

## Porównanie wybranych metod normalizacji zmiennych pod kątem podobieństwa uzyskiwanych rankingów

### Streszczenie:

Normalizacja zmiennych jest niezbędnym krokiem m. in. podczas konstrukcji mierników taksonomicznych opisujących analizowane zjawisko czy grupowania obiektów. Istnieje wiele formuł normalizacyjnych. Różne formuły mogą prowadzić do różnych rankingów czy grup obiektów. Celem niniejszej pracy było porównanie wybranych metod normalizacji zmiennych pod kątem podobieństwa uzyskiwanych rankingów obiektów na podstawie mierników taksonomicznych, w konstrukcji których wykorzystano rozważane metody normalizacji. Analiza zależności przeprowadzona została na przykładzie konstrukcji taksonomicznych mierników atrakcyjności giełd kryptowalut. W rozważanych przykładach mierniki konstruowano metodą sum standaryzowanych, a także z wykorzystaniem metody wzorca rozwoju. Uzyskiwane wyniki zależą od rozważanego zbioru formuł normalizacyjnych. W przypadku porównywania pięciu wybranych formuł normalizacyjnych przekształcenie ilorazowe z podstawą równą maksimum najczęściej prowadziło do uzyskiwania najbliższych (w stosunku do pozostałych metod) rankingów. W przypadku uwzględnienia trzech wybranych formuł normalizacyjnych unitaryzacja zerowana miała taką własność. Porównując metody normalizacji pod kątem podobieństwa uzyskiwanych rankingów nie można (nawet dla konkretnego zbioru danych) wyciągać wniosków, że dana metoda normalizacji daje najczęściej najbliższe rankingi w stosunku do wszystkich innych metod normalizacji, gdyż uzyskiwane rezultaty zależą od uwzględnionych formuł normalizacyjnych.

### Słowa kluczowe:

giełdy walut kryptograficznych, normalizacja, przekształcenie ilorazowe, standaryzacja, unitaryzacja zerowana, miernik taksonomiczny

## Wprowadzenie

Metody normalizacji zmiennych są wykorzystywane m. in. podczas konstrukcji mierników taksonomicznych czy analizy skupień, w celu sprowadzenia wartości rozważanych zmiennych do porównywalności. Istnieje wiele różnych formuł normalizacyjnych. Różne formuły mogą prowadzić do różnych wyników liniowego uporządkowania czy grupowania obiektów. Wybrane metody normalizacji zmiennych przedstawiają m. in. K. Kukuła i M. Walesiak<sup>1</sup>.

K. Kukuła dokonał porównania wybranych metod normowania pod kątem spełnienia siedmiu postulatów, takich jak<sup>2</sup>:

- pozbawienie mian (jednostek), w których są wyrażone cechy diagnostyczne;
- sprowadzenie rzędu wielkości zmiennych diagnostycznych do stanu porównywalności;
- równość rozpiętości przedziałów zmienności wartości wszystkich cech unormowanych oraz równość dolnej i górnej granicy ich przedziału zmienności;
- możliwość normowania cech diagnostycznych przyjmujących wartości zarówno dodatnie, jak i ujemne;
- możliwość normowania cech przyjmujących wartość zero;
- nieujemność wartości cech unormowanych;
- istnienie prostych formuł ujednocających charakter zmiennych.

<sup>1</sup>K. Kukuła, *Metoda unitaryzacji zerowanej na tle wybranych metod normowania cech diagnostycznych*, Acta Scientifica Academiae Ostroviensis nr 4, 1999, s. 6-7; M. Walesiak, *Wybór grup metod normalizacji wartości zmiennych w skalowaniu wielowymiarowym*, Przegląd Statystyczny, 1/2016, s. 10.

<sup>2</sup>K. Kukuła, *Metoda unitaryzacji zerowanej...*, s. 6-8.

W sensie liczby spełnionych powyższych postulatów najlepsza spośród rozważanych przez K. Kukułę metod normalizacyjnych okazała się metoda unitaryzacji zerowanej. M. Walesiak z kolei zaproponował procedurę pozwalającą na wyodrębnienie grup metod normalizacji wartości zmiennych prowadzących do zbliżonych wyników skalowania wielowymiarowego.

Celem niniejszej pracy jest porównanie wybranych metod normalizacji zmiennych pod kątem podobieństwa uzyskiwanych rankingów obiektów na podstawie wartości mierników taksonomicznych, w konstrukcji których wykorzystano rozważane metody normalizacji. Sformułowano następującą hipotezę badawczą: Żadna z rozważanych metod normalizacyjnych nie daje zawsze rezultatów najbliższych (w sensie podobieństwa rankingów) w porównaniu do pozostałych analizowanych metod. Analiza zależności przeprowadzona zostanie na przykładzie konstrukcji taksonomicznych mierników atrakcyjności giełd kryptowalut<sup>3</sup>. W niniejszej pracy uwaga skupiona będzie tylko na problemie metod normalizacyjnych, natomiast zagadnienia wyboru finalnego miernika taksonomicznego opisującego analizowane zjawisko (spośród wielu jego wariantów, które są potencjalnie akceptowane przez badacza), tworzenia na jego podstawie rankingów i klasyfikacji obiektów nie będą poruszane. Problematyka ta była dyskutowana w literaturze. Przedstawiono różne podejścia do problemu wyboru finalnego miernika taksonomicznego. Przykładowo, K. Kukuła i L. Luty zaproponowali, aby spośród wielu rozważanych rankingów (uzyskanych na podstawie odpowiednich mierników taksonomicznych) wybrać ten, dla którego odpowiednio zdefiniowana funkcja podobieństwa rankingów przyjmuje wartość maksymalną<sup>4</sup>. K. Kądziołka zaproponowała dwuetapową procedurę wyboru finalnego miernika w oparciu o podobieństwo uzyskiwanych rankingów oraz zdolności dyskryminacyjne miernika<sup>5</sup>. Istniejąca literatura przedmiotu oferuje też propozycje agregatowych mierników oceny poprawności metod porządkowania liniowego, na podstawie wartości których można dokonać wyboru finalnego miernika spośród rozważanych wariantów. Mierniki takie wykorzystali m. in.: A. Bąk, A. Sompolska – Rzechuła, M. Trojanowska i T. Nęcka<sup>6</sup>. Wspomniani Autorzy proponują konstrukcję miernika agregatowego będącego pierwiastkiem sumy kwadratów wartości mierników cząstkowych odnoszących się do różnych aspektów poprawności metod porządkowania liniowego. Jednakże takie podejście może prowadzić do błędnych rezultatów, gdyż wartości niektórych z wykorzystywanych mierników cząstkowych<sup>7</sup> mogą być ujemne, co po podniesieniu do kwadratu może wpływać na odwrócenie kolejności uporządkowania obiektów. Ponadto wartości niektórych mierników cząstkowych nie są ograniczone do pewnego (tego samego) przedziału, np. [0, 1], co może powodować większy wpływ danego miernika cząstkowego na finalny ranking niż pozostałych mierników. Propozycję modyfikacji agregatowego miernika oceny poprawności metod porządkowania liniowego, pozwalającą na wyeliminowanie tych problemów, przedstawiono w pracy K. Kądziołki<sup>8</sup>. Należy mieć jednak na uwadze, że nie opracowano jak dotąd „jedyną słuszną” metody wyboru finalnego miernika taksonomicznego, a różne podejścia mogą prowadzić do różnych wyników.

<sup>3</sup>Prezentowane w niniejszym artykule wyniki uzyskano z wykorzystaniem darmowego programu R i ogólnodostępnych danych publikowanych na portalach [www.coingecko.com](http://www.coingecko.com), [www.coinmarketcap.com](http://www.coinmarketcap.com), [www.cryptocompare.com](http://www.cryptocompare.com) [dostęp z dnia: 30/08/2020]

<sup>4</sup>K. Kukuła, L. Luty, *Propozycja procedury wspomagającej wybór metody porządkowania liniowego*, Przegląd Statystyczny, R. LXII, Zeszyt 2, 2015, s. 223.

<sup>5</sup>K. Kądziołka, *Propozycja metody wspomagającej wybór miernika taksonomicznego na przykładzie oceny atrakcyjności giełd kryptowalut*, Firma i Rynek, 59/2021, s. 65 – 76.

<sup>6</sup>A. Bąk, *Analiza porównawcza wybranych metod porządkowania liniowego*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 508, 2018, s. 19-28; A. Sompolska – Rzechuła, *Selection of the Method of Linear Ordering using the Example of Assessing the Level of Socio – Economic Development of European Union Countries*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 64(7), 2020, s. 118-129; M. Trojanowska M. i T. Nęcka, *Selection of the Multiple-Criteria Decision-Making Method for Evaluation of Sustainable Energy Development: A Case Study of Poland*, Energies, nr 13, 1-24, 2020, doi: 10.3390/en13236321

<sup>7</sup>Postacie analityczne cząstkowych mierników oceny poprawności metod porządkowania liniowego są przedstawione m. in. w pracy I. Muller – Frączek, *Propozycja miary syntetycznej*, Przegląd Statystyczny nr 4/2017, s. 430 – 432.

<sup>8</sup>K. Kądziołka, *Propozycja miernika oceny poprawności metod porządkowania liniowego*, Zeszyty Naukowe Polskiego Towarzystwa Ekonomicznego w Zielonej Górze, 15, 2021, s. 19-33.

## Analizowane metody normalizacji

Rozważano następujące metody normalizacji<sup>9</sup>:

1. Standaryzacja:  $z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$ , gdzie  $\bar{x}_j$  oznacza wartość średnią  $j$ -tej zmiennej a  $s_j$  – jej odchylenie standardowe
2. Metoda unitaryzacji zerowanej:  $z_{ij} = \frac{x_{ij} - \min_i\{x_{ij}\}}{\max_i\{x_{ij}\} - \min_i\{x_{ij}\}}$
3. Przekształcenie ilorazowe 1:  $z_{ij} = \frac{x_{ij}}{\max_i\{x_{ij}\}}$
4. Przekształcenie ilorazowe 2:  $z_{ij} = \frac{x_{ij}}{\bar{x}_j}$
5. Przekształcenie ilorazowe 3: ,  $z_{ij} = \frac{x_{ij}}{\text{med}_j}$  gdzie  $\text{med}_j$  oznacza medianę  $j$ -tej zmiennej.

Rozważane metody normalizacyjne wybrano w sposób arbitralny. Są wśród nich często stosowane metody takie jak standaryzacja czy unitaryzacja zerowana. Ponadto, wybierając metody normalizacyjne ich liczbę zredukowano w ten sposób, aby nie było wśród nich metod prowadzących zawsze do identycznych wyników porządkowania obiektów z wykorzystaniem mierników taksonomicznych. Poszczególne formuły normalizacji mają określone własności. Standaryzacja powoduje ujednoczenie wartości wszystkich zmiennych pod względem zmienności mierzonej odchyleniem standardowym, co oznacza wyeliminowanie zmienności jako podstawy różnicowania obiektów. Unitaryzacja zerowana zapewnia znormalizowanym wartościom zmiennych zróżnicowaną zmienność i jednocześnie stały rozstęp dla wszystkich zmiennych. Przekształcenie ilorazowe z podstawą równą maksimum zapewnia znormalizowanym wartościom zmiennych zróżnicowaną zmienność, średnią arytmetyczną irozsęp. Z kolei przekształcenia ilorazowe z podstawą normalizacji równą średniej i medianie zapewniają znormalizowanym wartościom zmiennych zróżnicowaną zmienność i rozstęp oraz stałą dla wszystkich zmiennych średnią arytmetyczną (medianę)<sup>10</sup>.

## Charakterystyka analizowanych danych

W prezentowanym przykładzie uwzględniono następujące charakterystyki giełd<sup>11</sup>:

- wskaźnik *Security Rank*, mierzący poziom zabezpieczeń giełdy przeciw zagrożeniom z cyberprzestrzeni ( $x_1$ )
- wskaźnik płynności ( $x_2$ )
- liczba różnych walut, które można nabyć/sprzedać na giełdzie ( $x_3$ )
- wskaźnik jakości świadczonych usług ( $x_4$ )
- KYC/Ryzyko transakcji ( $x_5$ )

<sup>9</sup>Rozważania ograniczono do sytuacji, gdy wszystkie zmienne diagnostyczne (składowe konstruowanych mierników taksonomicznych) są stymulantami. Stymulanta jest zmienną, której większa wartość wskazuje na większy poziom rozpatrywanego zjawiska. Z kolei destymulanta to zmienna mająca na to zjawisko negatywny wpływ, por. A. Młodak, *Analiza taksonomiczna w statystyce regionalnej*, Difin, Warszawa, 2006, s. 33. W przypadku, gdyby wśród zmiennych diagnostycznych występowały destymulanty, konstruując mierniki metodą sum standaryzowanych, należałoby je przekształcić na stymulanty.

<sup>10</sup>M. Walesiak, *Przegląd formuł normalizacji wartości zmiennych oraz ich własności w statystycznej analizie wielowymiarowej*, Przegląd Statystyczny, 4/2014, s. 367.

<sup>11</sup>Ocenie poddano 80 giełd, dla których dostępne były wartości wszystkich rozważanych zmiennych. Zmienne  $x_1 - x_6$  zostały również opisane i wykorzystane w pracy K. Kądziołki, *Metoda PROMETHEE II w wielokryterialnej ocenie giełd kryptowalut*, *Economic and Regional Studies*, 14(2), 2021, s. 131-145, a zmienne  $x_1 - x_2$  (oraz dodatkowa zmienna dotycząca popularności giełd) w pracy K. Kądziołki, *Propozycja metody wspomagającej wybór miernika taksonomicznego...*, s. 65 – 76.

- czas funkcjonowania giełdy ( $x_6$ )

Wskaźnik *Security Rank* dotyczy oceny poziomu zabezpieczeń giełdy przeciw zagrożeniom z cyberprzestrzeni. Wyznaczany jest dla giełd kryptowalut na platformie cer.live. Wskaźnik ten jest pewnym miernikiem agregatowym, oceniającym podatność użytkowników i serwerów na zagrożenia z cyberprzestrzeni<sup>12</sup>. Im wyższa wartość wskaźnika tym wyższy poziom zabezpieczeń giełdy przeciwko zagrożeniom z cyberprzestrzeni.

Wskaźnik płynności wyznaczany jest przez portal coinmarketcap.com. Wskaźnik ten jest również pewnym miernikiem agregatowym i uwzględnia różne zmienne zawarte w tzw. księdze zamówień (ang. *order book*), jak np. wielkość zamówienia czy odległość od średniej ceny<sup>13</sup>. Im wyższa wartość tego wskaźnika tym wyższa płynność, rozumiana tu jako możliwość szybkiej sprzedaży lub kupna kryptowaluty po cenie zbliżonej do rynkowej.

Wskaźnik jakości świadczonych usług oraz wskaźnik dotyczący implementacji procedur KYC (ang. *Know Your Customer*) i ryzyka transakcji wyznaczone są przez portal cryptocompare.com. Wartość pierwszego powstaje przez zsumowanie punktów przyznanych w ramach różnych kategorii dotyczących jakości świadczonych usług przez platformę giełdową, jak np. średni czas odpowiedzi aplikacji (ms) czy gromadzenie i udostępnianie danych historycznych. Wartość drugiego jest wyznaczana przez zsumowanie punktów przyznanych w ramach różnych kategorii dotyczących obszaru związanego z przeciwdziałaniem praniu pieniędzy, jak np. identyfikacja użytkowników, monitorowanie i identyfikacja podejrzanych transakcji<sup>14</sup>.

Współczynnik zmienności każdej ze zmiennych wynosił powyżej 10% oraz nie występowało silne ich skorelowanie<sup>15</sup> (patrz tab. nr 1), w związku z czym nie dokonywano ich redukcji.

**Tab. 1. Współczynniki korelacji Spearmana między zmiennymi diagnostycznymi**

	x1	x2	x3	x4	x5	x6
x1	1,0000	0,3003	0,0257	0,3245	0,3983	0,2954
x2	0,3003	1,0000	0,3786	0,3920	0,2466	0,1433
x3	0,0257	0,3786	1,0000	0,1388	-0,2573	-0,1718
x4	0,3245	0,3920	0,1388	1,0000	0,0674	0,0322
x5	0,3983	0,2466	-0,2573	0,0674	1,0000	0,2854
x6	0,2954	0,1433	-0,1718	0,0322	0,2854	1,0000

Źródło: opracowanie własne.

<sup>12</sup>Szczegóły dotyczące metodologii dostępne są na stronie <https://cer.live/certified>, data dostępu: 30.08.2020.

<sup>13</sup>Szczegółowe informacje na temat tego wskaźnika dostępne są na stronie internetowej <https://support.coinmarketcap.com/hc/en-us/articles/360043836931-Liquidity-Score-Market-Pair-Exchange>, [dostęp z dnia: 30/08/2020]

<sup>14</sup>Szczegóły wyznaczania wskaźnika jakości usług i KYC opisane są w dokumencie: [www.cryptocompare.com/media/37072188/cryptocompare-exchange-benchmark-july-2020.pdf](http://www.cryptocompare.com/media/37072188/cryptocompare-exchange-benchmark-july-2020.pdf), [dostęp z dnia: 30/08/2020].

<sup>15</sup>Do oceny skorelowania zmiennych diagnostycznych wykorzystano współczynnik korelacji Spearmana, który w odróżnieniu od współczynnika korelacji liniowej mierzy również nieliniowe zależności monotoniczne.

## Konstrukcja mierników taksonomicznych

W niniejszej pracy rozważano dwa rodzaje mierników taksonomicznych: mierniki konstruowane z wykorzystaniem metody sum standaryzowanych oraz mierniki konstruowane z wykorzystaniem tzn. wzorca rozwoju z uwzględnieniem wag poszczególnych składowych. W przypadku metody sum standaryzowanych miernik taksonomiczny definiowany jest następująco:

$$TM_i = \sum_{j=1}^m w_j z_{ij}$$

gdzie  $z_{ij}$  – znormalizowane wartości zmiennych diagnostycznych (przekształconych do postaci stymulant),  $i=1, \dots, n$ ,  $j=1, \dots, m$ ,  $w_j$  oznacza wagę  $j$  – tej zmiennej diagnostycznej,  $w_j > 0$ ,  $j=1, \dots, m$ .

Konstruując mierniki metodą wzorca rozwoju, zastosowano metodę Hellwiga<sup>16</sup>,  $\sum_{j=1}^m w_j = 1$  uwzględniając dodatkowo wagi zmiennych diagnostycznych. Konstrukcja mierników obejmowała w tym przypadku następujące kroki:

1. Normalizacja zmiennych diagnostycznych<sup>17</sup>.
2. Konstrukcja współrzędnych tzw. wzorca rozwoju wg wzoru:

$$z_{0j} = \begin{cases} \max_i \{z_{ij}\}, & \text{w przypadku stymulant} \\ \min_i \{z_{ij}\}, & \text{w przypadku destymulant} \end{cases}$$

3. Wyznaczenie dla każdego obiektu odległości od wzorca wg wzoru:

$$d_{i0} = \sqrt{\sum_{j=1}^m w_j (z_{ij} - z_{0j})^2}, \text{gdzie } w_j \text{ oznacza wagę } j \text{ – tej zmiennej diagnostycznej.}$$

4. Wyznaczenie wartości zmiennej syntetycznej wg wzoru:  $TM_{-H_i} = 1 - \frac{d_{i0}}{d_0}$ , gdzie  $d_0 = \bar{d}_0 + 2s_d$  gdzie  $\bar{d}_0$  i  $s_d$  oznaczają odpowiednio wartość średnią i odchylenie standardowe odległości obiektów od wzorca. Im wyższa wartość tak skonstruowanej zmiennej syntetycznej tym dany obiekt lepszy w sensie przyjętych kryteriów.

W analizowanym przykładzie wszystkie zmienne diagnostyczne były stymulantami.

## Ocena podobieństwa rankingów dla pięciu metod normalizacji

Podobieństwo uzyskiwanych rankingów do innych rankingów oceniane było za pomocą funkcji  $Mean\_CS$ , zdefiniowanej następująco<sup>18</sup>:  $Mean\_CS_i = \frac{1}{k-1} \sum_{j=1, j \neq i}^k \rho_{ij}$ , gdzie  $\rho_{ij}$  oznacza współczynnik korelacji Spearmana między wartościami  $i$  – tego i  $j$  – tego miernika taksonomicznego,  $k$  oznacza liczbę rozważanych mierników taksonomicznych. W pierwszej kolejności porównano rezultaty w sytuacji przyjęcia równych wag i konstrukcji mierników metodą sum standaryzowanych. Najbardziej podobny do pozostałych

<sup>16</sup> Z. Hellwig, *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*, Przegląd Statystyczny, z. 4, 1968, s. 307 – 327.

<sup>17</sup> W przypadku mierników konstruowanych zgodnie z metodą Hellwiga, w celu normalizacji zmiennych diagnostycznych często wykorzystywana jest standaryzacja

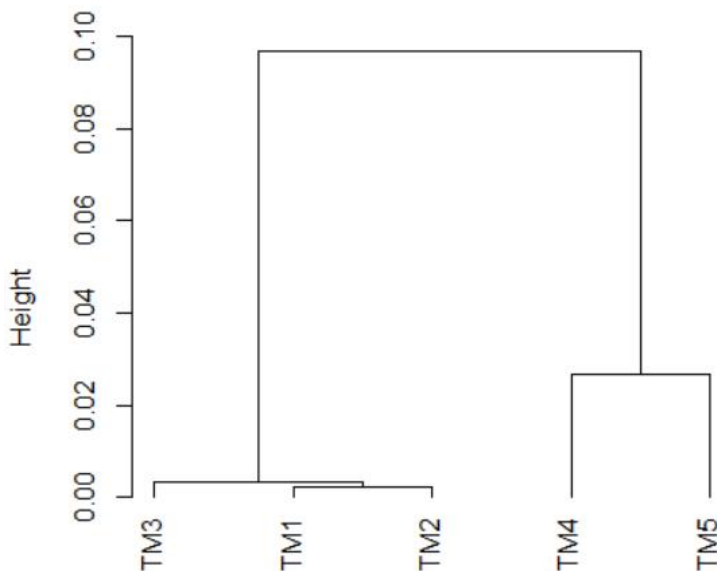
<sup>18</sup> Funkcja ta została zaproponowana i wykorzystana w pracy K. Kądziołki, *Metoda PROMETHEE II w wielokryterialnej ocenie giełd kryptowalut...*, s. 136 w kontekście oceny podobieństwa rankingów giełd kryptowalut uzyskanych z wykorzystaniem podejścia wielokryterialnego.

w sensie maksymalizacji wartości funkcji *Mean\_CS* był ranking uzyskany za pomocą miernika taksonomicznego, w którym normalizacji zmiennych dokonano wykorzystując przekształcenie ilorazowe 1 określone wzorem<sup>19</sup>:  $Z_{ij} = \frac{x_{ij}}{\max_i\{x_{ij}\}}$ . W tab. nr 2 przedstawiono wartości funkcji *Mean\_CS* w przypadku wykorzystania równych wag. TMi oznacza miernik taksonomiczny, w którym zastosowano *i* - tą formułę normalizacyjną<sup>20</sup>. Rys. 1 przedstawia grupy rankingów podobnych na podstawie rozważanych mierników taksonomicznych. Jako miarę niepodobieństwa wykorzystano formułę  $d=1-\rho_{ij}$  gdzie  $\rho_{ij}$  oznacza współczynnik korelacji Spearmana między wartościami *i* - tego i *j* - tego miernika. Do łączenia skupień wykorzystano metodę pojedynczego wiązania, w której odległość między dwoma skupieniami jest równa minimalnej odległości między dwoma obiektami należącymi do różnych skupień. Uwidocznił się podział na dwie grupy mierników. Pierwszą grupę stanowiły mierniki, w których wykorzystano pierwsze trzy analizowane formuły normalizacyjne. Pozostałe mierniki utworzyły drugie skupienie.

Tab. 2. Wartości funkcji *Mean\_CS* (równe wagi, metoda sum standaryzowanych)

Miernik taksonomiczny	Mean_CS
TM3	0,9274
TM1	0,9178
TM2	0,9145
TM4	0,9105
TM5	0,8404

Źródło: opracowanie własne.



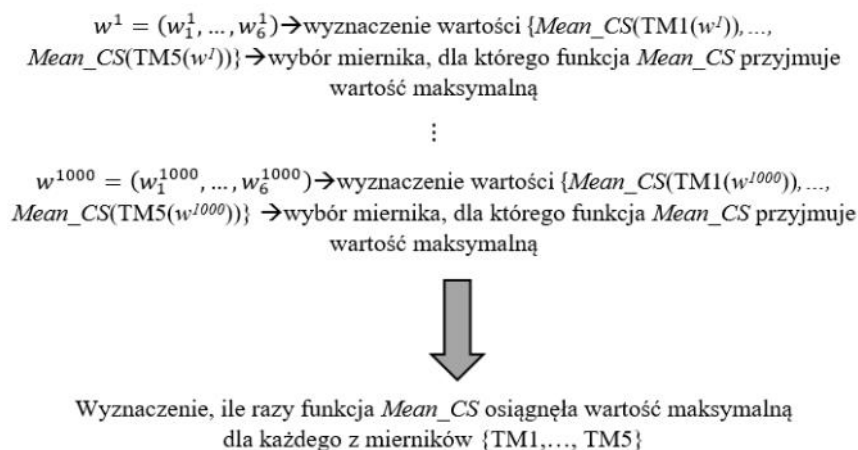
Rys. 1. Grupy podobnych rankingów wg wartości mierników taksonomicznych

Źródło: opracowanie własne.

<sup>19</sup> W kontekście przyjętych oznaczeń, w tym przypadku  $k=5$  (liczba porównywanych mierników taksonomicznych).

<sup>20</sup> TM1 – oznacza miernik taksonomiczny, w którym normalizacji zmiennych dokonano wykorzystując standaryzację. TM2 oznacza miernik, w którym normalizacji zmiennych dokonano wykorzystując unitaryzację zerowaną. TM3, TM4, TM5 oznaczają odpowiednio mierniki, w których normalizacji zmiennych dokonano wykorzystując odpowiednio przekształcenie ilorazowe 1, 2 i 3. Analogiczna jest interpretacja oznaczeń mierników 1 – 5 w dalszej części pracy.

Następnie wygenerowano losowo 1000 wektorów wag  $w^r = (w_1^r, \dots, w_6^r) = 1, \dots, 1000$ . Dla każdego wektora utworzono 5 mierników taksonomicznych konstruowanych z uwzględnieniem rozważanych metod normalizacji. W kolejnym kroku dla każdego wektora wag porównano wartości funkcji *Mean\_CS* dla utworzonych mierników i wyznaczono (na podstawie wyników dla wszystkich wektorów wag), ile razy każdy z mierników TM1,...,TM5 charakteryzował się maksymalną wartością funkcji *Mean\_CS*. Na rys. nr 2 przedstawiono opisany schemat postępowania.



**Rys. 2. Procedura oceny mierników pod względem podobieństwa rankingów**  
Źródło: opracowanie własne

Dla 873 przypadków najlepszy rezultat (maksymalna wartość funkcji *Mean\_CS*) został osiągnięty przy przekształceniu ilorazowym 1:  $z_{ij} = \frac{x_{ij}}{\max_i \{x_{ij}\}}$ .

Dla 93 przypadków najlepszy rezultat był przy przekształceniu ilorazowym 2 a dla 34 dla klasycznej standaryzacji. Pozostałe rozważane metody (unitaryzacja zerowana i przekształcenie ilorazowe 3) ani razu nie dały najlepszego rezultatu w sensie podobieństwa rankingów.

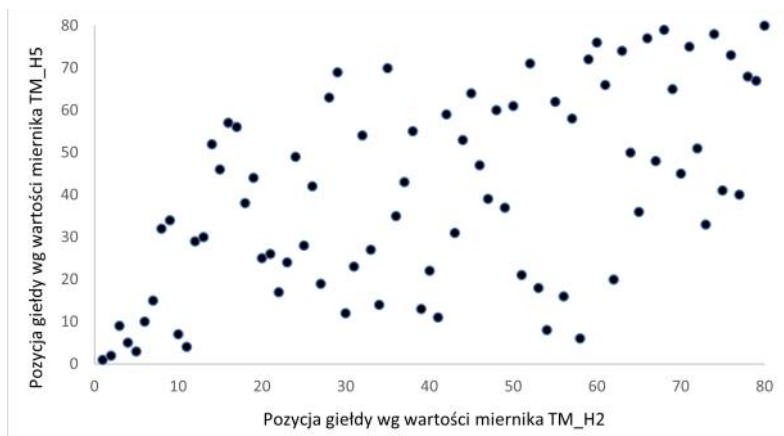
Następnie procedury powtórzono dla mierników konstruowanych metodą wzorca rozwoju. W przypadku przyjęcia równych wag uzyskano podobny rezultat, jak dla mierników konstruowanych metodą sum standaryzowanych (patrz tab. nr 3).

**Tab 3. Wartości funkcji *Mean\_CS* (równe wagi, metoda wzorca rozwoju)**

Miernik taksonomiczny	Mean_CS
TM_H3	0,8548
TM_H1	0,8216
TM_H2	0,8199
TM_H4	0,8075
TM_H5	0,6865

Źródło: opracowanie własne.

W przypadku mierników konstruowanych z wykorzystaniem wzorca rozwoju, można zauważyć większe różnice w uporządkowaniach giełd przy wykorzystaniu różnych metod normalizacji zmiennych niż w przypadku mierników konstruowanych metodą sum standaryzowanych. Sytuacja ta została zobrazowana na przykładzie pozycji giełd wg wartości miernika TM\_H2 oraz TM\_H5 (patrz rys. nr 3). Współczynnik korelacji Spearmana między wartościami tych mierników wyniósł 0,5666. Przykład ten obrazuje, że wybór formuły normalizacyjnej może mieć istotny wpływ na wynik porządkowania obiektów.



**Rys. 3. Pozycje giełd wg wartości mierników TM\_H2 i TM\_H5**

Źródło: opracowanie własne.

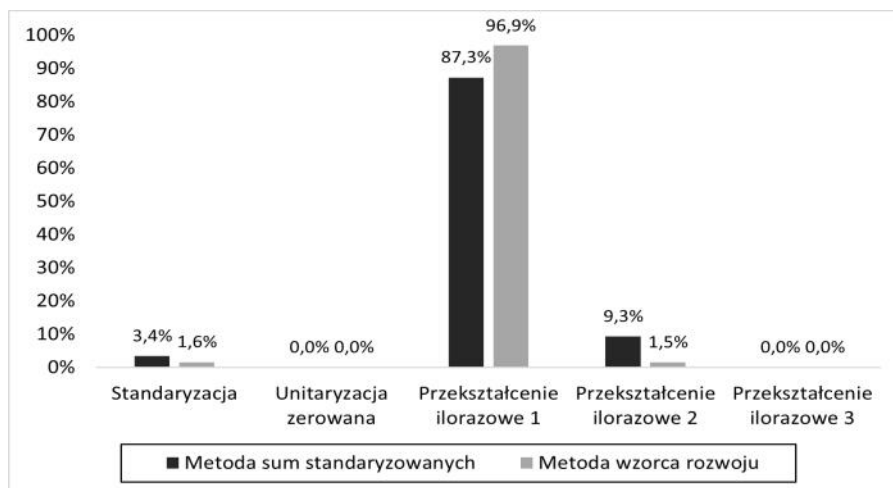
Następnie rozważano 1000 losowo wygenerowanych wektorów wag  $w^r = (w_1^r, \dots, w_6^r) = 1, \dots, 1000$ . Dla każdego wektora utworzono 5 mierników taksonomicznych konstruowanych z uwzględnieniem rozważanych metod normalizacji. W kolejnym kroku dla każdego wektora wag porównano wartości funkcji *Mean\_CS* dla utworzonych mierników.

Dla 969 przypadków najlepszy rezultat (maksymalna wartość funkcji *Mean\_CS*) został osiągnięty przy przekształceniu ilorazowym 1:  $z_{ij} = \frac{x_{ij}}{\max_i\{x_{ij}\}}$ .

Dla 15 przypadków najlepszy rezultat był przy przekształceniu ilorazowym 2 a dla 16 dla klasycznej standaryzacji. Pozostałe rozważane metody (unitaryzacja zerowana i przekształcenie ilorazowe 3) ani razu nie dały najlepszego rezultatu w sensie maksymalizacji wartości funkcji *Mean\_CS*.

Na rys. nr 4. przedstawiono odsetek najlepszych (w sensie podobieństwa rankingów) wyników poszczególnych metod normalizacji w przypadku opisanego wcześniej losowego generowania wag zmiennych diagnostycznych i konstrukcji mierników metodą sum standaryzowanych oraz metodą wzorca rozwoju. W analizowanym przykładzie maksymalna wartość funkcji *Mean\_CS* wyraźnie częściej była osiągnięta w przypadku przekształcenia ilorazowego z podstawą równą maksimum.





Rys. 4. Odsetek najlepszych rezultatów dla poszczególnych metod normalizacji

Źródło: opracowanie własne.

### Rozważane metody normalizacji a uzyskiwane wyniki

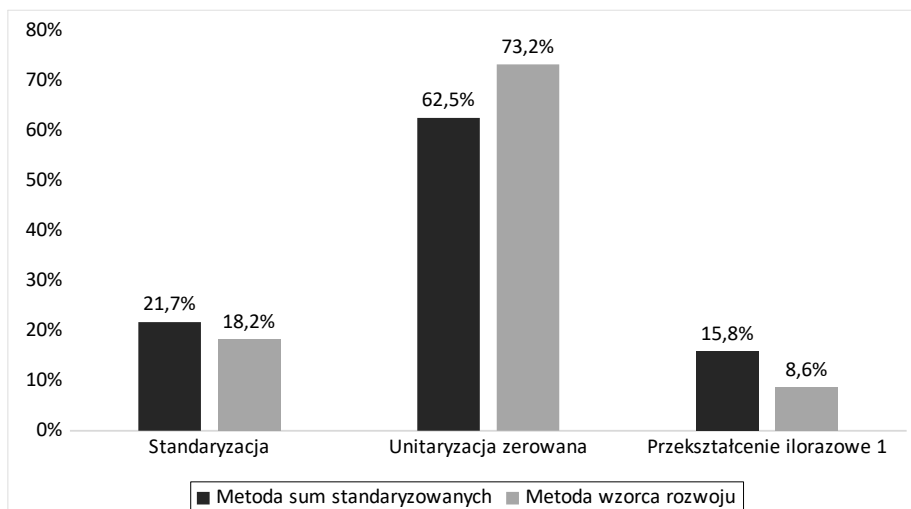
Prezentowane obliczenia powtórzono rozważając tylko pierwsze trzy metody normalizacji. Tym razem metodą, która najczęściej dawała najlepszy wynik była metoda unitaryzacji zerowanej. W tab. nr 4 przedstawiono współczynniki korelacji Spearmana między wartościami mierników oraz wartość funkcji *Mean\_CS* w przypadku mierników konstruowanych metodą sum standaryzowanych i przy przyjęciu równych wag składowych mierników. W rozważanym przypadku najwyższą wartością funkcji *Mean\_CS* charakteryzował się miernik, w którym normalizację zmiennych wykonano metodą unitaryzacji zerowanej, a najgorszą miernik, gdzie wykorzystano przekształcenie ilorazowe z podstawą maksimum.

Tab. 4. Wartości funkcji *Mean\_CS* (równe wagi, metoda sum standaryzowanych)

	TM1	TM2	TM3	Mean_CS
TM1	1,00000	0,99768	0,99667	0,99718
TM2	0,99768	1,00000	0,99669	0,99719
TM3	0,99667	0,99669	1,00000	0,99668

Źródło: opracowanie własne.

Na rys. nr 5 przedstawiono rezultat analogicznych obliczeń jak wcześniej, przy tych samych, co w poprzednim przykładzie wektorach wag  $w^r = (w_1^r, \dots, w_6^r), =1, \dots, 1000$ . Zarówno w przypadku mierników konstruowanych metodą sum standaryzowanych, jak i mierników konstruowanych z wykorzystaniem metody wzorca rozwoju, najczęściej najlepszy rezultat (w sensie podobieństwa rankingów) dawała metoda unitaryzacji zerowanej.



**Rys. 5. Odsetek najlepszych rezultatów dla poszczególnych metod normalizacji**

Źródło: opracowanie własne.

### Podsumowanie

Postawiona na wstępie hipoteza została potwierdzona. Co więcej, pokazano, że nie jest nawet prawdą, że dla dowolnego zbioru danych (zmiennych diagnostycznych, z których konstruowany jest miernik taksonomiczny) istnieje wśród rozważanych formuł normalizacyjnych taka, która najczęściej daje rezultaty (rankingi) najbliższe w porównaniu z pozostałymi rozważanymi formułami. W prezentowanym przykładzie, przy porównaniu pięciu metod normalizacji, najbardziej podobne rankingi do innych metod dawała metoda wykorzystująca przekształcenie ilorazowe z podstawą maksimum. W analizowanym przykładzie przy pięciu metodach normalizacji, unitaryzacja zerowana ani razu nie umożliwiła uzyskania najlepszego rezultatu. Z kolei przy ograniczeniu się do trzech metod normalizacji, najczęściej najlepszy rezultat był uzyskiwany przy wykorzystaniu unitaryzacji zerowanej. W tym przypadku z kolei przekształcenie ilorazowe z podstawą maksimum najrzadziej prowadziło do najlepszych rezultatów. Ocena danej metody normalizacyjnej pod kątem maksymalizacji podobieństwa uzyskiwanych rankingów, zależy od innych rozważanych metod normalizacji. Uzyskiwane zależności nie mają uniwersalnego charakteru nawet dla konkretnego zbioru danych. Nie można więc, nawet dla danego, konkretnego zbioru danych, wyciągać ogólnych wniosków, że dana metoda normalizacji częściej niż wszystkie inne umożliwia uzyskiwanie najbardziej podobnych rankingów (w porównaniu do wszystkich innych metod), gdyż uzyskiwane wyniki zależą od rozważanego zbioru metod normalizacji. Na przykładzie zobrazowano, że redukcja liczby rozważanych metod normalizacji wpłynęła na kolejność pozostałych metod pod względem odsetka najbardziej podobnych wyników w stosunku do pozostałych rozważanych metod.

Wybór metody normalizacji zmiennych może mieć istotny wpływ na finalny ranking obiektów uzyskany na podstawie wartości mierników taksonomicznych. W przykładzie zaprezentowanym na rys. 3 maksymalna różnica w uporządkowaniu wg wartości rozważanych mierników wyniosła 52 pozycje. Pojawia się więc pytanie, które uporządkowanie giełd jest prawidłowe? Nie opracowano jak dotąd metody pozwalającej rozstrzygnąć, w jaki sposób dla danego zbioru danych wybrać „najodpowiedniejszą” formułę normalizacji<sup>21</sup>. Wyniki prezentowanego w niniejszej pracy podejścia opartego na podobieństwie uzyskiwanych rankingów zależą od uwzględnionych formuł normalizacyjnych, ale tu pojawia się problem, na podstawie jakich kryteriów wyznaczyć dla danego zbioru danych zbiór rozważanych formuł normalizacyjnych, z których finalnie wybrana zostanie jedna.

<sup>21</sup>W przypadku, gdy w analizowanym zbiorze danych występują obserwacje nietypowe, zasadne wydaje się wykorzystanie odpowiednich metod normalizacji, uwzględniających ten problem. Jednakże istnieją różne metody tego typu. Przykładowe metody normowania zmiennych opisujących nietypowe obiekty przedstawia m. in.: K. Nermend, *Normowanie zmiennych opisujących obiekty nietypowe*, *Metody Ilościowe w Badaniach Ekonomicznych*, X/2009, s. 185 – 193.

---

**Bibliografia**

---

- Bąk A., *Analiza porównawcza wybranych metod porządkowania liniowego*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 508, 2018, doi: 10.15611/pn.2018.508.02
- Hellwig, *Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr*, Przegląd Statystyczny, z. 4, 1968.
- Kądziołka K., *Propozycja metody wspomagającej wybór miernika taksonomicznego na przykładzie oceny atrakcyjności giełd kryptowalut*, Firma i Rynek, 59/2021.
- Kądziołka K., *Metoda PROMETHEE II w wielokryterialnej ocenie giełd kryptowalut*, Economic and Regional Studies, 14(2), 2021, doi: 10.2478/ers-2021-0010.
- Kądziołka K., *Propozycja miernika oceny poprawności metod porządkowania liniowego*, Zeszyty Naukowe Polskiego Towarzystwa Ekonomicznego w Zielonej Górze, 15, 2021, doi: 10.26366/PTE.ZG.2021.201.
- Kukuła K., Luty L., *Propozycja procedury wspomagającej wybór metody porządkowania liniowego*, Przegląd Statystyczny, 2/2015.
- Kukuła K., *Metoda unitaryzacji zerowanej na tle wybranych metod normowania cech diagnostycznych*, Acta Scientifica Academiae Ostroviensis nr 4, 1999.
- Młodak A., *Analiza taksonomiczna w statystyce regionalnej*, Difin, Warszawa, 2006.
- Muller – Frączek I., *Propozycja miary syntetycznej*, Przegląd Statystyczny nr 4/2017.
- Nermend K., *Normowanie zmiennych opisujących obiekty nietypowe*, Metody Ilościowe w Badaniach Ekonomicznych, nr X, 2009.
- Sompolska – Rzechuła A., *Selection of the Method of Linear Ordering using the Example of Assessing the Level of Socio – Economic Development of European Union Countries*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, 64(7), 2020.
- Trojanowska M., Nęcka T., *Selection of the Multiple-Criteria Decision-Making Method for Evaluation of Sustainable Energy Development: A Case Study of Poland*, Energies, nr 13, doi: 10.3390/en13236321.
- Walesiak M., *Przegląd formuł normalizacji wartości zmiennych oraz ich własności w statystycznej analizie wielowymiarowej*, Przegląd Statystyczny, 4/2014.
- Walesiak M., *Wybór grup metod normalizacji wartości zmiennych w skalowaniu wielowymiarowym*, Przegląd Statystyczny, 1/2016.
- [www.coingecko.com](http://www.coingecko.com)
- [www.coinmarketcap.com](http://www.coinmarketcap.com)
- [www.cryptocompare.com](http://www.cryptocompare.com)

## Comparison of selected normalisation methods in terms of rankings similarity

### Summary:

The normalization process is a necessary step, among others during the construction of taxonomic measures describing the analyzed phenomenon or clustering of objects. There are many different normalisation formulas. Different formulas can lead to different rankings or groups of objects. The aim of this study was to compare selected methods of normalization in terms of the similarity of the obtained rankings of objects according to values of taxonomic measures. The analysis was carried out on the example of taxonomic measures of attractiveness of cryptocurrency exchanges.

### Keywords:

cryptocurrency exchanges, normalisation, quotient transform, standardisation, unitarisation, composite indicator