediów społecznościowych na przykładzie wojny w Ukrainie**

### Streszczenie

Badanie wzbogaca literaturę porównawczą klasycznych kwestionariuszy (KK) i metryk pozyskanych z mediów społecznościowych (DMS) porównując trendy w 9-rzutowym badaniu lęku przed wojną w kontekście napaści Rosji na Ukrainę, z metrykami pozyskanymi z Polskiego Twittera. Dane Twitterowe były zbierane pomiędzy marcem, a sierpniem 2022 roku. W tym czasie również, w nierównych odcinkach czasu zebrane zostały dane kwestionariuszowe. Jako metryk użyto metody latent semantic scaling (LSS), i afektywnych norm słownych, w szczególności norm walencji i pobudzenia. LSS polega na stworzeniu leksykonu słów związanych i przeciwnych do danego konceptu za pomocą techniki word embeddings, i obliczenie wysokości danego konceptu w każdym z tekstów, zliczając występujące w nim słowa. Afektywne normy słowne, z kolei są gotowymi leksykonami słów, ocenionych pod względem różnych komponentów emocji. Podobnie do LSS, wynik dla danego tekstu oblicza się uśredniając oceny występujących w nim słów. Analiza wykazała istotne różnice pomiędzy dwoma źródłami danych. Badanie zostało prerejestrowane, co sprawia, że jego wyniki są wartościowe pomimo niskiej mocy statystycznej.

Słowa kluczowe: NLP, SMD, cyfrowe nauki społeczne, analiza sentymentu, walidacja metryk



## **Comparative analysis of classical questionnaires and metrics extracted from social media on the example of the war in Ukraine**

### Abstract

This study enriches scientific literature by comparing classic questionnaires (KK) and metrics extracted from social media (DMS). It compares trends in a 9-item survey of fear of war in the context of Russia's attack on Ukraine, with metrics extracted from Polish Twitter. The Twitter data was collected between March and August 2022. During this time, questionnaire data was collected in uneven segments. The latent semantic scaling (LSS) method, and affective word norms, in particular valence and arousal norms were used as metrics. LSS involves creating a lexicon of words related to and opposed to a given concept using the word embeddings technique and calculating the metric of a given concept in each text by counting the words occurring in it. The affective word norms, in turn, are ready-made lexicons of words, evaluated in terms of various components of emotion. Similar to the LSS, the score for a given text is calculated by averaging the ratings of the words occurring in it. The analysis showed significant differences between the two data sources. The study was preregistered, which makes the results valuable despite their low statistical power.

Keywords: NLP, SMD, computational social science, sentiment analysis, metrics validation