

Kamil Bednarz

Akademia Górniczo-Hutnicza w Krakowie

ORCID: 0009-0005-2210-4825

Dominik Jaśkowiec

Uniwersytet Komisji Edukacji Narodowej w Krakowie

ORCID: 0000-0002-3492-2439

Mateusz Zaręba

Akademia Górniczo-Hutnicza w Krakowie

ORCID: 0000-0002-9663-6593

Sztuczna inteligencja w analizie politycznej: nowe horyzonty, możliwości i wyzwania

Streszczenie

Analizy wykorzystujące techniki sztucznej inteligencji (SI) stają się coraz istotniejszym elementem badań naukowych, co wynika z rosnącej dostępności mocy obliczeniowej oraz wysokiej jakości danych. Dzięki zastosowaniu algorytmów SI możliwe jest wykonywanie trafnych predykcji, zaawansowanej klasyfikacji wieloklasowej oraz analizowanie podobieństw między zbiorami cech o różnorodnych wymiarach, które charakteryzują poszczególne obserwacje danych. W niniejszej publikacji wykorzystano techniki eksploracyjnej analizy danych oraz nienadzorowanego uczenia maszynowego w celu przestrzennego zbadania preferencji wyborczych w Polsce w latach 2019–2024. Wykorzystano dane z wyborów parlamentarnych, samorządowych oraz do Parlamentu Europejskiego. Wykazano występowanie obszarów o wysokiej frekwencji wyborczej, zidentyfikowanych w klastrze 4, obejmującym duże aglomeracje miejskie oraz Polskę centralną, a także obszarów o niskiej frekwencji, zlokalizowanych głównie na terenach Polski wschodniej oraz północno-zachodniej. Pokazano również interesującą zależność pomiędzy wysokością frekwencji a poparciem dla poszczególnych ugrupowań politycznych w dwóch kolejnych cyklach wyborczych. Dodatkowo przeanalizowano etyczne aspekty wykorzystania technik SI w badaniach nad procesami demokratycznymi, zwracając uwagę na przyszłe kierunki rozwoju i stojące przed nimi wyzwania.

Artificial intelligence in political analysis: new horizons, opportunities, and challenges

Abstract

The application of artificial intelligence (AI) techniques is increasingly becoming a crucial component of scientific research, driven by greater computational power and the availability of high-quality data. AI enables precise forecasting, advanced multi-class classification,

and the analysis of similarities across feature sets with varying dimensions that characterize individual data observations. This study employs exploratory data analysis and unsupervised machine learning methods to spatially examine electoral preferences in Poland from 2019 to 2024. The analysis is based on data from parliamentary, local government, and European Parliament elections. High voter turnout was observed in Cluster 4, which encompasses major metropolitan regions and central Poland, while areas with lower turnout were primarily located in eastern and northwestern Poland. Furthermore, the study reveals an intriguing relationship between voter turnout levels and support for specific political groups across two consecutive electoral cycles. Ethical considerations surrounding the application of AI in research on democratic processes were also explored, along with a discussion of future directions and related challenges.

Słowa kluczowe: analiza danych, frekwencja wyborcza, preferencje polityczne, sztuczna inteligencja, wybory

Key words: artificial intelligence; data analysis; elections; electoral turnout; political preferences

Otrzymano: [31.10. 2024]

Received: [31.10.2024]

Zaakceptowano: [10.12.2024]

Accepted: [10.12.2024]

Sugerowana cytacja / Suggested citation: Bednarz, K., Jaśkowiec, D., Zaręba, M. (2024). Sztuczna inteligencja w analizie politycznej: nowe horyzonty, możliwości i wyzwania. *Annales Universitatis Paedagogicae Cracoviensis Studia Geographica*, 21, doi: 10.24917/20845456.21.3

Wprowadzenie

W ostatnich latach obserwowany jest wyraźny wzrost zastosowania technik analizy danych opartych na algorytmach sztucznej inteligencji (SI) w różnych dziedzinach nauki. Pierwsze analizy wpływu mediów cyfrowych oraz tego, jak analiza dużych zbiorów danych może wpływać na postrzeganie społeczne, pochodzą z początku dziesięciolecia XXI wieku. Za symboliczny początek można przyjąć pracę na temat ideologicznej segregacji użytkowników sieci internetowej oraz mediów społecznościowych (Gentzkow, Shapiro, 2011). W 2012 roku ukazała się kolejna publikacja w tym temacie, pokazująca wpływ mediów społecznościowych na mobilizację społeczeństw (Bennett, Segerberg, 2012). Warto nadmienić, że media społecznościowe korzystają z algorytmów rekomendacyjnych treści opartych na sztucznej inteligencji (Arkhipova, Janssen, 2024). W 2018 roku opublikowano pracę, w której przestudiowano wpływ analizy danych na prowadzenie kampanii politycznej w Wielkiej Brytanii. Autorka podkreśliła trzy kluczowe czynniki: kto jest podmiotem korzystającym z danych, z jakich źródeł pochodzą dane oraz jak te dane przekładają się na komunikację polityczną (Dommett, 2019).

Celem badań jest analiza możliwości wykorzystania algorytmów SI i uczenia maszynowego w analizach politycznych oraz przewidywaniu zachowań wyborców na przykładzie Polski. W szczególności przanalizowano wpływ frekwencji wyborczej. Nowatorskim wkładem tej pracy jest zastosowanie tych technik nie tylko do analizy pojedynczych zachowań wyborczych, ale także do czasowo-przestrzennej analizy szeregu głównych aktów wyborczych na przestrzeni zaledwie kilku lat,

w czasie silnej polaryzacji wyborczej w Polsce. Praca łączy statystyczną analizę populacji wyborców z wykorzystaniem technik sztucznej inteligencji, uzupełnioną o komentarz wyników w kontekście specyficznej sytuacji wyborczej w Polsce. Poza analizą deskryptywną przeprowadzono również analizę diagnostyczną wyników. Jest to swoistego rodzaju wstęp do dalszych badań z użyciem badań socjologicznych i SI w celu przeprowadzania analiz preskryptywnej i predyktywnej. Co najważniejsze, prezentowane w pracy techniki SI opatrzone zostały komentarzem dotyczącym etycznego użycia tego rodzaju algorytmów, co jest niezwykle istotnym, lecz często pomijanym aspektem w badaniach.

Sztuczna inteligencja

Sztuczna inteligencja (SI, ang. *artificial intelligence*, AI) przez wiele osób jest nadal postrzegana jako zagadnienie bardzo młode, które dopiero niedawno zyskało powszechne zastosowanie i uznane jako kluczowe dla dalszego rozwoju społeczeństwa (Gerlich, 2023). Jednym z najbardziej znanych narzędzi generatywnej SI jest ChatGPT, chatbot służący do generowania odpowiedzi na wprowadzane przez użytkownika dane. Wydany pod koniec 2022 roku, zdobył popularność na całym świecie (Ray, 2023). Chociaż SI postrzegana jest jako technologia, która zrewolucjonizuje przemysł i codzienne życie, badania nad nią rozpoczęły się już w latach 60. XX wieku, a za jej narodziny uznaje się rok 1956 (Haenlein, Kaplan, 2019). Co ciekawe, wiele metod wykorzystywanych w SI bazuje na teorii, która powstała jeszcze przed erą komputerów. Przykładowo, sztuczne sieci neuronowe, próbujące symulować strukturę i funkcjonowanie biologicznych neuronów w mózgu, zostały po raz pierwszy zdefiniowane jeszcze w trakcie II wojny światowej (Flasiński, 2011).

W ostatnich latach postęp technologiczny i łatwiejszy dostęp do mocy obliczeniowej sprawiły, że SI rozwija się wyjątkowo dynamicznie, oferując praktyczne zastosowania, które były nieosiągalne dekady temu. Dzięki temu SI przestała być jedynie przedmiotem badań naukowych i stała się integralną częścią narzędzi codziennego użytku, zmieniając oblicze takich branż jak medycyna, edukacja, logistyka i wiele innych. Obserwując ten rozwój, łatwo dostrzec, że chociaż SI ma długą historię, jej potencjał zaczyna być w pełni wykorzystywany dopiero w obecnych czasach.

Samo pojęcie sztucznej inteligencji nie jest ściśle określone. Komisja Europejska definiuje je jako: „Systemy (...) zaprojektowane przez ludzi – oprogramowania (i ewentualnie również sprzętu), które, biorąc pod uwagę złożony cel, działają w wymiarze fizycznym lub cyfrowym, postrzegając swoje środowisko poprzez pozyskiwanie danych, interpretując zebrane ustrukturyzowane lub nieustrukturyzowane dane, rozumując na podstawie wiedzy lub przetwarzając informacje, uzyskane z tych danych i decydując o najlepszym działaniu (działaniach), jakie należy podjąć, aby osiągnąć dany cel” (*Czym jest sztuczna inteligencja?*...). Definicja ta jest bardzo szeroka, do SI zalicza się m.in.: przetwarzanie języka naturalnego, systemy eksperckie, przetwarzanie obrazów i mowy, a także planowanie, robotykę, algorytmy oparte na biologii (ewolucji) oraz uczenie maszynowe. Jedną z metod uczenia maszynowego zostanie zaprezentowana w analizach wyborczych w dalszej części artykułu.

Uczenie maszynowe

Uczenie maszynowe (ang. *machine learning*, ML) jest jedną z kluczowych dziedzin sztucznej inteligencji, łączącą informatykę, programowanie oraz statystykę. Jego celem jest tworzenie modeli stanowiących matematyczną reprezentację zależności pomiędzy danymi oraz umożliwiających naukę wzorców i relacji, które mogą być następnie zastosowane do przewidywania lub klasyfikacji nowych informacji. Proces budowy systemu uczenia maszynowego obejmuje kilka etapów, takich jak przygotowanie danych, wybór odpowiedniego modelu, jego trening, czyli naukę wzorców i relacji w danych, oraz testowanie modelu pod kątem jego skuteczności na nowych, wcześniej nieznanach danych.

Można wyróżnić dwa główne typy uczenia maszynowego: uczenie nadzorowane i nienadzorowane. W uczeniu nadzorowanym dane wykorzystywane do treningu modelu posiadają przypisane etykiety, które reprezentują poprawne rozwiązania. Modele tego typu stosuje się do rozwiązywania problemów predykcyjnych (przewidywanie wartości) oraz klasyfikacyjnych (przypisywanie obiektów do odpowiednich kategorii). Przykładowo, przy budowie modelu prognozującego ceny sprzedaży samochodów na podstawie historii transakcji w danym salonie, etykiety będą odpowiadały cenom sprzedaży tych pojazdów. Cechy (zmienne), które algorytm wykorzysta do przewidywania cen, mogą obejmować przebieg, liczbę poprzednich właścicieli, rok produkcji i inne parametry.

W uczeniu nienadzorowanym dane nie posiadają przypisanych etykiet, a zadaniem algorytmu jest samodzielne odkrywanie powiązań między danymi. Modele tego typu są przede wszystkim stosowane w analizie skupień, która obejmuje grupowanie oraz klasteryzację. Zadania te polegają na przypisaniu obiektów (np. klientów) do odpowiednich grup na podstawie podobieństw w danych, takich jak zachowania zakupowe czy preferencje produktów. Trudnością w stosowaniu tej metody jest dobór liczby klastrow, która często musi być podana na wstępie. Wymaga to wiedzy eksperckiej i stosowania technik analizy podobieństw, które mogą wspomóc ten wybór (Zaręba i in., 2023).

Wykorzystanie SI w badaniach wyborczych

Uczenie maszynowe, jako część SI, staje się kluczowym narzędziem w analizach politycznych. Może umożliwić precyzyjniejsze prognozowanie wyników wyborczych i dokładniejsze zrozumienie preferencji wyborców. Dzięki zaawansowanej analizie zbiorów danych, takich jak badania sondażowe, dane demograficzne oraz interakcje w mediach społecznościowych, modele ML są w stanie identyfikować wzorce zachowań i przewidywać czynniki determinujące decyzje wyborcze. Ponadto algorytmy te mogą być wykorzystywane do segmentacji wyborców, co umożliwia lepsze zrozumienie różnorodności preferencji w społeczeństwie (Dommett, 2019). W niniejszej pracy zbadano dodatkowo wpływ zmian w frekwencji wyborczej na poparcie dla partii politycznych w kolejnych cyklach wyborczych, wykorzystując techniki SI. Analizy takie dostarczają cennych informacji na temat długoterminowych trendów i dynamiki politycznej.

Zagrożenia związane z SI

Stosowanie i dalszy rozwój systemów sztucznej inteligencji ma niewątpliwie wiele zalet. Mogą one ułatwić, zoptymalizować, a nawet całkowicie zastąpić wykonywanie pewnych, często żmudnych czynności przez ludzi. SI przynosi również wiele korzyści w takich dziedzinach jak diagnostyka medyczna, bezpieczeństwo i ekologia transportu, czy odnawialne źródła energii (Rattan i in., 2019).

Niestety, niewiele osób zdaje się dostrzegać zagrożenia związane z nieetycznym zastosowaniem i tworzeniem systemów SI. Owe zagrożenia są szczególnie istotne w kontekście kształtowania krajobrazu politycznego w społeczeństwach demokratycznych. Na przykład w procesach rekrutacyjnych model, który opiera się na danych z wcześniejszych procesów zatrudniania, mógłby wykazywać uprzedzenia wobec danej grupy, jeśli w przeszłości dane treningowe były niebilansowane. Taki model nieproporcjonalnie faworyzowałby kandydatów np. ze względu na płeć, co mogłoby doprowadzić do marginalizacji niektórych grup w miejscu pracy. Innym przykładem może być wykorzystanie algorytmów SI w ocenie kredytowej. Jeżeli model trenowany byłby na danych historycznych odzwierciedlających nierówności społeczne, mógłby dyskryminować pewne grupy społeczne, które w przeszłości miały niższe wyniki kredytowe ze względów systemowych, takich jak ograniczony dostęp do edukacji.

Ponadto rozwój SI może naruszać podstawowe prawa obywatelskie. Na przykład powszechne wykorzystanie systemu rozpoznawania twarzy w miejscach publicznych, choć oczekiwane z perspektywy wymiaru sprawiedliwości i służb mundurowych jako pomoc w łapaniu przestępców, może jednocześnie stworzyć bezprecedensową możliwość inwigilacji i zbierania informacji o obywatelach. Innym ryzykiem są obrazy, teksty oraz materiały wideo generowane przez narzędzia SI, które już teraz z łatwością są w stanie tworzyć fałszywe treści trudne do odróżnienia od autentycznych. Tworzenie materiałów kompromitujących może zająć zaledwie kilka minut, a czasami nawet sekund (Pei i in., 2024).

Ważnym aspektem jest także transparentność i wiarygodność systemów SI. Przykładowo, modele bazujące na sztucznych sieciach neuronowych są swego rodzaju „czarną skrzynką”¹ – wagi parametrów wejściowych do sieci są modyfikowane przez odpowiedni aparat matematyczny (Flasiński, 2011). Po okresie uczenia modele te często potrafią osiągać bardzo dobre rezultaty w rozwiązywaniu danego zagadnienia, problemem jednak jest fakt, że człowiek nie jest zawsze w stanie określić, dlaczego sieć tak zareagowała na dane wejściowe.

Zasady etycznego wytwarzania i użytkowania SI

Jedną z firm kładących nacisk na etyczne wytwarzanie systemów sztucznej inteligencji jest Microsoft (Morandín-Ahuerma, 2023). Zdefiniowała ona sześć głównych zasad, które mogą służyć jako wytyczne w zakresie tworzenia i użytkowania SI. Są to:

1 W ujęciu teorii systemów „czarna skrzynka” (ang. *black box*) to model/system, który działa, ale nie jest znana zawartość jego budowy wewnętrznej, przez co nie jest jasny sam proces przetwarzania danych wejściowych na dane wyjściowe.

- **sprawiedliwość** – systemy sztucznej inteligencji powinny traktować wszystkich sprawiedliwie i unikać wpływania na grupy osób w podobnej sytuacji w różny sposób,
- **niezawodność i bezpieczeństwo** – systemy SI muszą działać niezawodnie, bezpiecznie i spójnie zarówno w normalnych, jak i nieoczekiwanych warunkach,
- **prywatność i zabezpieczenia** – w przypadku sztucznej inteligencji kwestie dotyczące zabezpieczeń danych wymagają szczególnej uwagi, ponieważ SI musi mieć dostęp do danych, aby mogła tworzyć celne prognozy i podejmować słuszne decyzje dotyczące ludzi,
- **brak wykluczeń** – technologia musi uwzględniać szeroki zakres ludzkich potrzeb i doświadczeń,
- **przezroczystość** – ważne jest, aby ludzie zrozumieli, w jaki sposób systemy sztucznej inteligencji wyciągają wnioski, gdy są używane do podejmowania decyzji mających wpływ na życie ludzi,
- **odpowiedzialność** – osoby, które projektują i wdrażają systemy sztucznej inteligencji, muszą być odpowiedzialne za sposób działania ich systemów (*Identyfikowanie wytycznych dla odpowiedzialnego używania sztucznej inteligencji...*).

Warto zaznaczyć, że już powstały pierwsze akty prawne dotyczące regulacji wytwarzania i używania systemów sztucznej inteligencji w Unii Europejskiej, a kolejne są już w trakcie opracowywania. Komisja Europejska kategoryzuje systemy SI według poziomów ryzyka na cztery kategorie: stwarzające niedopuszczalne, wysokie, ograniczone lub minimalne ryzyko. Za niedopuszczalnie ryzykowne uznawane jest wszystko, co stanowi wyraźne zagrożenie dla obywateli Unii Europejskiej: od prowadzonej przez rządy oceny punktowej obywateli (tzw. *social scoring*), po zabawki z asystentem głosowym, które zachęcają dzieci do niebezpiecznych zachowań (*Doskonałość i wiarygodność sztucznej inteligencji...*).

Zagrożenia związane z nieetycznym wykorzystaniem SI w analizach wyborczych

Nieetyczne wykorzystanie sztucznej inteligencji w badaniach politycznych niesie ze sobą poważne zagrożenia, takie jak stronnictwo wyników, która może prowadzić do manipulacji opinią publiczną oraz dezinformacji (Jungherr, 2023). Wykorzystanie tendencyjnych danych i algorytmów może dodatkowo zwiększać polaryzację w społeczeństwie, ponieważ modele oparte na danych odzwierciedlających istniejące uprzedzenia wzmocnią podziały między różnymi grupami, co skutkuje bardziej ekstremalnymi poglądami i osłabieniem dialogu społecznego.

Kontekst analizy

Niespodziewanie wysoka frekwencja wyborcza (74,38%) w wyborach parlamentarnych z 2023 roku wywołała wiele dyskusji na temat jej przyczyn i mechanizmów motywowania wyborców do udziału w głosowaniu.

Przyczyny tego zjawiska, tej bezprecedensowej mobilizacji wyborców, którzy dotychczas nie uczestniczyli w procesie wyborczym albo uczestniczyli w nich nieregularnie, są niewątpliwie złożone. Trudno wskazać jeden czynnik odpowiedzialny za to, że 15 października 2023 roku przy urnach pojawiło się najwięcej

wyborców w całej historii Polski po 1989 roku. Najwyższą frekwencję w wyborach do Sejmu odnotowano dotąd w 1989 roku, wyniosła ona wówczas 62,7%. Frekwencja w przedostatnich wyborach w 2019 roku osiągnęła natomiast poziom 61,74%, zbliżając się do tej wartości (Jaworska-Surma, 2023: 3).

Literatura przedmiotu potwierdza, że sprzyjającymi czynnikami uczestnictwa wyborczego są wysoka pozycja zawodowa i powiązane z nią wysokie dochody. Sytuację taką tłumaczy się, po pierwsze, większą troską niższej klasy społecznej o sprawy bytowe własnego gospodarstwa domowego niż całego kraju. Po drugie, osoby zamożniejsze częściej mogą zyskać bądź stracić na zmianie układu władzy, co skłania je do większego zainteresowania sprawami publicznymi oraz większego poczucia sprawstwa.

Kolejnym istotnym czynnikiem warunkującym wyższy poziom aktywności wyborczej jest wykształcenie. Zależność ta ma charakter liniowy – im wyższe wykształcenie, tym większe prawdopodobieństwo czynnego uczestnictwa w procesie wyborczym. Uzasadnieniem takiej sytuacji jest m.in. większe zrozumienie systemu politycznego oraz większa świadomość odpowiedzialności za jego funkcjonowanie.

Związek aktywności wyborczej z wiekiem zauważono już w latach 60. XX wieku. Począwszy od badań B.R. Berelsona, P. Lazarsfelda i W.N. McPhee, jest to jedno z najważniejszych uwarunkowań większego zaangażowania obywatelskiego (Berelson i in., 1954). Niski poziom aktywności wyborczej w młodym wieku potwierdzają liczne badania. Według teorii cyklu życiowego młodzi ludzie, pochłonięci wieloma sprawami, począwszy od nauki, szukania pracy, czy na zakładaniu rodziny skończywszy, mają ograniczony czas na zainteresowanie się polityką (Murawska, 2017: 101).

W dotychczasowej literaturze badania nad polaryzacją wyborczą koncentrowały się przede wszystkim na polaryzacji ideologicznej, ewentualnie opartej na innych kryteriach socjopolitycznych. W literaturze istnieje wiele wskaźników mierzących zróżnicowanie terytorialne wyników komitetów wyborczych, od klasycznych miar statystycznych, takich jak odchylenie standardowe poparcia czy indeks Giniego, po specjalistyczne wskaźniki, takie jak indeksy unarodowienia partii autorstwa Golosova czy Kasuyi i Moeniusa (Golosov, 2010; Kasuya, Moenius, 2008).

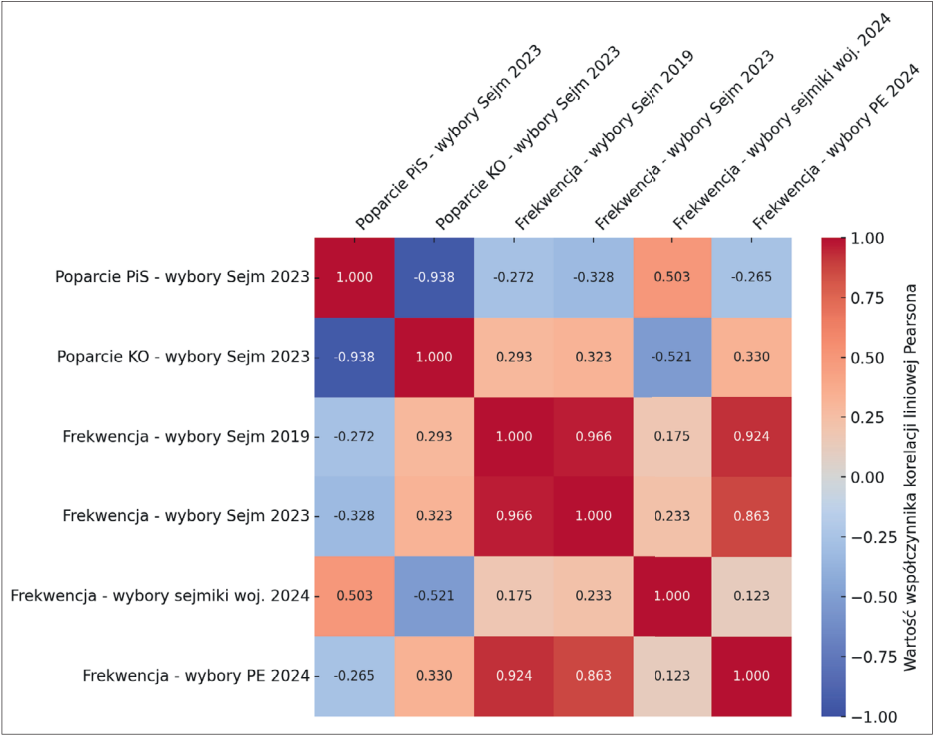
Polaryzacja jest tym szczególnym typem zróżnicowania, polegającym na skupieniu punktów wokół dwóch wyraźnych, lecz przeciwstawnych biegunów (Flis, Stolicki, 2017: 66).

W kontekście polskiej sceny politycznej tymi przeciwstawnymi biegunami będą wyniki wyborcze Koalicji Obywatelskiej oraz Prawa i Sprawiedliwości – dwóch ugrupowań politycznych znajdujących się po przeciwnych stronach konfliktu politycznego w Polsce. Na potrzeby niniejszego artykułu szczegółowej analizie poddana została korelacja wyników wyborczych tych ugrupowań z frekwencją wyborczą, zarówno w skali ogólnokrajowej (wybory z lat 2019–2024), jak i w ujęciu terytorialnym na poziomie poszczególnych powiatów.

Eksploracyjna analiza danych wyborczych

Fundamentalną częścią procesu tworzenia algorytmów SI jest przeprowadzenie eksploracyjnej analizy dostępnych zbiorów danych (ang. *Exploratory Data Analysis*, EDA) (Zaręba i in., 2019). Głównym celem tego procesu jest zrozumienie

struktury danych, weryfikacja ich poprawności, identyfikacja potencjalnych wartości odstających oraz analiza zależności pomiędzy zmiennymi. Kluczowym narzędziem w EDA są różnorodne wizualizacje, które umożliwiają głębsze zrozumienie charakterystyki danych oraz wzajemnych relacji pomiędzy atrybutami. Jednym z przykładów jest macierz korelacji, czyli tabela, która pokazuje współczynniki korelacji Pearsona, Tau Kendalla czy Spearmana między wieloma zmiennymi (Waskom, 2021). Współczynniki te pokazują zarówno siłę, jak i kierunek powiązania pomiędzy dwoma atrybutami. Zakres wartości mieści się w przedziale od -1 do 1, gdzie 1 oznacza silną dodatnią korelację (gdy jedna zmienna rośnie, druga również); 0 brak korelacji (zmienne są niezależne); a -1 silną ujemną korelację (gdy jedna zmienna rośnie, druga maleje).



Ryc. 1. Macierz korelacji frekwencji wyborczej w wybranych wyborach i poparcia dla komitetów KW Prawo i Sprawiedliwość (PiS) oraz KKW Koalicja Obywatelska (PO). N IPL Zieloni (KO) uzyskanego w wyborach do Sejmu RP w 2023 roku na poziomie powiatów

Źródło: opracowanie własne na podstawie danych Państwowej Komisji Wyborczej (PKW)

Poddając analizie macierz korelacji przedstawioną na rycinie 1, należy zwrócić uwagę na wartość współczynnika wyrażającego zależność pomiędzy frekwencją wyborczą z lat 2019, 2023 i 2024 a wynikiem wyborczym Prawa i Sprawiedliwości

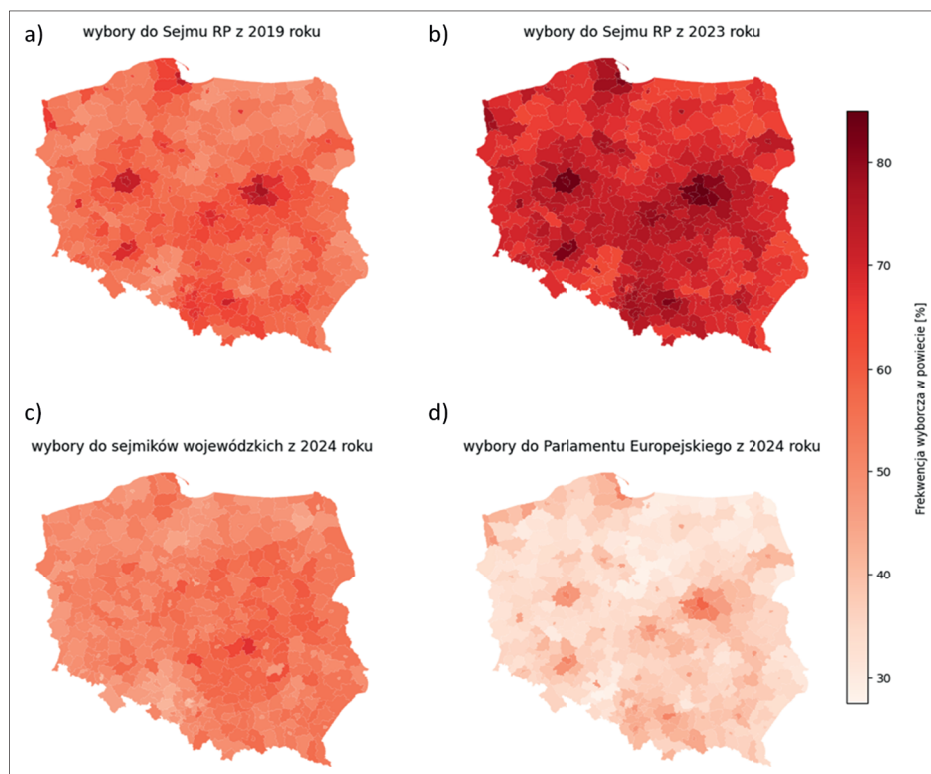
uzyskanym w danym powiecie w wyborach do Sejmu RP w 2023 roku. Współczynnik ten ma wartość lekko ujemną i kształtuje się pomiędzy $-0,328$ a $-0,265$.

W przypadku poparcia dla Koalicji Obywatelskiej korelacja jest dodatnia, co oznacza, że wyższemu wynikowi wyborczemu tego ugrupowania w danym powiecie towarzyszyła wyższa frekwencja wyborcza.

Jednak w wyborach do sejmików wojewódzkich przeprowadzonych w 2024 roku ów trend się zmienił. W powiatach, w których Prawo i Sprawiedliwość w 2023 roku zdobyło większy odsetek głosów, frekwencja wyborcza osiągnęła wyższe wartości. Jak należy interpretować to zjawisko?

Zasadniczo wyższy trend mobilizacyjny wyborców sprzyja ugrupowaniom opozycyjnym, a nie rządzącym. Z kolei mniejsza frekwencja wyborcza jest korzystniejsza dla partii rządzących, ponieważ zmniejsza udział w procesie wyborczym tych grup wyborców, którzy są niezadowoleni ze sprawowanych rządów.

Posiadając dane dotyczące frekwencji wyborczej z lat 2019–2024 oraz współrzędne geograficzne granic powiatów, warto zwizualizować zmiany frekwencji wyborczej na przestrzeni tych wyborów, tak aby w łatwy sposób można było dostrzec prawidłowości i trendy na poziomie regionalnym i ogólnokrajowym (ryc. 2).



Ryc. 2. Frekwencja wyborcza w wybranych wyborach w poszczególnych powiatach

Źródło: opracowanie własne na podstawie danych Państwowej Komisji Wyborczej (PKW)

Analiza z wykorzystaniem metod uczenia nienadzorowanego

Analizując zmiany dotyczące frekwencji (ryc. 2) oraz preferencji wyborczych Polaków w podziale na powiaty, zauważono wyraźny charakter regionalny tych zjawisk. W celu przeanalizowania podobieństwa między obszarami kraju zastosowano metodę nienadzorowanego uczenia maszynowego, wykorzystując algorytm k-średnich. Schemat działania tego algorytmu jest następujący:

1. Wybiera się liczbę grup k .
2. Losowo wyznacza się k punktów, które będą środkami tych grup.
3. Każdy punkt danych jest przypisywany do najbliższego środka grupy.
4. Środki grup są przesuwane na podstawie średniej punktów przypisanych do każdej z grup.
5. Kroki 3 i 4 powtarza się, aż środki grup przestaną się zmieniać.

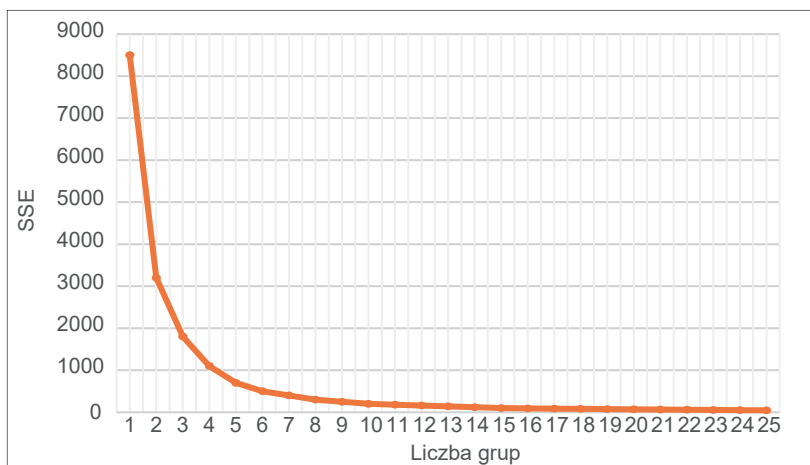
Celem algorytmu k-średnich jest to, aby punkty danych wewnątrz każdego klastra były jak najbardziej zbliżone do siebie, a klastry względem siebie odseparowane. Metoda ta jest szybka oraz pozwala na ogół uzyskać dobrze odseparowane klastry. Opiera się w głównej mierze na obliczaniu dystansów pomiędzy poszczególnymi elementami analizowanej zbiorowości. Dzięki temu, że metoda ta jest stosunkowo wysoko stabilna i niezawodna, stała się uniwersalną i wykorzystywaną w różnych zagadnieniach jako punkt odniesienia w analizowaniu skupień (Ahmed i in., 2020).

Pewną trudnością w stosowaniu tego rozwiązania jest fakt, że liczba klastrów musi być arbitralnie określona przed rozpoczęciem pracy algorytmu. Jedną z metod wspomagających dobór optymalnej liczby grup jest metoda łokcia (ang. *elbow method*), która polega na wielokrotnym przeprowadzaniu klasteryzacji danych przy różnych liczbach klastrów i ocenie jakości uzyskanych grup (Umargono i in., 2020). Ocena ta opiera się na mierze, która wskazuje, jak blisko siebie znajdują się dane w ramach klastrów – zazwyczaj jest to suma kwadratów odchyłeń (SSE, ang. *Sum of Squared Errors*), która mierzy odległości danych od środka swojego klastra; im mniejsza wartość SSE, tym lepiej dane są zgrupowane, co oznacza, że grupy są bardziej jednorodne (Shi i in., 2021).

Na wykresie punktowym lub liniowym przedstawiającym wyniki tej miary dla różnej liczby klastrów poszukuje się punktu, w którym dalsze zwiększanie liczby klastrów przestaje przynosić znaczącą poprawę – wygląda to jak „łokieć”. Punkt ten sugeruje optymalną liczbę klastrów do użycia. Metoda łokcia pomaga znaleźć równowagę między zbyt małą a zbyt dużą liczbą klastrów, co umożliwia lepsze zrozumienie struktury danych (Shi i in., 2021).

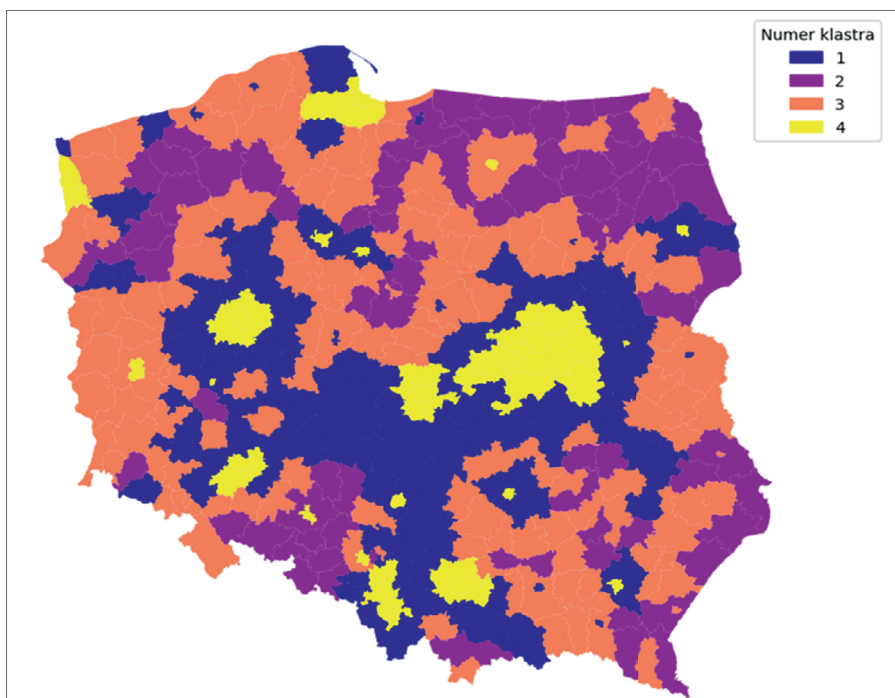
Rycina 3 przedstawia wyniki zastosowania metody łokcia na danych dotyczących frekwencji wyborczej w każdym powiecie w wyborach do Sejmu RP w 2023 roku. W wielu przypadkach wybór punktu przegięcia pozostaje subiektywny, tak też i tutaj za taki punkt można uznać liczbę klastrów od 3 do 5 (ryc. 3). Ostatecznie do klasteryzacji powiatów na podstawie tych danych przy użyciu algorytmu k-średnich zastosowano 4 klastry.

Klasteryzacja wyników wyborów do Sejmu RP z roku 2023 (ryc. 4) dostarcza nowych zmiennych w zakresie obszarowego kształtowania się frekwencji wyborczej. W klastrze 4 znajdują się powiaty o największej frekwencji wyborczej, wynoszącej $\pm 77\%$. Są to największe polskie miasta wraz z otaczającymi je powiatami



Ryc. 3. Wykres przedstawiający rezultaty metody łokcia wspomagającej wyznaczenie optymalnej liczby klastrow dla klasteryzacji powiatów na podstawie frekwencji w wyborach do Sejmu RP w 2023 roku. Początkowa liczba klastrow została ustawiona na 1, a końcowa na 25

Źródło: opracowanie własne na podstawie danych Państwowej Komisji Wyborczej (PKW)



Ryc. 4. Mapa wyników klasteryzacji powiatów na podstawie frekwencji w wyborach do Sejmu RP w 2023 roku za pomocą algorytmu k-średnich z liczbą klastrow ustaloną na 4

Źródło: opracowanie własne na podstawie danych Państwowej Komisji Wyborczej (PKW)

tworzącymi wspólne aglomeracje oraz miasta średnie, które nie są otoczone powiatami. Klaster 1, z frekwencją na poziomie $\pm 74\%$, to obszary położone wokół aglomeracji warszawskiej, konurbacji śląskiej oraz powiaty sąsiadujące z miastami wojewódzkimi, w tym klastrze wyróżnić możemy również jednolity obszar centralnej Polski: Wielkopolskę, Górny Śląsk, Ziemię Łódzką, środkowe Mazowsze oraz zachodnią i środkową Małopolskę. Klaster 2 grupuje powiaty o najniższej frekwencji wyborczej, wynoszącej $\pm 65\%$; w tym klastrze przeważają rejony wschodniej Polski, tradycyjnie głosujące na Prawo i Sprawiedliwość, Opolszczyzna oraz część Pomorza Zachodniego. Klaster 3, z frekwencją na poziomie $\pm 70\%$, obejmuje powiaty głównie z województw: podkarpackiego, świętokrzyskiego, lubelskiego, lubuskiego i kujawsko-pomorskiego, jednak bez miast na prawach powiatów oraz powiatów stanowiących aglomerację większych ośrodków miejskich.

Podsumowanie i wnioski

W niniejszej pracy przeanalizowano możliwości zastosowania SI w badaniu tendencji wyborczych w Polsce, wykorzystując dane z lat 2019–2024. Analizy przeprowadzono na podstawie danych publikowanych przez Państwową Komisję Wyborczą, obejmujących wybory do Sejmu RP, samorządowe oraz Parlamentu Europejskiego.

Celem badania było sprawdzenie, jak algorytmy SI i uczenia maszynowego mogą wspierać analizy polityczne i przewidywanie zachowań wyborców, ze szczególnym uwzględnieniem frekwencji wyborczej. Pozwoliło to na uchwycenie istotnych trendów przestrzennych i krótkookresowych zmian czasowych w preferencjach politycznych obywateli Polski. Wyniki wskazują na duży potencjał tych technologii, które mogą nie tylko poprawić precyzję analiz, lecz także przyczynić się do lepszego zrozumienia dynamiki społeczno-politycznej, co jest istotne zarówno dla analityków, decydentów politycznych, jak i dla całego społeczeństwa demokratycznego.

Podkreślono również wyzwania i zagrożenia związane z zastosowaniem SI w analizie wyborczej, szczególnie w kontekście ochrony prywatności oraz przeciwdziałania potencjalnym manipulacjom. Wskazano, jak istotne jest stosowanie wytycznych etycznych przy tworzeniu i wdrażaniu systemów SI, w celu zapewnienia ich zrozumiałości, przewidywalności działania, transparentności, uczciwości oraz bezpieczeństwa. Wnioski z przeprowadzonych analiz mogą stanowić solidną podstawę do dalszych badań nad wykorzystaniem sztucznej inteligencji w analizach politycznych, jednocześnie podkreślając potrzebę odpowiedzialnego podejścia do nowych technologii w kontekście procesów demokratycznych.

Do najważniejszych wniosków z przeprowadzonych badań należą:

1. Korelacja poparcia dla partii z frekwencją wyborczą:
 - przy użyciu metod EDA oraz wizualizacji (mapy, macierz korelacji) wykazano dodatnią korelację (około 0,3 dla współczynnika korelacji Pearsona) pomiędzy poparciem dla KO w powiatach w wyborach do Sejmu RP z 2023 roku a frekwencją wyborczą w wyborach do Sejmu RP z lat 2019 i 2023 oraz do Parlamentu Europejskiego z 2024 roku,
 - zidentyfikowano ujemną korelację ($-0,26$ do $-0,32$ dla współczynnika korelacji Pearsona) pomiędzy poparciem dla PiS w powiatach w wyborach do

Sejmu RP z 2023 roku a frekwencją wyborczą w wyborach do Sejmu RP z lat 2019 i 2023 oraz do Parlamentu Europejskiego z 2024 roku,

- w wyborach do sejmików wojewódzkich w 2024 roku korelacje te uległy zmianie – współczynnik korelacji Pearsona między frekwencją a poparciem dla obu partii wzrósł w wartościach bezwzględnych do ponad 0,5 oraz zmienił swój znak, co oznacza odwrócenie tendencji zaobserwowanej w pozostałych analizowanych wyborach,
- odnotowano silną polaryzację między poparciem dla KO i PiS; korelacja ich poparcia na poziomie powiatów była bardzo silnie ujemna, osiągając poziom $-0,9$.

2. Klasteryzacja powiatów według frekwencji wyborczej:

- przeprowadzono klasteryzację powiatów przy użyciu algorytmu k-średnich dla optymalnej liczby klastrów wyznaczonej metodą łokciową, opierając się na frekwencji wyborczej w wyborach do Sejmu RP z 2023 roku, co pozwoliło na wyodrębnienie 4 klastrów. Algorytm zidentyfikował, że największa frekwencja występowała w największych aglomeracjach i miastach, które przypisano do jednego klastra (klaster 4), natomiast klaster o najniższej tendencji skupione były we wschodniej oraz północno-zachodniej części kraju.

Artykuł został przygotowany w ramach badań statutowych Katedry Geoinformatyki i Informatyki Stosowanej, Wydziału Geologii, Geofizyki i Ochrony Środowiska Akademii Górniczo-Hutniczej w Krakowie.

Literatura

- Ahmed, M., Seraj, R., Islam, S.M.S. (2020). The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics*, 9(8), 1295, doi: 10.3390/electronics9081295
- Arkhipova, D., Janssen, M. (2024). AI Recommendations' Impact on Individual and Social Practices of Generation Z on Social Media: A Comparative Analysis between Estonia, Italy, and the Netherlands. *Semiotica*, 261, 61–86, doi: 10.1515/sem-2023-0089
- Bennett, W.L., Segerberg, A. (2012). The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics. *Information, Communication & Society*, 15(5), 739–768, doi: 10.1080/1369118X.2012.670661
- Berelson, B.R., Lazarsfeld, P.F., McPhee, W.N. (1954). *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago: University of Chicago Press.
- Dommett, K. (2019). Data-driven Political Campaigns in Practice: Understanding and Regulating Diverse Data-Driven Campaigns. *Internet Policy Review*, 8(4), 1–18. doi: 10.14763/2019.4.1432
- Flasiński, M. (2011). *Wstęp do sztucznej inteligencji*. Warszawa: Wydawnictwo Naukowe PWN.
- Flis, J., Stolicki, D. (2017). Przechyły terytorialne – różnicowania wyborczej bazy samorządowych władarzy. W: T. Koziełło, D. Szczepański (red.), *Geografia wyborcza Polski. Interpretacje postaw i zachowań obywateli*. Rzeszów: Wydawnictwo Uniwersytetu Rzeszowskiego.
- Gentzkow, M., Shapiro, J.M. (2011). Ideological Segregation Online and Offline. *The Quarterly Journal of Economics*, 126(4), 1799–1839, doi: 10.1093/qje/qjr043

- Gerlich, M. (2023). Perceptions and Acceptance of Artificial Intelligence: A Multi-Dimensional Study. *Social Sciences*, 12(9), 502, doi: 10.3390/socsci12090502
- Golosov, G.V. (2010). The Effective Number of Parties: A New Approach. *Party Politics*, 16(2), 171–192, doi: 10.1177/1354068809342522
- Haenlein, M., Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future Of Artificial Intelligence. *California Management Review*, 61(4), 5–14, doi: 10.1177/0008125619864925
- Jaworska-Surma, A. (2023). *Fenomen wyborczej mobilizacji: Przyczyny rekordowej frekwencji podczas wyborów parlamentarnych 2023 – wnioski z badań*. Warszawa: Fundacja im. Stefana Batorego.
- Jungherr, A. (2023). Artificial Intelligence and Democracy: A Conceptual Framework. *Social Media + Society*, 9(3), 1–14. doi: 10.1177/20563051231186353
- Kasuya, Y., Moenius, J. (2008). The Nationalization of Party Systems: Conceptual Issues and Alternative District-Focused Measures. *Electoral Studies*, 27(1), 126–135, doi: 10.1016/j.electstud.2007.09.004
- Morandín-Ahuerma, F. (2023). Ethics of AI from Global Companies: Microsoft, Google, Meta, and Apple. W: *Normative Principles for an Ethics of Artificial Intelligence*. Puebla: Council of Science and Technology of the State of Puebla (Concytép – Mexico), 137–161.
- Murawska, A. (2017). Uwarunkowania aktywności wyborczej mieszkańców Łodzi i Iwanowa. *Rocznik Nauk Społecznych. Annals of Social Sciences*, 9(45), 101, doi: 10.18290/rns.2017.9.1-6
- Pei, G., Zhang, J., Hu, M., Zhang, Z., Wang, C., Wu, Y., Zhai, G., Yang, J., Shen, C., Tao, D. (2024). Deepfake Generation and Detection: A Benchmark and Survey. *arXiv:2403.17881*, doi: 10.48550/arXiv.2403.17881
- Rattan, R., Kataria, T., Banerjee, S., Goyal, S., Gupta, D., Pandita, A., Bisht, S., Narang, K., Mishra, S.R. (2019). Artificial Intelligence in Oncology, its Scope and Prospects with Specific Reference to Radiation Oncology. *BJR Open*, 1(1), 20180031, doi: 10.1259/bjro.20180031
- Ray, P.P. (2023). ChatGPT: A Comprehensive Review on Background, Applications, Key Challenges, Bias, Ethics, Limitations and Future Scope. *Internet of Things and Cyber-Physical Systems*, 3, 121–154, doi: 10.1016/j.iotcps.2023.04.003
- Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., Liu, J. (2021). A Quantitative Discriminant Method of Elbow Point for the Optimal Number of Clusters in Clustering Algorithm. *EURASIP Journal on Wireless Communications and Networking*, 31, 1–16. doi: 10.1186/s13638-021-01910-w
- Umargono, E., Suseno, J.E., Gunawan, S.K.V. (2020). K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula. *Proceedings of the 2nd International Seminar on Science and Technology (ISSTEC 2019)*, 121–129. Atlantis Press, doi: 10.2991/assehr.k.201010.019
- Waskom, M.L. (2021). Seaborn: Statistical Data Visualization. *Journal of Open Source Software*, 6(60), 3021, doi: 10.21105/joss.03021
- Zaręba, M., Danek, T., Stefaniuk, M. (2019). Some Statistical Considerations of Azimuth and Inclination Angles Determination Based on Walk-Away VSP Data in Python. *E3S Web of Conferences*, 133, 01006, doi: 10.1051/e3sconf/201913301006
- Zaręba, M., Danek, T., Stefaniuk, M. (2023). Unsupervised Machine Learning Techniques for Improving Reservoir Interpretation Using Walkaway VSP and Sonic Log Data. *Energies*, 16(1), 493, doi: 10.3390/en16010493

Źródła internetowe

Komisja Europejska. *Doskonałość i wiarygodność sztucznej inteligencji*. Commission.europa.eu, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/excellence-and-trust-artificial-intelligence_pl (dostęp: 28.10.2024).

Microsoft. *Identyfikowanie wytycznych dla odpowiedzialnego używania sztucznej inteligencji*. Microsoft Learn, <https://learn.microsoft.com/pl-pl/training/modules/embrace-responsible-ai-principles-practices/3-identify-guiding-principles-responsible-ai> (dostęp: 25.10.2024).

Ministerstwo Cyfryzacji. *Czym jest sztuczna inteligencja?* Gov.pl, <https://www.gov.pl/web/ai/czym-jest-sztuczna-inteligencja2> (dostęp: 15.10.2024).

Źródła danych

Wybory do Parlamentu Europejskiego 2024 r., https://wybory.gov.pl/pe2024/pl/dane_w_arkuszach (dostęp: 1.07.2024).

Wybory do Sejmu i Senatu Rzeczypospolitej Polskiej 2019 r., https://sejmsenat2019.pkw.gov.pl/sejmsenat2019/pl/dane_w_arkuszach (dostęp: 1.07.2024).

Wybory do Sejmu i Senatu Rzeczypospolitej Polskiej 2023 r., https://sejmsenat2023.pkw.gov.pl/sejmsenat2023/pl/dane_w_arkuszach (dostęp: 1.07.2024).

Wybory Samorządowe 2024 r., https://samorząd2024.pkw.gov.pl/samorząd2024/pl/dane_w_arkuszach (dostęp: 1.07.2024).

Nota biograficzna o autorze: Kamil Bednarz jest specjalistą w zakresie sztucznej inteligencji i analizy danych, z sukcesem łącząc swoje zainteresowania naukami o Ziemi z zagadnieniami politycznymi. Studiuje inżynierię i analizę danych na Wydziale Geologii, Geofizyki i Ochrony Środowiska Akademii Górniczo-Hutniczej w Krakowie, w Katedrze Geoinformatyki i Informatyki Stosowanej, gdzie zgłębia wiedzę o geologii, geofizyce oraz możliwościach zastosowania SI w analizach przestrzennych. Aktywnie działa w Kole Naukowym Geopolityki SGH, realizując projekty dotyczące geopolitycznych i strategicznych wyzwań globalnych. Jest Przewodniczącym Koła Naukowego DataTeam Akademii Górniczo-Hutniczej, intensywnie angażuje się w rozwój sztucznej inteligencji, w tym w prace nad generatywnymi narzędziami podobnym do ChatGPT. Trzykrotnie laureat stypendium rektora za wysokie wyniki akademickie. Wyróżnia się szerokimi kompetencjami, łącząc wiedzę techniczną z interdyscyplinarną perspektywą.

Biographical note of author: Kamil Bednarz specializes in artificial intelligence and data analysis, combining his interests in Earth sciences and political issues. He studies engineering and data analysis at the Faculty of Geology, Geophysics, and Environmental Protection at AGH University of Science and Technology, focusing on AI applications in spatial analyses. He is actively involved in the Geopolitics Student Research Group at SGH, working on projects analyzing global geopolitical and strategic challenges. He also serves as the Chair of the DataTeam Student Research Group at AGH, where he is deeply engaged in the development of artificial intelligence, including work on generative tools similar to ChatGPT. A three-time recipient of the Rector's Scholarship for outstanding academic performance, he stands out as a talented student with extensive expertise, combining technical knowledge with a profound interdisciplinary perspective.

e-mail: kamil.bednarz02@gmail.com

Nota biograficzna o autorze: Dominik Jaśkowiec, poseł na Sejm RP, wieloletni radny Krakowa, doktor nauk politycznych, adiunkt na Uniwersytecie Komisji Edukacji Narodowej w Krakowie. Pomysłodawca innowacji miejskich, inicjator projektowania partycypacyjnego. Autor publikacji naukowych dotyczących partycypacji społecznej i samorządowi terytorialnemu.

Biographical note of author: Dominik Jaśkowiec, member of the Polish Parliament, long-time Kraków councilor, doctor of political science, and assistant professor at the University of the National Education Commission in Kraków. Initiator of urban innovations and author of scientific publications on public participation and local government.

e-mail: dominik.jaskowiec@uken.krakow.pl

Nota biograficzna o autorze: Mateusz Zaręba, adiunkt w Katedrze Geoinformatyki i Informatyki Stosowanej na Wydziale Geologii, Geofizyki i Ochrony Środowiska Akademii Górniczo-Hutniczej w Krakowie. Specjalizuje się w interdyscyplinarnych zastosowaniach sztucznej inteligencji, zwłaszcza w analizie danych w badaniach środowiskowych. Jego prace koncentrują się na wykorzystaniu zaawansowanych technik SI i big data do analizy czasowo-przestrzennych wzorców zanieczyszczeń powietrza oraz ich migracji. Bada również możliwości wsparcia, jakie sztuczna inteligencja oferuje w badaniach z zakresu geofizyki, ochrony środowiska i nauk społecznych. Jest laureatem konkursu Geologia 2019 organizowanego przez Głównego Geologa Kraju za innowacyjną technologię przetwarzania sygnałów. Jest autorem licznych publikacji naukowych, które przyczyniają się do rozwoju wiedzy w obszarze sztucznej inteligencji. Jego badania mają kluczowe znaczenie dla zrozumienia interakcji między działalnością człowieka a środowiskiem naturalnym, przyczyniając się do poprawy jakości życia w kontekście ochrony środowiska.

Biographical note of author: Mateusz Zaręba is an assistant professor in the Department of Geoinformatics and Applied Computer Science at the Faculty of Geology, Geophysics, and Environmental Protection at AGH University of Krakow. His research focuses on the interdisciplinary applications of artificial intelligence, particularly in data analysis for environmental studies. His work emphasizes the use of advanced AI and big data techniques to analyze spatiotemporal patterns of air pollution and its migration. He explores how AI can support research in geophysics, environmental protection, and the social sciences. He is a laureate of the Geology 2019 competition, organized by the Chief Geologist of Poland, for his innovative signal processing technology. As the author of numerous scientific publications, he has made significant contributions to the advancement of knowledge in artificial intelligence. His research plays a crucial role in understanding the interactions between human activity and the natural environment, ultimately contributing to improving quality of life within the framework of environmental protection.

e-mail: zareba@agh.edu.pl