

Zastosowanie eksploracji danych do identyfikacji oznaczonych wyników pomiaru jakości energii elektrycznej w ujęciu obszarowym

Streszczenie. Artykuł zawiera wykorzystanie eksploracji danych w analizie parametrów określających jakość energii elektrycznej (JEE) pod kątem identyfikacji danych podlegających regule oznaczania w rozumieniu normy PN EN 61000-4-30. Zaprezentowano możliwość wykorzystania analizy skupień jako narzędzia umożliwiającego podział zagregowanych danych pomiarowych na grupy reprezentujące wyniki pomiarów wolne od zdarzeń napięciowych oraz wyniki pomiarów, w trakcie których wystąpiło zdarzenie napięciowe. Artykuł zawiera wyniki badań wrażliwości wybranego algorytmu analizy skupień (*k*-średnich) na identyfikację danych zawierających przerwy, zapady, wzrosty oraz szybkie zmiany napięcia. Za zbiór danych testowych wykorzystano synchroniczne pomiary przeprowadzone w sieci zakładów górniczych. Uzyskane wyniki pozwalają na określenie skuteczności wykorzystania analizy skupień do identyfikacji danych zagregowanych zawierających zdarzenia napięciowe w ujęciu obszarowym. Badanym obszarem jest sieć elektroenergetyczna zasilająca wybrany zakład górniczy wydobywający rudy miedzi na Dolnym Śląsku.

Abstract. The article presents the use of data mining to power quality issue. The possibility to using cluster analysis as an appreciate tool to realize division into groups representing the measurement period for which aggregated data (within the meaning of PN EN 61000-4-30 standard) contain and do not contain aggregated voltage events is presented. The K-means algorithm sensitivity test to the identification of data containing interruptions, dips, increases and rapid voltage changes was presented. Synchronous measurements carried out in the mining plant network were used for the test data set. The obtained results allow determining the effectiveness of using cluster analysis to identify aggregated data containing voltage events. The area-related in this article is electrical power network of copper mining industry in Lower Silesia. (**Application of cluster analysis to identification flagged power quality measurements in area-related approach**).

Słowa kluczowe: jakość energii elektrycznej, analiza skupień, koncepcja oznaczania; analiza obszarowa.

Keywords: power quality, cluster analysis, flagging concept., area related analysis.

Wstęp

Rozwój systemów pomiarowo-kontrolnych stosowanych w elektroenergetyce spowodował, że dane pomiarowe pochodzące z tych systemów traktowane są jak duże zbiory danych (ang. big data). Ważnym zagadnieniem staje się opracowanie metod wspierających analizę dużych zbiorów danych. Jednym z rozwiązań jest wykorzystanie różnorodnych technik eksploracji danych. Poniższa praca jest przykładem wykorzystania analizy skupień (ang. cluster analysis) jako narzędzia wspierającego ocenę danych pochodzących z różnych punktów pomiarowych traktowanych jako obszar ze względu na wzajemne powiązania. Badaną siecią jest sieć zasilająca duży zakład górniczy rud miedzi na Dolnym Śląsku. Wybór sieci kopalnianej uwarunkowany jest charakterem obciążenia – duża zmienność w czasie oraz istotność problematyki jakości energii dla tak dużego odbiorcy [1]. Jako źródła danych wykorzystano rzeczywiste dane pomiarowe parametrów jakości energii elektrycznej, zarejestrowane w badanej sieci kopalnianej w trakcie synchronicznych pomiarów wielopunktowych z wykorzystaniem analizatorów klasy A. Przedstawione w artykule zastosowanie analizy skupień służy automatycznej identyfikacji danych podlegających regule oznaczania (ang. flagging concept) w rozumieniu normy PN EN 61000-4-30 [2]. Skuteczność identyfikacji danych oznaczonych z wykorzystaniem zaproponowanego narzędzia analizy skupień oceniana jest przez porównanie do identyfikacji danych oznaczonych w sposób klasyczny tj. oznaczeniu danych zagregowanych, w trakcie których wystąpiły zdarzenia typu zapad, wzrost lub przerwa. Artykuł wskazuje również na możliwość wykorzystania analizy skupień do identyfikacji danych oznaczonych w sposób synchroniczny dla wszystkich punktów objętych obszarem obserwacji. Zaproponowane podejście stanowi rozszerzenie możliwości klasycznej koncepcji oznaczania danych, dla której proces oznaczania odbywa się lokalnie dla punktu pomiarowego, bez ujęcia informacji o oznaczeniu danych pochodzących z innego punktu pomiarowego.

Eksploracja danych – analiza skupień

Eksploracja danych w literaturze określona jest przez wiele definicji m in.:

- eksploracja danych jest wydobywaniem ukrytych, wcześniej nieznanymi i potencjalnymi informacjami z danych [3],
- eksploracja danych jest krokiem w procesie odkrywania wiedzy w bazach danych, który składa się z zastosowania analizy danych i odkrycia algorytmu, które przy akceptowalnych ograniczeniach wydajności obliczeniowej generują określone wyliczenia wzorców (lub modeli) danych [4],
- eksploracja danych jest analizą (często dużych) obserwowalnych zbiorów danych w celu odnalezienia nieoczekiwanych związków i by podsumować dane w sposób zrozumiały i użyteczny dla użytkownika [5],

Podane powyżej definicje mimo swej różnorodności wskazują na najważniejszy element wspólny jakim jest pozyskanie informacji (wiedzy, zależności) z dużego zbioru danych. Istotnym elementem jest przeprowadzenie procesu, który zapewni wynik zrozumiały i przede wszystkim użyteczny dla użytkownika.

Eksploracja danych realizowana jest z wykorzystaniem wielu technik. Powszechną klasyfikacją jest podział na metody z nauczycielem (ang. supervised learning), oraz bez nauczyciela (ang. unsupervised learning) [6].

Do metod z nauczycielem zalicza się:

- regresję logistyczną (ang. logistic regression),
 - wielowarstwowa sieć perceptronowa (ang. multilayered perceptron network – MLP network),
 - zasady decyzyjne (ang. decision rules),
 - drzewa decyzyjne (ang. decision trees).
- Natomiast w przypadku metod bez nauczyciela wskazano:
- sztuczne sieci neuronowe (ang. artificial neural networks),
 - reguły asocjacyjne (ang. association rules),
 - analiza skupień (ang. clustering).

Zaproponowaną w artykule techniką eksploracji danych jest analiza skupień (w polskiej literaturze znana również jako klastrowanie, grupowanie oraz taksonomia). Ukierunkowana jest na uzyskanie homogenicznych grup (skupień) danych [7]. Istnieje wiele różnych możliwości doboru warunków podziału danych na skupienia. Za najczęściej stosowane kryteria można przyjąć podobieństwo obiektów przypisanych do danej grupy lub odmiennosć obiektów jednej grupy od elementów grup pozostałych.

Zasadniczym celem analizy skupień jest uzyskanie homogenicznych grup danych (skupień) [7]. Istnieje wiele różnych możliwości doboru warunków podziału danych na skupienia. Za najczęściej stosowane kryteria można przyjąć podobieństwo obiektów przypisanych do danej grupy lub odmiennosć obiektów jednej grupy od elementów grup pozostałych [8],[9]. Jednak ze względu na podejście do procesu grupowania można wyróżnić hierarchiczne i niehierarchiczne metody podziału. Hierarchiczne realizowane są poprzez tworzenie hierarchii klasyfikacji – dla n obserwacji tworzy się hierarchię składającą się z k klas, $k = 1, 2, 3, \dots, n$. Niehierarchiczne (kombinatoryczne) oparte są na przyporządkowaniu n obiektów do zadanej wcześniej k liczby skupień.

W przypadku niehierarchicznych metod analizy skupień za najpopularniejsze algorytmy można przyjąć [7], algorytm k -średnich (ang. k -mean algorithm), algorytm EM (ang. algorithm EM – Expectation Maximization), FCM: algorytm k -średnich rozmytych (ang. Fuzzy c-Means).

W przypadku zastosowanego w prezentowanym artykule algorytmu k -średnich istotnym elementem jest określenie najbliższych środków skupień (centroidów). Do realizacji tego celu możliwe jest zastosowanie wielu miar odległości elementów zbiorów danych. Do najpopularniejszych metod wybranych z [7, 10] należą m.in.:

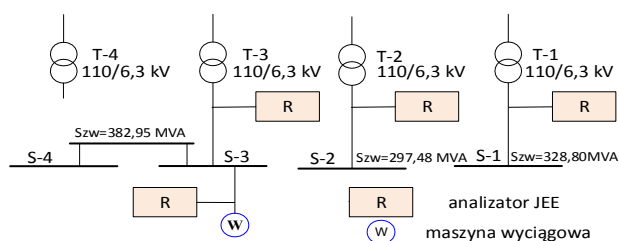
- odległość kosinusowa,
- odległość potęgowa,
- odległość euklidesowa,
- kwadrat odległości euklidesowej,
- odległość miejska (Manhattan),
- odległość Czebyszewa.

W badanym przypadku wybrano odległość Czebyszewa wyznaczaną następująco:

$$(1) \quad OC(x,y) = \text{Maksimum} |x_i - y_i|$$

gdzie: $OC(x,y)$ – odległość Czebyszewa między wektorami x i y , x_i – wektor obserwacji (elementów) należących do skupienia x , y_i – wektor obserwacji (elementów) należących do skupienia y .

Wybór odległości Czebyszewa podyktowany jest ukierunkowaniem na wskazanie maksymalnych różnic między grupami (skupieniami), które reprezentują różne uwarunkowania pracy sieci. Wybór metody poprzedzono badaniami wstępnymi dla różnych definicji odległości w środowisku Statistica. Wykonano analizy dla różnej kombinacji liczby końcowej skupień, odległości i sposobu doboru początkowych środków skupień.



Rys.1. Uproszczony schemat badanej sieci zakładów górniczych traktowanej jako obszar pomiarowy

Opis obiektu badań oraz źródło danych pomiarowych

Wykorzystane w artykule dane pomiarowe JEE pochodzą z elektroenergetycznych sieci zakładów górniczych wydobywających miedź [11]. Badania przeprowadzono z wykorzystaniem przyrządów klasy A w okresie od 27.04.2017-28.06.2019 w czterech punktach pomiarowych średniego napięcia, w tym trzech polach transformatorów oraz jednym polu odpływowym znaczącego odbioru - maszyny wyciągowej (MW). Uproszczonego schematu badanego fragmentu sieci wraz z lokalizacją rejestratorów został zaprezentowany na rysunku 1. Pomiarów prowadzono synchronicznie z wykorzystaniem systemu GPS.

W warunkach normalnej pracy sekcje 1 i 2 (S-1, S-2) pracują niezależnie natomiast sekcje 3 i 4 są sprzęgnięte. Badana sieć charakteryzuje się zmiennym obciążeniem pochodzącym przede wszystkim od maszyn wyciągowych, pomp odwadniania kopalń, wentylatorów, przenośników taśmowych oraz zmiennym poziomem generacji rozproszonej. Do lokalnych źródeł energii elektrycznej należą trzy bloki gazowo-parowe o mocach odpowiednio 15 MW, 13,5 MW, 10 MW. Lokalizację źródeł generacji oraz miejsca instalacji rejestratorów przedstawiono na rysunku 1. Taka sieć w ramach artykułu traktowana jest jako pewnego rodzaju obszar – sieć zasilająca jeden zakład.

Wyniki identyfikacji danych zagregowanych zawierających zdarzenia napięciowe

Podczas pomiarów odnotowano w każdym z punktów pomiarowych zdarzenia napięciowe. Zgodnie z regułą oznaczania (ang. flagging concept) w rozumieniu normy PN EN 61000-4-30 [2], jeżeli podczas danego przedziału czasu agregacji wystąpi zdarzenie typu zapad napięcia, wzrost lub przerwa, to każdą zagregowaną wartość parametrów jakości energii elektrycznej zawierającą ten znacznik czasowy należy oznaczyć i nie uwzględniać w ocenie jakości energii. Dodatkowo na potrzeby badań wrażliwości algorytmu analizy skupień poszerzono wskazaną grupę zdarzeń warunkujących zastosowanie oznaczania o szybkie zmiany napięcia [11, 12].

Wskazana reguła oznaczania obejmuje czasy agregacji właściwe dla poszczególnych parametrów jakości energii (10 sekund, 10 minut, 2 godziny) [2]. W związku z tym, że większość parametrów jakości energii wykorzystuje agregację 10-minutową [13] na potrzeby artykułu przyjęto, iż do celów identyfikacji danych podlegających regule oznaczania zastosowany zostanie zbiór parametrów 10-minutowych składający się z następujących parametrów:

- częstotliwość,
- wartość skuteczna napięcia,
- krótkookresowy wskaźnik migotania światła,
- asymetria napięcia,
- całkowity współczynnik zawartości harmonicznych w napięciu.

Przyjęcie wspólnego czasu agregacji 10-minut, pozwoliło uzyskać jednolity wymiar zbioru danych. Dla tak przygotowanego zbioru danych przeprowadzono w pierwszej kolejności analizę oznaczania wyników pomiaru w badanych punktach pomiarowych niezależnie z klasycznym podejściem zgodnym z przywołaną normą PN-EN 6100-4-30 [2]. Następnie dla danych z różnych punktów pomiarowych należących do wybranego fragmentu sieci elektroenergetycznej oznaczono dane, dla których w co najmniej w jednym punkcie pomiarowym oznaczono wartości parametrów JEE na skutek zdarzenia dynamicznego.

Omawiane podejście ma charakter rozszerzenia metody klasycznej przez synchronizację oznaczania danych od zdarzeń odnotowanych w innych punktach pomiarowych tej

samej sieci. Jest to możliwe pod warunkiem zastosowania synchronizacji czasowej rejestracji w punktach pomiarowych zastosowanej sieci, co w przypadku omawianych badań terenowych miało miejsce, dzięki zastosowaniu systemu GPS. Wtedy na podstawie zdarzeń odnotowanych w poszczególnych punktach pomiarowych możliwe jest oznaczenie znacznika czasowego lub okresu, na podstawie którego oznaczone zostaną dane we wszystkich punktach pomiarowych wspólnie tj. synchronicznie. Uzyskana łączność oznaczania danych we wszystkich punktach pomiarowych może mieć znaczenie praktyczne w przypadku analizy fragmentu sieci.

Jako alternatywną metodą oznaczania danych w sposób synchroniczny zaproponowano analizę skupień. Synchronizację oznaczania uzyskano przez scalenie zbiorów danych z poszczególnych punktów pomiarowych pracujących synchronicznie w jeden wspólny zbiór, a następnie wyznaczenie skupień zadaną liczbą skupień równą 2. Do podziału danych na oznaczone i nieoznaczone zastosowano zbiór danych zawierający 10-minutowe dane pomiarowe częstotliwości, wartość skutecznej napięcia, krótkookresowego wskaźnika migotania światła, współczynnika asymetrii napięcia, całkowitego współczynnika zawartości harmonicznych w napięciu, poddane standaryzacji zgodnie z wzorem (2):

$$(2) \quad x_i' = \frac{x_i - \bar{x}_i}{\sigma}$$

gdzie: x_i – wartość zmiennej, x_i' – wartość zmiennej znormalizowana, \bar{x}_i – wartość średnia zbioru X, σ – odchylenie standardowe danych zbioru X.

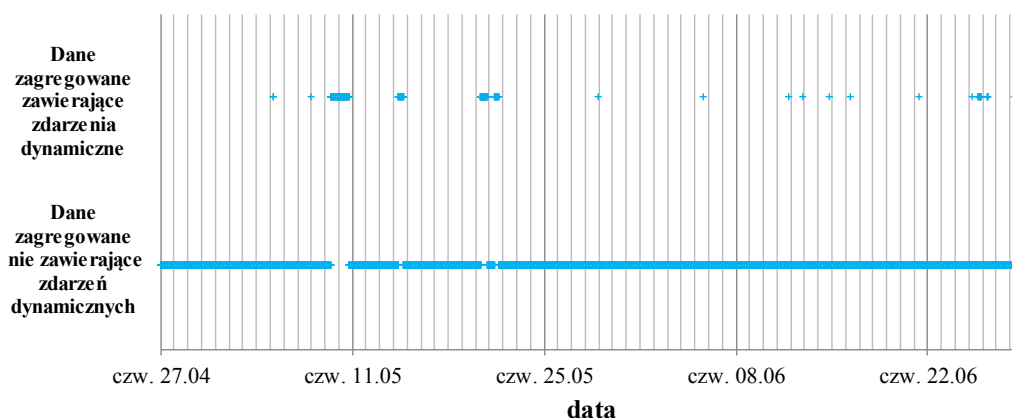
Zastosowano analizę skupień z zadaną liczbą skupień równą 2 wybrano, algorytm k-średnich oraz odległość Czebyszewa. Na podstawie tak przeprowadzonej analizy uzyskano podział wszystkich danych pomiarowych ze wszystkich punktów pomiarowych na dwa skupienia tj.:

- skupienie 1: reprezentujące zbiór danych nieoznaczonych, tj. dla których w żadnym punkcie pomiarowym nie wystąpiło zdarzenie napięciowe,
- skupienie 2: reprezentujące zbiór danych oznaczonych, tj. dla których w co najmniej jednym punkcie pomiarowym wystąpiło zdarzenie napięciowe.

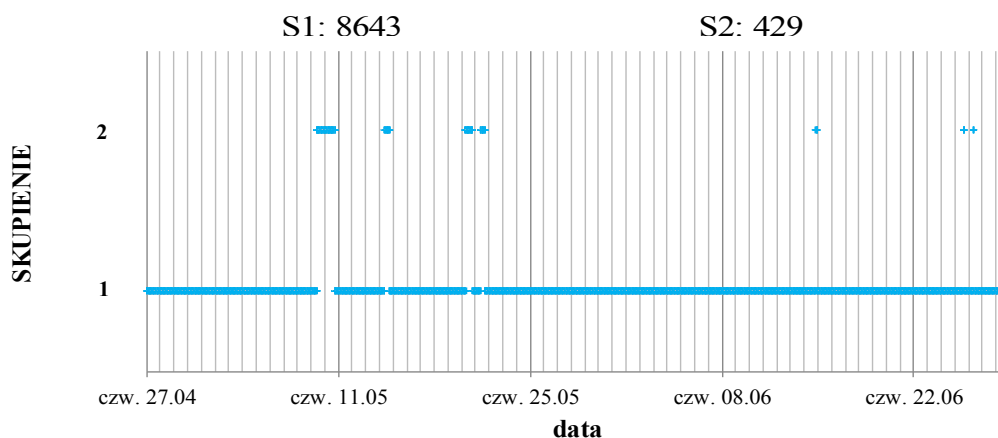
W celu porównania otrzymanych rezultatów metodą klasyczną i metodą analizy skupień przedstawiono wyniki oznaczania danych uzyskane przez odnotowanie wspólnych znaczników czasowych lub okresów, w których odnotowano zdarzenie w co najmniej jednym punkcie pomiarowym na podstawie klasyfikacji zdarzeń użytych rejestratorów (Rys. 2) oraz jako wynik zastosowania analizy skupień (Rys. 3).

Dane zagregowane nie zawierające zdarzeń dynamicznych: 8656

Dane zagregowane zawierające zdarzenia dynamiczne: 416



Rys.2. Wyniki oznaczania danych metodą klasyczną na podstawie identyfikacji zdarzeń w poszczególnych punktach pomiarowych (dolny zbiór reprezentuje dane nie zawierające zdarzeń, górny zbiór dane spełniające regułę oznaczania)



Rys.3. Wyniki oznaczania danych z wykorzystaniem analizy skupień danych zebranych z wszystkich punktów pomiarowych (dolny zbiór reprezentuje pierwsze skupienie S1 zawierające dane wolne od zdarzeń napięciowych, górny zbiór przedstawia drugie skupienie S2 zawierający dane, w trakcie których nastąpiły zdarzenia napięciowe)

Tabela 1. Wyniki oznaczania danych metodą klasyczną oraz z wykorzystaniem analizy skupień

Metoda	Łączna liczba danych	Liczba danych nieoznaczonych	Liczba danych oznaczonych
Klasyczna	9072	8656	416
Analiza skupień	9072	8643	429

Wyniki porównania procesu oznaczania danych metodą klasyczną oraz z wykorzystaniem analizy skupień zaprezentowano w Tab. 1.

Na podstawie przedstawionej analizy można stwierdzić, że analiza skupień z zastosowaniem algorytmu K-średnich, odległości Czebyszewa z zadaną wejściową ilością skupień równą 2 może być narzędziem wykorzystywanym do jednoczesnego (synchronicznego) oznaczania danych we wszystkich punktach pomiarowych na skutek zdarzenia odnotowanego w co najmniej w jednym punkcie pomiarowym. Skuteczność oznaczania danych metodą analizy skupień w odniesieniu do metody klasycznej wykorzystującej klasyfikację wynosi średnio 96.9%. Pozwala to stwierdzić, iż rezultat zastosowanie analizy skupień jako alternatywnego narzędzia identyfikacji synchronicznej danych oznaczonych w porównaniu do metody klasycznej można uznać za bardzo dobry.

Wnioski

W artykule zaprezentowano możliwość wykorzystania analizy skupień do identyfikacji oznaczonych wyników pomiaru jakości energii elektrycznej w ujęciu obszarowym. Wyniki uzyskane z zastosowaniem proponowanej metody porównano z wynikiem identyfikacji danych oznaczonych uzyskanych metodą klasyczną tj. zgodnie z koncepcją oznaczania przedstawioną w normie PN-EN 61000-4-30. Uzyskana średnią skuteczność metody na poziomie 96% dla rozszerzonego scenariusza identyfikacji zdarzeń tj. uwzględniającego zarówno zdarzenia typu zapad, wzrost lub przerwa, jak również szybkie zmiany napięcia. Warto podkreślić, iż w klasycznym podejściu identyfikację danych oznaczonych wykonuje się lokalnie dla każdego punktu pomiarowego. W przedstawionym podejściu dane z wszystkich punktów pomiarowych są scalane w jeden zbiór danych, a zastosowanie analizy skupień pozwala wyłonić dane oznaczone jednocześnie dla wszystkich punktów pomiarowych. Przedstawione rezultaty skłaniają do dalszych badań zaproponowanej metody nad możliwością zwiększenia wrażliwości identyfikacji oraz weryfikacji wyników na większej grupie danych pomiarowych z dłuższego okresu czasu oraz większego obszaru.

Autorzy: dr inż. Michał Jasiński, Politechnika Wrocławska, Katedra Podstaw Elektrotechniki i Elektrotechnologii, ul. Pl. Grunwaldzki 13, 50-370 Wrocław, E-mail: michal.jasinski@pwr.edu.pl; dr hab. inż. Tomasz Sikorski, prof. Uczelni, Politechnika Wrocławska, Katedra Podstaw Elektrotechniki i Elektrotechnologii, ul. Pl. Grunwaldzki 13, 50-370 Wrocław, E-mail: tomasz.sikorski@pwr.edu.pl; mgr inż. Klaudiusz Borkowski, KGHM Polska Miedź S.A., ul. Marii Skłodowskiej-Curie 48, 59-301 Lubin, E-mail: klaudiusz.borkowski@kghm.com.pl

LITERATURA

- [1] Lange A., Pasko M., Wybrane zagadnienia dotyczące jakości energii elektrycznej w kopalniach i hutach, *Przegląd Elektrotechniczny*, 88 (2012), nr. 5, 150–153
- [2] PN EN 6100-4-30: Kompatybilność elektromagnetyczna (EMC) - Część 4-30: Metody badań i pomiarów-- Metody pomiaru jakości energii, (2015)
- [3] Witten I.H., Frank E., *Data Mining: Practical Machine Learning Tools and Techniques*, (2011)
- [4] Fayyad U., Piatetsky-Shapiro G., Smyth P., From Data Mining to Knowledge Discovery in Databases, *AI Magazine*, 17 (1996), DOI: 10.1609/aimag.v17i3.1230
- [5] Hand D., Mannila H., Smyth P., *Principles of data mining* (2001), DOI:10.2165/00002018-200730070-00010
- [6] Kantardzic M., *Data Mining: Concepts, Models, Methods, and Algorithms: Second Edition*, *Data Mining: Concepts, Models, Methods, and Algorithms: Second Edition*, (2001) DOI:10.1002/9781118029145
- [7] Wierchoń S., Kłopotek, M., *Algorithms Of Cluster Analysis, Institute of Computer Science Polish Academy of Sciences*, (2015), http://rbc.ipipan.waw.pl/Content/40/Monografie_cz3_Wierchon_Kłopotek_ostateczny.pdf
- [8] Cabena P., Hadjinain P., Stadler R., Verhees J., Zanasi A., *Discovering data mining: from concept to implementation*. NJ: Upper Saddle River (New Jersey) : Prentice Hall PTR, (1998)
- [9] Jain A. K., Murty M. N., Flynn P. J., *Data clustering: a review, ACM Computing Surveys*, 31 (1999), 264–323, DOI:10.1145/331499.33150
- [10] Statsoft Polska, "StatSoft Electronic Statistic Textbook", (2016)
- [11] Jasiński M., Borkowski K., Sikorski T., Kostyla P., *Cluster Analysis for Long-Term Power Quality Data in Mining Electrical Power Network, Progress in Applied Electrical Engineering (PAEE)*, IEEE (2018). DOI:10.1109/PAEE.2018.8441089
- [12] Jasiński M., Sikorski T., Borkowski K., *Clustering as a tool to support the assessment of power quality in electrical power networks with distributed generation in the mining industry, Electric Power Systems Research*, 166 (2019), 52–60, DOI: 10.1016/j.epsr.2018.09.020
- [13] Jasiński M., Sikorski T., Kostyla P., Kaczorowska D., Leonowicz Z., Rezmer J., Szymańda J., Janik P., Bejmert D., Rybiański M., Jasińska E., *Influence of Measurement Aggregation Algorithms on Power Quality Assessment and Correlation Analysis in Electrical Power Network with PV Power Plant, Energies*, 12 (2019), 3547, DOI:10.3390/en12183547.