



Arkadiusz Kisiolek

Uniwersytet Ekonomiczny w Katowicach
Wydział Ekonomii
Katedra Informatyki Ekonomicznej
arkadiusz.kisiolek@ue.katowice.pl

ANALIZA WPISÓW NA PORTALU TWITTER Z WYKORZYSTANIEM NARZĘDZI BIG DATA ZAWARTYCH W PAKIECIE R

Streszczenie: Wraz z rozwojem internetu, mediów społecznościowych oraz technologii mobilnych znacznie wzrosła ilość generowanych danych. Dane te, zarówno w formie ustrukturalizowanej, jak i nieustrukturalizowanej, mogą nieść wartość biznesową dla przedsiębiorców. W danych big data można znaleźć m.in. informacje na temat klientów, konkurencji, rynku pracy, opinii na temat produktów danej firmy, czy aktualnych trendów. Dzięki dokładnej analizie internetu i mediów społecznościowych, interesariusze mogą pozyskać nową wartość, jaką są informacje na temat nastawienia i opinii konsumentów. Celem artykułu jest przedstawienie narzędzi big data jako jednego ze sposobów analizy mediów społecznościowych i wyciągania w ten sposób wartościowych informacji. Przedmiotem przeprowadzonej analizy były tysiące tweetów użytkowników portalu Twitter. Analiza została przeprowadzona przy wykorzystaniu technik text mining oraz sentyment analysis.

Słowa kluczowe: big data, social media, internet, text mining, analiza danych.

JEL Classification: C38, J24.

Wprowadzenie

Aby sprostać rosnącym wymaganiom klientów, organizacje poszukują nowych źródeł danych opisujących zachowania i preferencje konsumentów [Lawrence i in., 2010]. Ciekawym źródłem informacji wydają się media społecznościowe [Katal, Wazid, Goundar, 2013]. Równocześnie rozwój różnorodnych narzędzi do przetwarzania big data spowodował, że media społecznościowe stały się przedmiotem licznych analiz [He, Zha, Li, 2013].

Celem artykułu jest przedstawienie narzędzi big data w analizie mediów społecznościowych przy wykorzystaniu metod analizy wydźwięku wypowiedzi oraz text mining. Badania zostały przeprowadzone na podstawie danych pochodzących z portalu Twitter. Badania pozwalają na pozyskanie informacji dotyczącej tego, jakie słowa najchętniej były wykorzystywane przez użytkowników portalu Twitter do opisu marek Apple oraz Samsung, jak również tego, jakie jest nastawienie emocjonalne użytkowników do tych marek – pozytywne, neutralne czy negatywne.

1. Big data jako źródło danych

Ilość danych w sieci ciągle rośnie [Olszak, Gajowska, 2017]. Dane mogą być wytwarzane przez różnorodne systemy, czy urządzenia, np. sieci społecznościowe, Internet of Things, systemy sprzedażowe, sklepy internetowe, czy GPS [Pedrycz, Chen, 2015].

Rosnąca ilość danych skutkuje tym, że przedsiębiorstwa są zainteresowane coraz dokładniejszą eksploracją internetu i wyciąganiem wniosków na podstawie danych, które do tej pory nie były brane pod uwagę. Jednym z obszarów analiz mogą być portale społecznościowe. Portale te mogą być wykorzystywane przez użytkowników m.in. do wyrażania opinii na temat produktów, oceniania usług, czy poszukiwania pracy [Szewczyk, 2011]. Informacje zawarte na portalach społecznościowych są jednym z klasycznych przykładów zasobów big data.

Big data może być definiowana jako obszerny zbiór danych pozyskiwanych z różnorodnych obszarów. Jest to określenie mające zastosowanie do zbiorów danych o dużej objętości, różnorodności, strumieniowym napływie w czasie rzeczywistym, a także zmienności i złożoności [Sagiroglu, Sinanc, 2013]. Dane big data mogą być [Russom, 2011; Sagiroglu, Sinanc, 2013]:

- ustrukturalizowane – obejmują zbiory posiadające określoną strukturę, np. tradycyjne bazy danych,
- nieustrukturalizowane – dane, które nie posiadają określonej struktury. Do tego typu danych można zaliczyć np. język mówiony, pliki wideo, czy obrazy,
- semi-strukturalne – są podobne do strukturalnych, jednak są rozszerzone o dodatkowe informacje niestukturalne. Do tej kategorii można zaliczyć dane przechowywane w formacie XML, RSS, czy JSON.

Określenie big data może być stosowane do specyficznych zbiorów danych. Zbiory te muszą charakteryzować się pewnymi cechami. Wśród nich można wyróżnić tzw. koncepcję 5V [Olszak, 2016]. Charakterystyka cech składających się na koncepcję 5V została przedstawiona w tabeli 1.

Tabela 1. Charakterystyka cech składowych koncepcji 5V

Cecha	Charakterystyka
Volume	Ilość danych mierzona w petta- i zettabajtach
Velocity	Ogromna prędkość napływu danych i konieczność ich analizy w czasie rzeczywistym
Variety	Bardzo duża różnorodność danych (tekst, obraz, film, dźwięk, XML). Pochodzenie z różnego typu urządzeń i aplikacji (komputery, serwery, czujniki, urządzenia mobilne, social media, internet)
Veracity	Dane mogą być niespójne, niekompletne oraz niedokładne
Value	Wartość ukryta w danych

Źródło: Opracowanie własne na podstawie: [Manyika i in., 2011; Katal, Wazid, Goudar, 2013; Erl, Khattak, Buhler, 2015; Olszak, 2016].

Obecnie niektórzy autorzy rozszerzają koncepcję 5V o dwie dodatkowe składowe: Variability, Virtual Data [www 1].

Big data jest związana z intensywnym rozwojem i wzrostem popularności sieci społecznościowych. Sieci te są wykorzystywane do gromadzenia informacji o użytkownikach w celu coraz lepszej personalizacji przekazu marketingowego [Hsu, Chang, Hsu, 2017]. Motywacją do tego działania może być zwiększenie zysków, dotarcie do użytkownika zainteresowanego usługami, np. szukającego kredytu lub samochodu, czy obniżenie kosztów marketingowych poprzez redukcję nieefektywnych kanałów komunikacji marketingowej.

Big data ma zastosowanie w wielu branżach. Wśród nich można wymienić branżę detaliczną, bankowość, edukację, produkcję, prawo, telekomunikację, czy ochronę zdrowia [Olszak, Mach-Król, 2015; Frizzo-Barker, Chow-White, Mozafari, 2016; SAS].

Analiza big data obejmuje sztuczną inteligencję, analizę danych pochodzących z urządzeń mobilnych, sieci społecznościowych oraz Internet of Things. Analiza dużych zbiorów danych jest możliwa przy wykorzystaniu różnego rodzaju technik, wśród których można wymienić: text mining (analiza danych tekstowych), uczenie maszynowe, analizy predyktywne, drażnienie danych, statystyka oraz przetwarzanie języka naturalnego [IBM].

Przedstawiając informacje na temat big data, należy pamiętać o technologiach związanych z tą kategorią. W tabeli 2 zostały przedstawione wybrane narzędzia wykorzystywane w kontekście analizy, przetwarzania, przechowywania, czy innych działań wykonywanych na dużych zbiorach danych.

Tabela 2. Charakterystyka wybranych narzędzi big data

Nazwa narzędzia	Charakterystyka
Big Table	System rozproszonej bazy danych, zbudowany w systemie plików Google
Cassandra	System zarządzania bazami danych o otwartym kodzie przeznaczony do obsługi ogromnych ilości danych w systemie rozproszonym
Google File System	Autorski rozproszony system plików opracowany przez Google – część inspiracji dla Hadoop
Hadoop	Platforma oprogramowania open source do przetwarzania ogromnych zbiorów danych dotyczących niektórych problemów w systemie rozproszonym
Hbase	Rozproszona, nierelacyjna baza danych typu open source wzorowana na Big Table Google
MapReduce	Platforma programowa wprowadzona przez Google do przetwarzania ogromnych zbiorów danych dotyczących określonych problemów w systemie rozproszonym
Mashup	Aplikacja, która wykorzystuje i łączy prezentację danych lub funkcjonalność z dwóch lub więcej źródeł do tworzenia nowych usług

Źródło: [Olszak, 2016].

2. Media społecznościowe jako przykład big data

Każdego dnia miliony użytkowników korzystają z mediów społecznościowych z różnorodnych powodów. Charakterystyczną cechą tego typu platform jest to, że skupiają one wielu użytkowników wokół siebie.

Istnieje kilka podstawowych rodzajów mediów społecznościowych. Można wśród nich wymienić [Cavazza, 2017]:

- media społecznościowe nastawione na nawiązywanie kontaktów między użytkownikami, tzw. Networking, np. LinkedIn, Badoo, Meetup,
- media służące udostępnianiu treści i dzieleniu się nimi z innymi użytkownikami, np. YouTube, Instagram, Spotify,
- media, których celem jest umożliwienie kontaktu między użytkownikami poprzez wysyłanie wiadomości tekstowych, zdjęć lub filmów, np. Messenger, Viber, WhatsUp,
- media nastawione na publikację treści przez użytkowników, np. Wordpress, Wikipedia, mySpace,
- media ułatwiające i organizujące współpracę między użytkownikami, np. Dropbox, Reddit, Slack,
- platformy umożliwiające komunikację wideo między użytkownikami, np. Skype, Google Hangouts,

Raport Social Media Landscape 2017 przedstawia aktualny podział mediów społecznościowych wraz z przykładami najczęściej używanych portali (zob. rys. 1).



Rys. 1. Podział mediów społecznościowych

Źródło: [Cavazza, 2017].

Jak można zauważyć, w obecnych czasach istnieje bardzo wiele portali społecznościowych, które udostępniają różnorodne funkcjonalności. Część z nich należy także do kilku grup. Przykładami mogą być Facebook, Twitter, czy Google.

3. Wybrane metody i techniki analizy mediów społecznościowych

Analiza mediów społecznościowych może być dokonywana przy wykorzystaniu różnego rodzaju technik. Wśród nich można wyróżnić Social Media Mining, który jest procesem reprezentacji, analizy i wydobywania wzorców na podstawie danych pochodzących z Social Media [Zagarani, Abbasi, Liu, 2014].

Kolejną metodą mającą zastosowanie w analizie mediów społecznościowych jest także text mining, który pozwala na analizę danych tekstowych pochodzących z różnych źródeł [Olszak, Zurada, 2015]. Wśród przykładów można wymienić: komunikację mailową, dokumenty internetowe, jak również dane pochodzące z portali społecznościowych [Chen i in., 2012].

Jednym z narzędzi wykorzystywanych w analizach tekstu jest tzw. analiza wydźwięku wypowiedzi. Analiza ta ma za zadanie określić opinie, postawy i emocje ludzi dotyczące danego tematu [Medhat, Hassan, Korashy, 2014]. Efektem analizy sentymentalnej jest sklasyfikowanie wyrażen, np. pod względem tego, czy są pozytywne, negatywne, czy neutralne, oraz siły tych emocji.

4. Metodyka badawcza

4.1.1. Informacje o zastosowanej metodzie badawczej

Celem przeprowadzonych badań jest przedstawienie wybranych narzędzi big data jako metod analizy informacji płynących z mediów społecznościowych. Badania pozwoliły na znalezienie najczęściej powtarzających się słów na dany temat, a także na określenie, jakie jest nastawienie konsumentów do niego. Wiedza ta może wpłynąć na poprawę procesu obsługi klienta, lepsze dostosowanie produktu do jego potrzeb bądź zweryfikować, czy przeprowadzona kampania promocyjna uzyskała aprobatę wśród odbiorców.

Badania zostały przeprowadzone na podstawie danych pochodzących z serwisu Twitter, który jest specyficznym narzędziem służącym do mikroblogowania. Na potrzeby niniejszej pracy przeanalizowano frazy użyte przez użytkowników serwisu społecznościowego w tweetach na temat produktów firm Samsung oraz Apple. Do badań wykorzystano platformę programistyczną R.

4.1.2. Podstawowe informacje na temat przeprowadzonego badania

Badania zostały przeprowadzone w okresie 10.10.2017–18.11.2017 na próbie około 6000 tweetów. Dane do analizy zostały pozyskane przy wykorzystaniu programu R. Ważnym elementem analizy było wstępne przetworzenie danych, które z jednej strony ogranicza próbę badawczą, jednak z drugiej pozwala oczyścić dane, co skutkuje wyodrębnieniem istotnych dla badania informacji w odpowiedniej formie. Badania pozwoliły na określenie, które słowa na temat #Apple oraz #Samsung były najczęściej wykorzystywane przez internautów oraz jaki był wydźwięk tych wypowiedzi.

5. Analiza portali społecznościowych z wykorzystaniem narzędzi big data

Na potrzeby niniejszego opracowania wykorzystano dane z serwisu Twitter w postaci tzw. tweetów użytkowników. Tweety zostały zebrane za pomocą pakietu R („*twitteR*” library). Następnie dane zostały wstępnie przetworzone i połączone w tzw. korpus. Dane zostały także oczyszczone, np. usunięto znaki przestankowe i liczby oraz usunięto słowa, które nic nie wnoszą do analizy, np. spójniki. Kolejnym etapem było drażnienie danych.

Na podstawie analizy można wyodrębnić słowa najczęściej powtarzające się. Na rys. 2 przedstawiono 20 najczęściej używanych tweetów w odniesieniu do #Samsung, zaś na rys. 3 dla #Apple.

[1]	"black"	"smart"	"blackfriday"	"sale"	"refurbished"
[6]	"smartphone"	"price"	"screen"	"white"	"plus"
[11]	"best"	"gold"	"tablet"	"technology"	"discount"
[16]	"unlock"	"amazon"	"youtube"	"uhd"	"buy"

Rys. 2. 20 najczęstszych tweetów dla #Samsung

Źródło: Opracowanie własne na podstawie danych pobranych z serwisu Twitter.

[1]	"information"	"homepod"	"check"	"deal"	"listen"
[6]	"music"	"amazon"	"macbook"	"itunes"	"series"
[11]	"diversity"	"speaker"	"perfect"	"black"	"icloud"
[16]	"silver"	"iwatch"	"price"	"win"	"technology"

Rys. 3. 20 najczęstszych tweetów dla #Apple

Źródło: Opracowanie własne na podstawie danych pobranych z serwisu Twitter.

Pakiet R oferuje także graficzne metody prezentacji danych w postaci tzw. chmury słów, która zostanie przedstawiona poniżej.

Ze zbiorów słów zaprezentowanych na rys. 2 i 3 można wnioskować, iż klienci są zadowoleni z produktów firmy Samsung oraz szukają promocji na produkty oferowane przez firmę. W kontekście firmy Apple można zauważyć kilka postów na temat muzyki. Posty te są najprawdopodobniej związane z premierą nowego produktu tej firmy – głośnika HomePod [Apple]. Można więc stwierdzić, że temat ten jest interesujący dla użytkowników portalu Twitter.

Pakiet R daje także możliwość połączenia chmur słów w celu dokonania analizy porównawczej. Chmura porównawcza (tzw. *comparison cloud*) została przedstawiona na rys. 4.

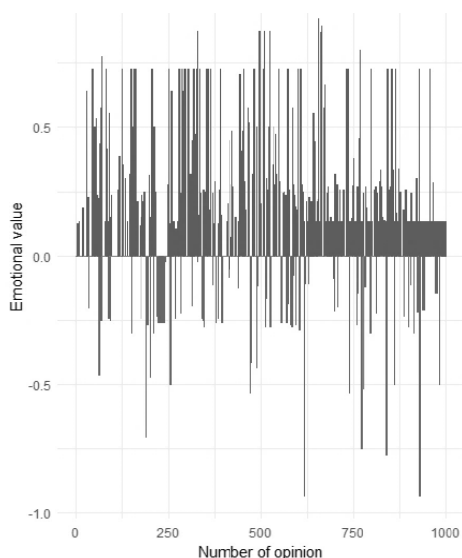


apple

Źródło: Opracowanie własne na podstawie danych pobranych z serwisu Twitter.

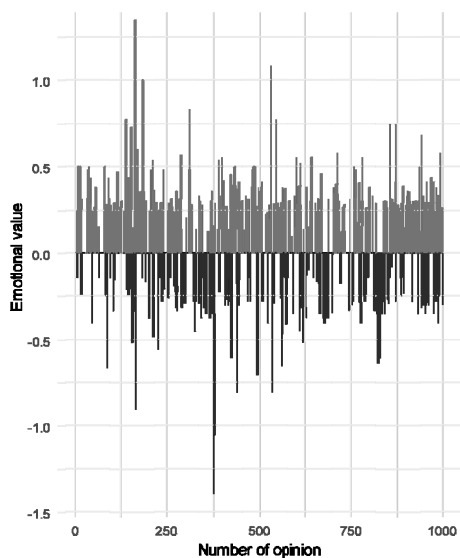
Podstawową zaletą tej metody analizy danych jest bardzo czytelne i przystępne przedstawienie wyników.

Wartości wag mogą być ujemne lub dodatnie. Opinie negatywne otrzymują wartości ujemne, zaś pozytywne – dodatnie. Opinie neutralne otrzymują wartość 0. Wyniki analizy można zobaczyć na rys. 5 i 6.



Rys. 5. Analiza opinii dla #Samsung z podziałem na pozytywne, neutralne oraz negatywne

Źródło: Opracowanie własne na podstawie danych pobranych z serwisu Twitter.



Rys. 6. Analiza opinii dla #Apple z podziałem na pozytywne, neutralne oraz negatywne

Źródło: Opracowanie własne na podstawie danych pobranych z serwisu Twitter.

W przypadku wypowiedzi na temat #Samsung amplituda pozytywnych emocji jest wyraźnie wyższa niż w przypadku #Apple. Pakiet R pozwala także na odnalezienie poszczególnych tweetów i wartości ich rang. Gotowy dokument jest eksportowany do pliku HTML, co pozwala na proste i szybkie wyszukanie interesujących informacji.

Analizy tego typu mogą być pomocne w procesie zarządzania relacjami z klientami i reagowania na zmiany, a także sprawdzenia, czy oferowane produkty spełniają oczekiwania konsumentów. Powyższe analizy mogą być także zastosowane do sprawdzenia, jakie są opinie na temat nowo wprowadzonego produktu na rynek, a więc wykorzystanie tego typu analiz wydaje się dość ciekawe z punktu widzenia analityków rynku.

Na podstawie przeprowadzonej analizy można stwierdzić, iż jednym z największych wyzwań w procesie analizy tekstu jest jego oczyszczenie ze słów, które nie stanowią wartości dodanej w procesie analizy. Należy także zwrócić uwagę na fakt, iż analizy są przeprowadzane za pomocą pewnego algorytmu, przez co istnieje zagrożenie błędnej interpretacji części wskaźników, ze względu na fakt występowania kontekstu wypowiedzi.

Z przeprowadzonej analizy można wyciągnąć następujące wnioski:

- większość opinii dotyczących badanych marek była pozytywna, jednak pojawiły się także opinie negatywne, o różnym stopniu emocjonalności,
- nastawienie emocjonalne użytkowników w badanych postach jest różne dla obydwu marek,
- wiele tweetów ma charakter reklamowy bądź konkursowy, zaś nie są opinią użytkowników,
- użytkownicy są zainteresowani nowymi produktami obu firm.

Podsumowanie

Big data jest coraz chętniej wykorzystywana na świecie. Dzięki rosnącej ilości danych analitycy mają możliwość wyciągania bardziej wiarygodnych informacji i dostosowywania oferty do specyficznych potrzeb użytkowników. Jednym z przykładów tego typu analiz może być tzw. text mining, czyli drążenie tekstu. Pozwala ono na wyciąganie wartościowych informacji z danych tekstowych, np. pochodzących z portali społecznościowych. Przeprowadzona analiza pozwoliła na odnalezienie najczęściej występujących słów z podziałem na kategorie. Zbadano także nastawienie użytkowników do marek na podstawie zamieszczanych przez nich wpisów. Niniejsze badania wskazują, że narzędzia big data, a zwłaszcza text mining, sentyment analysis, sprawdzają się w analizie opinii użytkowników portali społecznościowych.

Przeprowadzone badania zwracają uwagę na pewne ograniczenia w stosowaniu technik text mining, sentyment analysis. Wśród nich należy wymienić fakt, że techniki te opierają się na algorytmach, które mogą pomijać (nie uwzględniać) kontekst wypowiedzi. Dodatkowo należy pamiętać, że w analizach text mining bardzo ważna jest częstotliwość występowania słów i ona wpływa na ostateczny wynik. Rzadko stosowane frazy (jednak występujące) mogą zostać nieuwzględnione lub zmarginalizowane w całościowej analizie.

Zagadnienie analizy portali społecznościowych z wykorzystaniem metod big data jest obszernym i złożonym problemem. W ramach niniejszego artykułu omówiono tylko wybrane narzędzia z zakresu text mining, sentyment analysis, które mogą być stosowane do analizy opinii użytkowników. W kolejnych pracach autor zamierza skupić uwagę na analizach dotyczących powiązań między użytkownikami, ich zaangażowania w wybrane aktywności, a także na eksploracji ich szczegółowych zainteresowań.

Literatura

- [Apple] www.apple.com/homepod (dostęp: 19.12.2017).
- Cavazza F. (2017), *Social Media Landscape 2017*, <https://fredcavazza.net/2017/04/19/social-media-landscape-2017> (dostęp: 22.12.2017).
- Chen H., Chiang R., Storey V. (2012), *Business Intelligence and Analytics: From Big Data to Big Impact*, „MIS Quarterly”, Vol. 36, No. 4, s. 1165-1188.
- Erl T., Khattak W., Buhler P. (2015), *Big Data Fundamentals: Concepts, Drivers & Techniques*, Prentice Hall, Boston.
- Frizzo-Barker J., Chow-White P.A., Mozafari M. (2016), *An Empirical Study of the Rise of Big Data in Business Scholarship*, „International Journal of Information Management”, Vol. 36, No. 3, s. 403-413.
- He W., Zha S., Li L. (2013), *Social Media Competitive Analysis, Text Mining: A Case Study in the Pizza Industry*, „International Journal of Information Management”, Vol. 33, No. 3, s. 464-472.
- Hsu H., Chang C., Hsu C. (2017), *Big Data Analytics for Sensor-Network Collected Intelligence 1st Edition*, A volume in Intelligent Data-Centric Systems, Academic Press, 2017, s. 1- 306.
- [IBM] *Big Data Analytics. Employ the Most Effective Big Data Technology*, www.ibm.com/analytics/hadoop/big-data-analytics (dostęp: 10.01.2018).
- Katal A., Wazid M., Goudar R.H. (2013), *Big Data: Issues, Challenges, Tools, Good Practices*, http://www.stat.purdue.edu/~doerge/BIOINFORM.D/SPRING16/KatalWazidGoudar_2013.pdf (dostęp: 27.12.2017).
- Lawrence R., Melville P., Perlich C., Sindhwani V., Meliksetian S., Hsueh P.-Y., Liu Y. (2010), *Social Media Analytics, The Next Generation of Analytics – Based Marketing Seeks Insights from Blogs*, https://www.researchgate.net/publication/235769780_Social_Media_Analytics-The_Next_Generation_of_Analytics-Based_Marketing_Seeks_Insights_From_Blogs (dostęp: 20.12.2017).
- Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Byers A.H. (2011), *Big Data: The Next Frontier for Innovation, Competition, Productivity*, McKinsey Global Institute, https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.ashx/ (dostęp: 06.10.2017).
- Medhat W., Hassan A., Korashy H. (2014), *Sentiment Analysis Algorithms and Applications: A Survey*, „Ain Shams Engineering Journal”, 5(4), s.1093-1113.
- Olszak C.M. (2016), *Toward Better Understanding, Use of Business Intelligence in Organizations*, „Information Systems Management”, Vol. 33, No. 2, s. 105-123.
- Olszak C.M., Gajowska D. (2017), *Big Data Approach to Analyzing the IT Job Market*, Proceedings of the 11 European Conference on Information Systems Management, University of Genoa, 14-15 September 2017, Genoa, Italy, Ed. R.P. Dameri, R. Spinelli, Published by Academic Conferences, Publishing International Reading, UE, s. 242-251.

- Olszak C.M., Mach-Król M. (2015), *Big Data: a New Value for Organizations* [w:] A. Fošner (ed.), *Advances in Business-Related Scientific Research Conference*, GEA College, Ljubljana.
- Olszak C.M., Zurada J. (2015), *Information Technology Tools for Business Intelligence Development in Organizations*”, „Polish Journal of Management Studies”, Vol. 12, No. 1, s. 138-139.
- Pedrycz W., Chen S.-M. (2015), *Information Granularity, Big Data, Computational Intelligence*, Springer, London.
- Russom P. (2011), *Big Data Analytics*, TDWI best practices report, The Data Warehousing Institute (TDWI) Research.
- Sagiroglu S., Sinanc D. (2013), *Big Data: A Review*, Collaboration Technologies, Systems (CTS), International Conference on: IEEE, s. 42-47.
- [SAS] *What is big data*, SAS, https://www.sas.com/pl_pl/insights/big-data/what-is-big-data.html#modal2 (dostęp: 19.12.2017).
- Szewczyk A. (2011), *Popularność funkcji serwisów społecznościowych*, „Zeszyty Naukowe Uniwersytetu Szczecińskiego. Studia Informatica”, nr 29.
- Zagarani R., Abbasi M., Liu H. (2014), *Social Media Mining: An Introduction*, Cambridge University Press, <http://dmml.asu.edu/smm/SMM.pdf> (dostęp: 20.12.2017).
- [www 1] *The 7 pillars of Big Data*, Petroleum Review, 2015 https://www.landmark.solutions/Portals/0/LMSDocs/Whitepapers/The_7_pillars_of_Big_Data_Whitepaper.pdf (dostęp: 23.12.2017).

SOCIAL MEDIA ANALYSIS WITH BIG DATA TOOLS

Summary: Development of Internet, social media and databases has caused a huge increase of data. Structured, semi-structured and unstructured data has a high business value. It contains various information about customers, competition, labor market, and development trends for industries, products and services.

The internet and social media are places where customers express their opinions about various products and services. It is a valuable source of information for entrepreneurs. The aim of this paper is to explore the issue of big data and to propose a set of different techniques for the analysis of customer opinions on the example of Twitter.

Keywords: big data, social media, internet, text mining.