

Dariusz Filip

Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie

WNHiS, Instytut Socjologii

d.filip@uksw.edu.pl

***Big Data* — narzędzie instytucji finansowych w dotarciu do klienta**

JEL Classification: M15, G21, G22

Keywords: *Big Data*, data analysis of large volume, 3Vs, financial institutions, competitive advantages

Abstract

Big Data: A Financial Institutions' Tool for Reaching Out to Clients

Big Data is an analytical tool which uses a large volume of data of various and complex structure; identifying, gathering, storing and transforming them as well as drawing appropriate conclusions for improving business activity requires certain skills. The process of finding non-obvious relations, mining hidden knowledge, creating visualization of results and business implementation by means of new technologies is referred to as data analysis of large data sets. The efficient application of the results from the advanced analyses in business models might help entities, including financial institutions, to adjust the offer to real and individual needs of the customer and thus to gain competitive advantage. The aim of the article is to systematize the terms and review the possibilities and scope of methods related to data analysis of large volume as well as to refer to the challenges in the context of Big Data enforcement.

1. Wprowadzenie

Peter Drucker stwierdził kiedyś, że kluczem do sukcesu gospodarczego w XXI wieku będzie demografia oraz informacja wynikająca z technologii (2009, s. 52 i 106). Obecnie nowoczesne technologie uważa się właśnie za wiodące kierunki działania całej gospodarki, a wielu ekonomistów postrzega zarówno wiedzę, jak i wynikającą z niej technologię jako samodzielny czynnik produkcji¹ (zob. Nonneman, Vanhoudt, 1996; Thurow, 2006). Ponadto, technologia może zmie-

¹ Mankiw, Romer i Weil (1992) do klasycznego modelu Solowa (1956) dodali kapitał ludzki jako czynnik produkcji, a z kolei Nonneman i Vanhoudt (1996) rozszerzyli wspomniany model dodatkowo o nowy czynnik produkcji — technologię — i wskazali jego wpływ na wzrost gospodarczy.

niać model biznesowy, a w konsekwencji jest w stanie w znaczący sposób oddziaływać na pozycje poszczególnych graczy na rynku, stwarzając przy tym nowe możliwości działania.

Jednocześnie jednak część badaczy (np. Carr, 2004) uważa, że technologia i kolejne inwestycje w działy IT (ang. *information technology*) nie stanowią obecnie o przewadze konkurencyjnej, ponieważ nowoczesne rozwiązania zostały wystandaryzowane. W krótkim okresie jedynie na początku, po wprowadzeniu nowoczesnego rozwiązania, nowe technologie dają pionierom znaczące przewagi, jednak wraz ze zwiększaniem dostępności, zaspokajaniem kolejnych potrzeb IT poszczególnych organizacji i obniżaniem kosztów stają się one dobrem powszechnym. Zarządzanie sferą IT powinno ograniczać się jedynie do redukcji zagrożeń, a nie zajmować się poszerzaniem możliwości działania.

Jednakże nie wszystkie elementy infrastruktury IT stały się powszechne. W ostatnim okresie pojawiło się sporo nowych obszarów związanych z innowacyjnością przedsiębiorstw i konkurencją na dynamicznie zmieniających się rynkach. Jak podaje Płoszajski (2013, s. 5), znajdujemy się w kolejnej fazie rewolucji komputerowej, w której najważniejszymi czynnikami są szerokopasmowy Internet oraz właśnie nieograniczone przetwarzanie wielkich zbiorów danych o zachowaniu obiektów w czasie realnym. W tym miejscu warto przytoczyć słowa J. Schoendorfa, eksperta Światowego Forum Ekonomicznego (WEF), który stwierdził podczas jednej z konferencji², że rozwój technologiczny zbliża się do „połowy szachownicy”³. Oznacza to, że kolejne tworzone modele biznesowe będą bazowały na informacji, która stanie się ekwiwalentem kapitału i pracy.

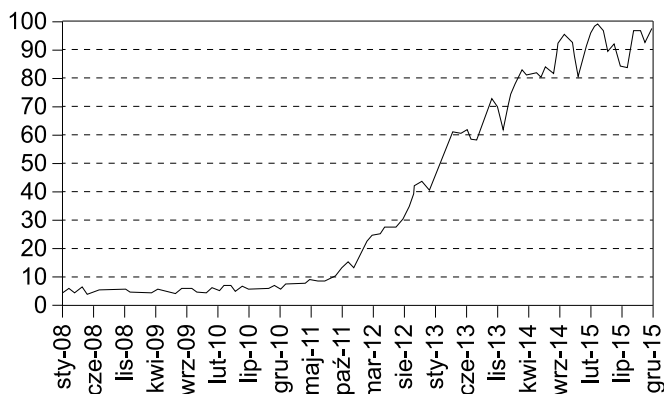
Obecnie każda organizacja zmaga się z niespotykanym wzrostem wielkości danych. Szacuje się, że od początku cywilizacji do 2003 roku ludzkość stworzyła tyle informacji, ile obecnie generowanych jest co dwa dni, tj. ok. 5 eksabajtów danych (EB, 10^{18} bajtów) (Dev, Patgiri, 2015, s. 538). Do 2012 roku wielkość danych wzrosła do blisko 2,72 zetabajtów (ZB, 10^{21} bajtów). Jak podają Sagiroglu i Sinanc (2013, s. 42), w najbliższym czasie wspomniane wielkości będą podwajane co dwa lata. Trudno jednak estymować, jak ten rozwój będzie wyglądał w kolejnej dekadzie.

Analiza dużej ilości danych staje się obszarem nauki, który niezwykle szybko się rozwija i zdobywa coraz większą popularność. Potwierdzeniem tego może być analiza trendów wyszukiwań, które ukazują zainteresowania użyt-

² 10. konferencja DLD (*Digital-Life-Design*), odbywająca się w Monachium w dniach 19–21 stycznia 2014 r. — branżowe forum poświęcone technologiom i innowacyjności.

³ Ilustracja ciągu geometrycznego pokazująca postęp technologiczny jako szachownicę. Każde kolejne pole oznacza podwojenie (np. szybkości mikroprocesorów) — według starej perskiej legendy o wynalazcy gry w szachy, który jako wynagrodzenia zażyczył sobie tyle ziaren zboża, ile może się nagromadzić: jeśli na pierwsze pole szachownicy położymy jedno, na drugie dwa, na trzecie cztery, na czwarte osiem i na każde następne z 64 pól szachownicy podwójną liczbę ziarenek w stosunku do położonych na poprzednim polu. Życzenie to okazało się nie do spełnienia, gdyż wymagana ilość zboża kilkakrotnie przewyższała możliwości zbiorów. Na samym ostatnim polu liczba ziaren wynosiła ponad 9 trylionów, a łączna liczba to 2^{64} , czyli ponad 18 trylionów.

kowników Internetu w danym okresie. Wykorzystując narzędzie, jakim jest serwis Google Trends, można ustalić zależność od czasu, ale również liczbę zapytań w wyszukiwarce, czyli samą popularność danego hasła. Trendy wyszukiwań terminu „Big Data” w okresie 2004–2015 zostały zaprezentowane na wykresie 1.



Wykres 1. Popularność terminu „Big Data” w zapytaniach Google

Źródło: opracowanie własne na podstawie Google Trends, <https://www.google.pl/trends/explore#q=big%20data> (dostęp: 01.2016).

Narzędzie pozwalające określić popularność danego hasła (Google Trends) w zależności od czasu uwzględnia wszelkie sezonowe trendy związane z częstotliwością wyszukiwania hasła „Big Data”. Jak wynika z wykresu 1, dynamiczny wzrost zainteresowania tym terminem jest odnotowywany od połowy 2011 roku, a jego popularność nie osiągnęła jeszcze swojej górnej granicy.

Skuteczne wykorzystywanie rezultatów zaawansowanych analiz w modelach biznesowych pomaga podmiotom gospodarczym, w tym instytucjom finansowym, w dostosowywaniu oferty do rzeczywistych i zindywidualizowanych potrzeb klienta, a tym samym w osiągnięciu przewagi konkurencyjnej. Biorąc to pod uwagę, niezwykle istotny wydaje się opis i wyjaśnienie możliwości, jakie dają narzędzia *Big Data*. Celem niniejszej pracy będzie usystematyzowanie pojęć oraz dokonanie przeglądu możliwości i zakresu metod związanych z analizą danych o wielkich wolumenach, jak również odniesienie się do wyzwań stojących za wdrożeniem *Big Data* w instytucjach finansowych.

Niniejszy artykuł składa się z kilku zasadniczych części. W części drugiej zostanie przedstawione pojęcie oraz istota *Big Data*. W części trzeciej omówione zostaną możliwości wykorzystania analizy danych o wielkich wolumenach w obszarze usług finansowych. Podsumowanie najważniejszych ustaleń stanowić będzie ostatnią część pracy.

2. Pojęcie i istota *Big Data*

Początkowo założenie dotyczące *Big Data* utożsamiane było z faktem, że wolumen informacji zwiększał się do takiego rozmiaru, że ich liczba nie była możliwa do przetworzenia w pamięci komputerów. W związku z tym zaistniała potrzeba zmiany wykorzystywanych narzędzi i dostosowywania ich do współczesnych potrzeb (Mayer-Schönberger, Cukier, 2013, s. 6). Podobnego podejścia w definiowaniu przedmiotowego terminu używa firma McKinsey, która nie określa limitu wielkości danych, np. w petabajtach [PB], podając jednak, że zarówno technologia ulega zmianie, jak i sama wielkość generowanych danych się zwiększa. Pierwsze definicje dotyczące wielkich wolumenów mówiły zatem o zbiorach, których rozmiar przekraczał możliwości klasycznych narzędzi do składowania, analizy i zarządzania danymi (MCI, 2011, s. 1).

Jedna z najczęściej przywoływanych w literaturze przedmiotu definicji⁴ określa jednak *Big Data* jako zbiory informacji o dużej objętości, dużej zmienności oraz dużej różnorodności, które wymagają innowacyjnych form przetwarzania w celu wspomagania podejmowania efektywnych decyzji, odkrywania nowych zjawisk oraz automatyzacji procesów. We wspomnianej konwencji zawarte jest podejście związane z „trzema V”, oznaczającymi *volume* (objętość), *velocity* (zmienność) i *variety* (różnorodność), jednak nowsze podejścia uwzględniają inne czynniki, takie jak *value* (wartość) czy choćby *veracity* (wiarygodność), dla pokreślenia np. wzrostu socjoekonomicznej wartości uzyskanej poprzez wykorzystanie *Big Data* jako czynnika produkcji (zob. ITU, 2013, s. 8).

Z kolei Przanowski (2014, s. 13) definiuje *Big Data* w szerszym ujęciu, jako układ składający się z danych opisanych własnościami wspomnianych 3V (lub też 5V), metod składowania i przetwarzania danych, technik zaawansowanej analizy danych oraz wreszcie całego środowiska sprzętu informatycznego. Obszary te stanowią połączenie nowoczesnej technologii i teorii analitycznych, które pomagają optymalizować masowe procesy związane z dużą liczbą klientów.

Klasyczne cechy (zob. Ohlhorst, 2012, s. 3; Płoszajski, 2013, s. 6) wywodzą się z 3V, opisujące zbiory typu *Big Data*, określa się jako:

- ilość (ang. *volume*) związaną z przetwarzaniem dużych rozmiarów danych, zaczynających się od rozmiarów mierzonych w terabajtach [TB], a nawet w petabajtach [PB];

- zmienność (ang. *velocity*) opisującą dynamikę (szybkość aktualizacji) przetwarzanych danych w czasie zbliżonym do rzeczywistego (np. danych sen-

⁴ Słownik Gartner, <http://www.gartner.com/it-glossary/?s=big+data> (dostęp: 22.12.2015).

sorycznych, strumieniowych, pochodzących z Internetu) w celu ich wykorzystania;

— różnorodność (ang. *variety*) odnoszącą się do wielorakości danych nieustrukturyzowanych, np. pliki dźwiękowe i wideo, dokumenty, dane referencyjne, takie jak geolokalizacja oraz połączenia telefoniczne, logowania sieciowe, posty w sieciach społecznościowych, dla których tradycyjne bazy są niewłaściwie dostosowane.

W ramach pozostałych „V”, dodawanych do klasycznego podejścia (zob. ISO, 2015, s. 9–11; Wieczorkowski, 2014, s. 568), wymienia się:

— wiarygodność danych (ang. *veracity*) związaną z wymienionymi cechami opisującymi dane, których jakość, wynikająca z ich przetworzenia, może jednak prowadzić do pewnych błędów i niewłaściwej interpretacji;

— wartość dodaną (ang. *value*) osiąganą przez instytucje w wyniku umiejętnego przetwarzania wielkich wolumenów i wykorzystywania ich w podejmowaniu decyzji gospodarczych.

Zbiory typu *Big Data* przeważnie pochodzą z wielu nieskorelowanych źródeł często o różnym formacie. Zasadniczo dane można podzielić na wewnętrzne i zewnętrzne. Dane pochodzące z wewnętrznych źródeł zazwyczaj przedstawiane są jako dane strukturalne, które wywodzą się z relacyjnych baz danych (SQL), np. systemu finansowo-księgowego, funkcjonujących rozwiązań CRM (zarządzania relacjami z klientami), systemów transakcyjnych wykorzystywanych w danej organizacji (w tym również platform B2B) itp. Dane strukturalne są stosunkowo łatwe do eksploracji, ponieważ mają one ściśle określoną strukturę i są zgromadzone w postaci relacyjnych baz, na które składają się pliki rekordów, a ich analiza i interpretacja otrzymanych wyników wydaje się szybka i bardziej efektywna niż innych rodzajów danych. Szacuje się, że udział danych strukturalnych w całości danych dostępnych z wszelakich źródeł wynosi 5–10% (Marr, 2015, s. 59–60).

Dane pochodzenia zewnętrznego możemy podzielić na dwie podkategorie: dane semistrukuralne oraz niestrukuralne. Dane semistrukuralne (częściowo strukturalne) są pomostem między danymi strukturalnymi a niestrukuralnymi. Dane te obowiązuje powszechny standard składowania i wymiany pomiędzy heterogenicznymi systemami dużej ilości różnorodnych danych (np. język XML), których struktura nie jest sztywna (ukryta jest wewnątrz danych). Podobnie jak w wypadku danych strukturalnych, nierelacyjne bazy, do jakich należą dane semistrukuralne, stanowią 5–10% wszystkich dostępnych danych (Marr, 2015, s. 61). Z kolei dane niestrukuralne (słabo ustrukturalizowane) należą do największej kategorii danych (ok. 80%). Bazy danych w ramach tej kategorii są całkowicie nierelacyjne (NoSQL). Dane niestrukuralne nie mają określonego schematu informacji. Mogą to być różnego rodzaju dokumenty w postaci tekstowej (e-maile, pliki tekstowe, np. doc; skany dokumentów, np. pdf), multimedialnej (grafika,

dźwięki, filmy), dane sensoryczne i geolokacyjne czy choćby informacje pochodzące ze źródeł internetowych (portale społecznościowe, sklepy internetowe, fora dyskusyjne) oraz hurtowni danych. Zazwyczaj opisane są językiem HTML (ITU, 2013, s. 9).

Do niedawna tradycyjne przewagi konkurencyjne były osiągane dzięki właściwej organizacji procesów biznesowych oraz m.in. dzięki wdrożeniu systemów ERP (planowania zasobów przedsiębiorstwa). Przełomem w konkutowaniu analityką okazała się analiza trendów rynkowych, eksploracja danych dzięki rozwojowi systemów *Business Intelligence* oraz badaniu zachowań klientów w kontekście możliwości ich migracji czy choćby optymalizacji działań logistycznych oraz marketingowych. Obecnie nowe trendy wymuszają na organizacjach uwzględnienie czynnika niestukturalnego — m.in. aktywności klientów w sieci (np. na portalach społecznościowych, forach internetowych), który razem z danymi strukturalnymi stanowią źródło wiedzy potrzebnej do podejmowania decyzji biznesowych we współczesnym środowisku technologii cyfrowych.

Na zmieniających się szybko rynkach dynamicznie rośnie wolumen informacji kreowanych przez różnorodne źródła cyfrowe. Ma to istotne konsekwencje dla funkcjonowania organizacji i podmiotów gospodarczych. Analizę danych o wielkich wolumenach stosuje się do rozwiązywania problemów m.in. ekonomicznych i gospodarczych. Wykorzystanie zbiorów typu *Big Data* pozwala organizacjom sprawdzać rzeczywiste preferencje i zachowania klientów, konfrontując je z informacjami ujawnianymi przy podpisywaniu umów handlowych czy też z informacjami zawartymi w wypełnianych przy różnej okazji ankietach, a w konsekwencji umożliwia opracowanie w pewnym stopniu spersonalizowanych ofert oraz lepsze dotarcie do klienta. Umiejętne wykorzystanie narzędzi informatycznych umożliwia zdobycie przewagi konkurencyjnej, zbudowanie nowych modeli biznesowych oraz zwiększenie sprzedaży w dłuższym okresie.

Jak podaje Płoszajski (2013, s. 6–7), przetwarzanie wielkich danych tworzy wartość dla ich posiadaczy przez:

- czynienie informacji przejrzystymi i dostępnymi z większą częstotliwością;
- tworzenie i składowanie większej liczby informacji o transakcjach w formie cyfrowej dla lepszego badania efektywności działań, tworzenie bardziej precyzyjnych nisz klienckich i lepiej dopasowanych do nich produktów i usług;
- wspomaganie rozwoju następnych generacji produktów i usług;
- prowadzenie kontrolowanych eksperymentów.

Podsumowując, można pokusić się o stwierdzenie, że możliwości wykorzystywania *Big Data* są praktycznie nieograniczone. Umiejętność gromadzenia, przetwarzania i analizowania ogromnych zbiorów danych oraz wyciągania

na ich podstawie odpowiednich wniosków stanowi przyszłość działalności gospodarczej. Zaangażowanie organizacji w technologię *Big Data* widoczne jest przede wszystkim w obszarach IT i marketingu. W związku z tym najczęściej implementacja właściwych metod i narzędzi informatycznych wykorzystywana jest w biznesie (przedsiębiorstwach produkcyjnych, bankach, zakładach ubezpieczeń, firmach telekomunikacyjnych, energetycznych, transportowych itd.), ale też w sektorze publicznym (np. opiece zdrowotnej, administracji, służbach bezpieczeństwa) oraz w centrach badawczych wyspecjalizowanych w zaawansowanej analityce danych.

3. Możliwości wykorzystania analizy danych o wielkich wolumenach w obszarze usług finansowych

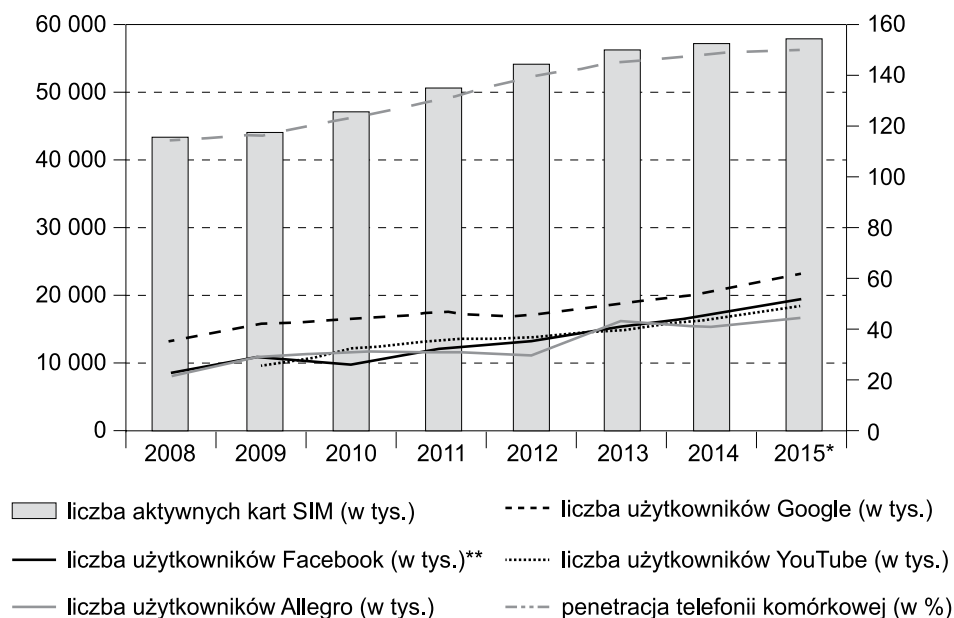
Jak wspomniano w poprzedniej części, zbiory typu *Big Data* charakteryzują się m.in. różnorodnością źródeł. Wsparciem procesu rozwiązywania problemów zarządzania mogą stać się nie tylko dane pochodzące z publicznych baz, ale także transakcje kupna/sprzedaży, informacje o postach zamieszczanych przez klientów w sieciach społecznościowych, zamieszczane zdjęcia oraz pliki audio i wideo czy choćby dane o sygnałach GPS z telefonów komórkowych.

Z uwagi na przedstawione fakty uzasadnione jest przybliżenie możliwości wykorzystania m.in. danych niestrukturalnych z perspektywy lepszego dostosowywania oferty produktowej przez instytucje finansowe w Polsce. Razem z rozwojem nowoczesnych technologii oraz samego Internetu poszerzyły się perspektywy dotarcia do klienta wykorzystującego cyberprzestrzeń do komunikowania się, spędzania wolnego czasu czy choćby robienia zakupów. Na wykresie 2 zestawiono liczbę użytkowników popularnych serwisów internetowych ze wskaźnikami telekomunikacyjnymi.

Jak wynika z wykresu 2, na przestrzeni ostatnich ośmiu lat liczba użytkowników (ang. *real users*) najpopularniejszych portali społecznościowych (np. Facebook), platform transakcyjnych (np. Allegro), serwisów oferujących materiały multimedialne (np. YouTube), czy choćby wyszukiwarki Google podwoiła się. Wynika to przede wszystkim z większego dostępu do Internetu.

* Dane za III kwartały 2015 r.

** Dla 2008 roku wartości użytkowników portali społecznościowych oparte były na danych serwisu Nasza-klasa, liczba użytkowników Facebooka zestawiona jest od roku 2009.



Wykres 2. Liczba użytkowników popularnych serwisów internetowych w Polsce (w tys. osób — skala lewa) w zestawieniu ze wskaźnikami telekomunikacyjnymi (wskaźnik penetracji rynku — skala prawa)

Źródło: opracowanie własne na podstawie danych Megapanel PBI/Gemius oraz danych GUS dotyczących telekomunikacji w Polsce.

Jak podaje Bank Światowy w swoim raporcie (World Bank, 2016, s. 177) dotyczącym technologii informacyjnych i telekomunikacyjnych, odsetek polskich gospodarstw domowych posiadających komputer zwiększył się z 40% w 2005 roku do prawie 78% w 2014 roku, natomiast odsetek gospodarstw domowych z dostępem do Internetu — z 30% w 2005 roku do blisko 75% w 2014 roku. Ważnym miernikiem społeczeństwa informacyjnego może być też odsetek osób używających Internetu: w 2005 roku było to prawie 39%, a w 2014 roku już blisko 67%. Popularność serwisów internetowych wpisuje się w ogólny trend przedstawionej informatyzacji społeczeństwa. Przykładowo, na koniec III kwartału 2015 roku:

- Facebook, główny serwis społecznościowy, umożliwiający tworzenie sieci i grup, dzielenie się wiadomościami i zdjęciami oraz korzystanie z różnego rodzaju aplikacji, miał w Polsce ponad 19,5 mln zarejestrowanych użytkowników;
- Allegro — platforma handlowa — umożliwiła przeprowadzanie transakcji blisko 16,8 mln użytkowników;

— Google, najbardziej popularna na całym świecie wyszukiwarka oraz dostarczyciel takich produktów, jak poczta Gmail oraz serwisy, m.in. GoogleMaps, zarejestrował w Polsce 23,3 mln użytkowników;

— YouTube, serwis internetowy umożliwiający bezpłatne umieszczanie i odtwarzanie filmów (w tym własnych), miał w Polsce 18,6 mln użytkowników.

Biorąc te tendencje pod uwagę, w odpowiedzi na negatywne lub też pozytywne sygnały dotyczące postrzegania marki oraz samych produktów, poprzez koncentrowanie się na aktywności osób opiniotwórczych w mediach społecznościowych, instytucje finansowe są w stanie w szybki sposób reagować na zaistniałe sytuacje. Działania takie polegają na podejmowaniu odpowiednich kroków medialnych związanych z wprowadzeniem usprawnień w produktach lub usługach, wydawaniu mniej lub bardziej formalnych komunikatów prostujących, ewentualnie rozpowszechnianiu korzystnych informacji na temat oferowanych produktów i usług.

Innym obszarem zainteresowania instytucji finansowych powinna być telekomunikacja. Liczba klientów sieci komórkowych w Polsce w 2015 roku wynosiła 55,4 mln i w porównaniu z 2008 rokiem wzrosła o 1/3. Wskaźnik penetracji rynku⁵ telefonii komórkowej w Polsce zwiększył się on w analizowanym okresie z 115% do 152% (patrz skala prawa na wykresie 2). Relatywnie wysoki poziom tego wskaźnika określa rynek telekomunikacyjny jako dosyć nasycony. Niemniej jednak w kontekście przekazywanych informacji szczególne znaczenie może mieć łączna długość rozmów krajowych przeprowadzona przez telefony komórkowe. Przeciętna długość rozmów mierzona w minutach w ramach abonamentu w ujęciu miesięcznym zwiększyła się z 46,7 w 2005 roku do 110,7 w 2014 roku (World Bank, 2016, s. 177). Wraz z rozmowami telefonicznymi zasadniczo wydłuża się czas korzystania ze smartfonów przez pisanie wiadomości tekstowych czy choćby korzystanie z Internetu za pomocą właśnie urządzeń mobilnych. Zwiększenie liczby urządzeń telefonicznych oraz częstotliwości komunikowania się klientów może skłaniać podmioty gospodarcze do zaangażowania użytkowników Internetu oraz abonentów usług telefonicznych w zadania związane z projektowaniem produktów, wykonywane do tej pory przez pracowników organizacji (ang. *crowdsourcing*). Partycypacja ta polega na pytaniu klientów, co myślą na temat danej oferty i zbieraniu pomysłów mających na celu podniesienie jej atrakcyjności.

W związku z tym, że *Big Data* to proces gromadzenia, analizowania i wyciągania wniosków z dużej ilości różnorodnych i zmiennych danych, jego zastosowanie może odgrywać znaczącą rolę m.in. w działalności instytucji finansowych. Banki czy choćby zakłady ubezpieczeń są podmiotami, które mogą

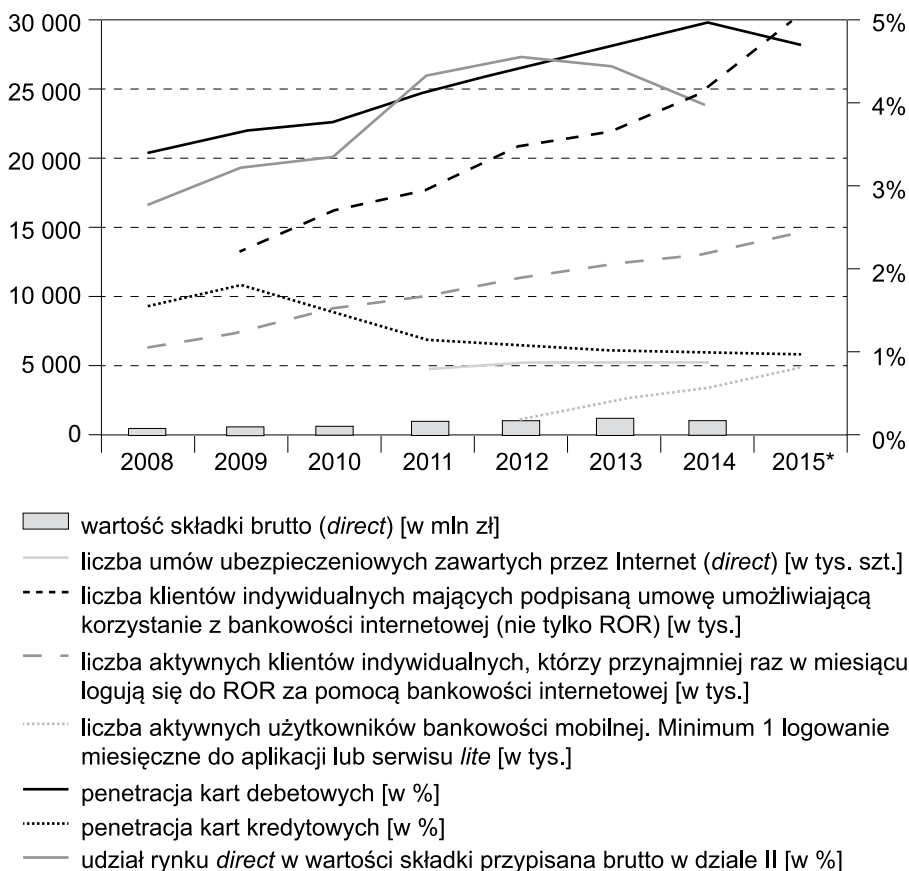
⁵ Penetracja rynku (ang. *penetration index*) mierzona jest poprzez stosunek liczby klientów, którzy kupili lub korzystali z produktów rynkowych, do całej populacji klientów.

i powinny stosować koncepcję *Big Data* w celu dopasowania produktów i usług do realnych potrzeb klienta, a tym samym wzrostu sprzedaży usług. Zastosowanie odpowiednich narzędzi do zbierania danych z wszelakich źródeł, od tych pochodzących z wewnętrznych baz klientów po zewnętrzne informacje zamieszczone w sieci przez samych klientów w sposób świadomy lub mniej świadomy (np. na portalach społecznościowych, forach internetowych), sprzyjać będzie opracowywaniu spersonalizowanych ofert oraz lepszemu dotarciu do klienta. Ich przetwarzanie na drodze *data mining*⁶ umożliwia budowę modeli statystycznych, których efektem jest podział klientów na segmenty o różnych charakterystykach, np. według poziomu ryzyka. Do wspomnianych segmentów opracowuje się strategie sprzedażowe, przygotowuje promocje wynikające z bieżących danych rzeczywistych czy choćby podejmuje się decyzje kredytowe związane z analizą zdolności kredytowej kredytobiorców. Bezpośrednimi korzyściami stosowania technologii *Big Data* są efekty biznesowe, natomiast pośrednimi — zdobycie nowej wiedzy o klientach. Predykcja, wynikająca z przetwarzania dużych zbiorów danych, będzie dotyczyć zachowań klientów, przyszłej wielkości sprzedaży lub prawdopodobieństwa utraty klienta.

Ze względu na ograniczenia objętościowe oraz specyfikę rynków finansowych w Polsce instytucje finansowe, w stosunku do których przedstawione zostaną możliwości wykorzystania dużych zbiorów danych, zawężone będą do wiodących uczestników systemu finansowego. W Polsce najbardziej rozpowszechnioną formą usług proponowanych przez instytucje finansowe, polegającą na umożliwieniu dostępu klientom do oferowanego produktu za pomocą urządzenia elektronicznego, są produkty bankowe. Wdrożenie przez banki scentralizowanego systemu informatycznego umożliwiło wprowadzenie jednolitych standardów obsługi oraz pozwoliło na korzystanie z dostępu do rachunku bankowego z dowolnego miejsca oraz bez ograniczeń czasowych (Niczyporuk, Talecka, 2011, s. 213). Pozostali pośrednicy finansowi również korzystają na szeroką skalę z nietradycyjnych kanałów dystrybucji (zob. Filip, 2006; Wiczorek-Kosmala, Staszal, 2014). Na wykresie 3 przedstawiono liczbę użytkowników bankowości elektronicznej z podziałem na liczbę klientów i liczbę wydanych kart płatniczych oraz wielkość i wartość rynku ubezpieczeń w systemie *direct*.

* Dane charakteryzujące bankowość za III kwartały 2015 r.

⁶ Eksploracja danych (ang. *data mining*) oznacza proces analityczny przeznaczony do badania dużych zasobów danych w poszukiwaniu regularnych wzorców oraz systematycznych współzależności pomiędzy zmiennymi. Proces ten składa się z trzech głównych etapów: wstępnego zagłębienia, budowania modelu i weryfikacji jakości danych oraz wdrożenia i stosowania modeli dla nowych danych w celu uzyskania przewidywanych wartości.



Wykres 3. Liczba użytkowników bankowości elektronicznej w zestawieniu z wielkością oraz wartością rynku ubezpieczeń w systemie *direct*

Źródło: opracowanie własne na podstawie danych Związku Banków Polskich (raporty Net@Bank) o liczbie klientów bankowości elektronicznej, danych Narodowego Banku Polskiego o liczbie wydanych kart płatniczych oraz danych Polskiej Izbie Ubezpieczeń o rynku *direct*.

Obecnie najpopularniejszym kanałem bankowości elektronicznej w Polsce jest bankowość internetowa. Jak wynika z wykresu 3, liczba klientów indywidualnych mających podpisaną umowę umożliwiającą korzystanie z bankowości internetowej (nie tylko ROR), a w tym również tych aktywnych klientów, od 2008 roku wzrosła ponad dwukrotnie. Na koniec III kwartału liczba bankowych klientów z internetowym dostępem do rachunku wynosiła ponad 30,3 mln. Z kolei liczba aktywnych klientów indywidualnych, którzy przynajmniej raz w miesiącu logują się do ROR za pomocą bankowości internetowej, była o połowę mniejsza i w III kwartale 2015 roku wynosiła 14,6 mln. Zapewne część z nich to klienci, którzy korzystają również z bankowości mobilnej. Dynamiczny rozkwit banko-

wości wykorzystującej tzw. systemy *lite* można obserwować od 2012 roku, kiedy to z odpowiednich aplikacji na telefony komórkowe korzystało niecałe 1,2 mln osób. Na koniec III kwartału 2015 roku liczba aktywnych (minimum z jednym logowaniem miesięcznie do aplikacji lub serwisu *lite*) użytkowników bankowości mobilnej wynosiła 4,9 mln osób. Taki stan rzeczy wynika przede wszystkim z popularyzacji mobilnych systemów płatności, lecz także z dalszego wzrostu rynku smartfonów i spadających cen Internetu. Jak podają Filip i Komorowski (2016), na całym świecie obserwuje się trend ograniczania tradycyjnej sieci sprzedaży, który jest wywołany m.in. rosnącym znaczeniem elektronicznych kanałów dystrybucji usług. Wiedza o preferencjach ofertowych klientów umożliwia stosowanie segmentacji klientów przez instytucje finansowe. Segmentowanie przy użyciu technologii *Big Data* jest doskonałym narzędziem w dostosowywaniu produktów do potrzeb klienta, w dostosowywaniu obsługi oraz systemu transakcyjnego do wymagań klientów, w poprawie relacji z kluczowymi klientami, w lepszym przygotowywaniu oferty dla nowych klientów na podstawie doświadczeń całego segmentu czy choćby opracowywaniu całkowicie nowych produktów.

Nieco inne tendencje można zaobserwować na polskim rynku kart płatniczych. O ile liczba kart debetowych w Polsce powoli rośnie (na koniec III kwartału 2015 roku było ich ponad 28 mln) — rynek ten w okresie 2008–2015 zwiększył się o 40%, co daje średnie roczne tempo na poziomie 5%, o tyle rynek kart kredytowych zmniejsza się. Szczyt popularności tych drugich można było odnotować w 2009 roku (10,8 mln kart). W III kwartale 2015 roku wydanych było 5,8 mln kart kredytowych. Jednak wzrost znaczenia innych produktów, w tym również kart debetowych, świadczy o pewnej ich uniwersalności (np. w związku z możliwością realizowania nimi zakupów w Internecie, zwiększającą się liczbą bankomatów oraz terminali POS). Niemniej jednak banki lub też instytucje pośredniczące w płatnościach, poprzez urządzenia uwierzytelniające płatność kartami (terminale POS), instalowanymi w punktach handlowo-usługowych, monitorują zwyczaje zakupowe swoich klientów w czasie rzeczywistym. Oprócz historii transakcji zachowywana jest również informacja o geolokalizacji zakupów oraz częstotliwości i skali transakcji, a to z kolei przekłada się na możliwość wyciągania wniosków na temat miejsca przebywania, stylu życia, upodobań zakupowych czy choćby płynności finansowej klientów.

Drugim analizowanym segmentem rynku finansowego jest rynek ubezpieczeń w systemie *direct*, czyli umożliwiającym zakup lub obsługę ubezpieczeń przez Internet lub telefon. Jak podają Wieczorek-Kosmala i Staszek (2014, s. 244), oprócz regulacji prawnych, dotyczących możliwości sprzedaży ubezpieczeń w kanałach nietradycyjnych, istotne znaczenie w sprzedaży ubezpieczeń poprzez wymienione kanały dystrybucji ma rozwój technologii. W działalności ubezpieczeniowej znaczenie systemu *direct* systematycznie rośnie, co potwierdza wzrost udziału sprzedaży w kanałach Internet, telefon i *call centers* w wartości składki przypi-

sanej brutto w dziale II⁷ (zob. skala prawa na wykresie 3) oraz samej wartości składki według grup ryzyka w dziale II. Udział rynku *direct* na koniec 2014 roku wyniósł 4%, natomiast składka przypisana brutto (dział II) — 1,1 mld zł. Informacja o klientach usług ubezpieczeniowych, stojąca za liczbą umów zawartych przez Internet w konkurencyjnych warunkach rynkowych, ma istotne znaczenie dla powodzenia prowadzonej przez te podmioty działalności. Wiele instytucji finansowych wykorzystuje informacje płynące z *call center* i w powiązaniu z danymi niestrukturalnymi dąży do redukcji migracji klientów, poprawy wyników sprzedażowych, wprowadzenia sprzedaży produktów dodatkowych (ang. *cross-selling*) czy monitorowania niepokojących zgłoszeń szkód oraz wykrywania oszustw.

W wyniku analizy danych o wielkich wolumenach informacji o zachowaniach i preferencjach klientów instytucji finansowych możliwe jest precyzyjne oferowanie klientom produktów i usług finansowych. Wśród najpopularniejszych typów wykorzystania technologii *Big Data* w bankowości i finansach (SAP, 2014) wymienia się:

- wykrywanie oszustw finansowych oraz dążenie do zapobiegania ich występowaniu poprzez zwiększenie monitorowania działalności finansowej klientów, identyfikując powierzchownie niewidoczne niebezpieczne zjawiska oraz różniąc normalną działalność od tej nielegalnej;

- segmentację klientów polegającą na określaniu grup klientów w celu lepszego do nich dotarcia, a tym samym wsparcia sprzedaży, promocji oraz kampanii reklamowych, jak również rozwój produktów i usług, np. poprzez analizę zachowania segmentu odbiorców przy użyciu m.in. plików *cookies*, czyli tzw. ciasteczek;

- personalizowanie oferowanych produktów m.in. poprzez wykorzystanie narzędzi targetowania behawioralnego, polegającego na tym, że oferuje się właściwe produkty odpowiednim klientom. Jest to związane z wykorzystaniem technologii pomagającej w zrozumieniu zachowań zakupowych klientów poprzez kanały, które do nich docierają, oraz w utrzymaniu klienta i wypracowaniu lojalności (np. poprzez analizę częstotliwości i kierunku wykonywanych przelewów, obserwowanie aktywności w porównywaniu ofert);

- zarządzanie ryzykiem związane z sygnalizowaniem centralnym organom nadzorczym wprowadzania nowych zarządzeń oraz rekomendacji, jak również elastyczne dostosowywanie wewnętrznych regulacji do potrzeb rynkowych z zachowaniem wymaganej ostrożności operacyjnej oraz wczesnego ostrzegania w sytuacjach ryzykownych (np. poprzez wykrywanie zachowań klienta mogących mieć wpływ na ograniczenie jego zdolności do regulowania swoich zobowiązań w przyszłości).

⁷ Podział typów ubezpieczeń na dział I (ubezpieczenia na życie) oraz dział II (pozostałe ubezpieczenia osobowe oraz ubezpieczenia majątkowe) wynika z przekroju rodzajowego prezentowania podstawowej działalności zakładów ubezpieczeń.

Podsumowując, należy stwierdzić, że *Big Data* to zbiory danych o zachowaniach klientów, ich aktywności, preferencjach, nawykach — pochodzące zarówno z wewnętrznych źródeł instytucji finansowych (np. systemy centralne, CRM, systemy transakcyjne, bankowość internetowa, geolokalizacja urządzenia mobilnego), jak i tych zewnętrznych (np. KIK, KRD, lecz także portale społecznościowe, sklepy internetowe, fora dyskusyjne). Technologię wielkich wolumenów danych wykorzystuje się w celu opracowywania spersonalizowanych ofert oraz lepszego dotarcia do klienta, a w konsekwencji zdobycia przewagi konkurencyjnej. W warunkach polskich staje się to możliwe dzięki rozwojowi nowoczesnych technologii oraz samego Internetu, który poszerzył perspektywę dotarcia do klienta wykorzystującego cyberprzestrzeń do komunikowania się, spędzania wolnego czasu oraz robienia zakupów.

4. Wnioski

Big Data jest obszarem nauki, który niezwykle szybko się rozwija i zdobywa coraz większą popularność. Omawiana tematyka w polskim piśmiennictwie dopiero zaczyna być poruszana. W związku z tym celem pracy było dokonanie przeglądu możliwości i zakresu metod związanych z analizą danych o wielkich wolumenach. Niniejszy tekst jest jedynie wprowadzeniem do tejże tematyki, dokonaniem w celu określenia perspektyw, jakie niesie z sobą wdrożenie *Big Data* w instytucjach finansowych. Ponadto usystematyzowanie pewnych zagadnień stanowi podstawę do dalszych rozważań i analiz w przedstawionym zakresie.

Do niedawna problemem w sprawnym zarządzaniu przedsiębiorstwem było nadążanie za bieżącymi trendami. Początkowo dane, często pochodzące z systemów centralnych, rozwiązań CRM oraz ankiet wypełnianych przez klientów, były używane do zwiększenia w krótkim okresie przychodów ze sprzedaży. Obecnie analiza danych o wielkich wolumenach umożliwia przede wszystkim nadążanie za zmieniającymi się warunkami rynkowymi i zachowaniami klientów oraz, w niektórych wypadkach, kreowanie nowych trendów. Dzięki technologii *Big Data* instytucje finansowe dążą do redukcji migracji klientów, poprawy wyników sprzedażowych oraz tworzą zindywidualizowane oferty uwzględniające rzeczywiste potrzeby klienta. Ponadto instytucje finansowe, które są zaangażowane społecznie, przy wykorzystaniu nowoczesnych narzędzi, takich jak *Big Data*, są w stanie stworzyć całkowicie nowe podejścia biznesowe.

Literatura

- Carr N.C. (2004), *Does IT Matter?: Information Technology and the Corrosion of Competitive Advantage*, Harvard Business School Publishing, Boston.
- Dev D., Patgiri R. (2015), *A Survey of Different Technologies and Recent Challenges of Big Data*, [w:] *Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics*, red. A. Nagar, D.P. Mohapatra, N. Chaki, ICACNI, t. 2, Springer, New Delhi, s. 537–548.
- Drucker P.F. (2009), *Zarządzanie XXI wieku. Wyzwania*, MT Biznes, Warszawa.
- Filip D. (2006), *Zmiany w zarządzaniu kanałami dystrybucji na rynku funduszy inwestycyjnych*, [w:] *Organizacje, przedsiębiorstwa i rynki kapitałowe w otoczeniu globalnym*, red. M. Lisiecki, Wydawnictwo KUL, Lublin 2006, s. 408–417.
- Filip D., Komorowski P. (2016), *Konkurencyjność rynków finansowych. Czynniki efektywności*, „Przedsiębiorstwo i Finanse”, WSFiZ w Białymstoku.
- ISO (2015), *Big Data: Preliminary Report 2014*, ISO/IEC JTC 1, Information technology.
- ITU (2013), *Big Data: Big Today, Normal Tomorrow*, ITU-T Technology Watch Report, International Telecommunication Union, November 2013.
- Mankiw G.N., Romer D., Weil D.N. (1992), *A Contribution to the Empirics of Economic Growth*, „The Quarterly Journal of Economics” 107, nr 2, s. 407–437.
- Marr B. (2015), *Big Data: Using SMART Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance*, Wiley.
- Mayer-Schönberger V., Cukier K. (2013), *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, An Eamon Dolan Book, Houghtin Mifflin Harcourt, Boston, New York.
- MCI (2011), *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, May 2011.
- Niczyporuk P., Talecka A. (2011), *Bankowość. Podstawowe zagadnienia*, Wyd. Temida 2, Białystok.
- Nonneman W., Vanhoudt P. (1996), *A Further Augmentation of the Solow Model and the Empirics of Economic Growth for OECD Countries*, „The Quarterly Journal of Economics” 111, nr 3, s. 943–953.
- Ohlhorst F.J. (2012), *Big Data Analytics: Turning Big Data into Big Money*, SAS Institute Inc., Cary, North Carolina.
- Płoszajski P. (2013), *Big Data: nowe źródło przewag i wzrostu firm*, „E-mentor” nr 3 (50), s. 5–10.
- Przanowski K. (2014), *Credit Scoring w erze Big Data. Techniki modelowania z wykorzystaniem generatora losowych danych portfela Consumer Finance*, Oficyna Wydawnicza SGH, Warszawa.
- Sagiroglu S., Sinanc D. (2013), *Big Data: A Review*, [w:] *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS)*, IEEE, San Diego, s. 42–47.
- SAP (2014), *Top 5 Big Data Use Cases in Banking and Financial Services*, https://www.sap.com/bin/sapcom/fi_fi/downloadasset.2014-03-mar-05-23.top-5-big-data-use-cases-in-banking-and-financial-services-pdf.html (dostęp: 13.12.2015).
- Solow R. (1956), *A Contribution to the Theory of Economic Growth*, „The Quarterly Journal of Economics” 70, nr 1, s. 65–94.
- Thurow L.C. (2006), *Nowe reguły gry w gospodarce opartej na wiedzy*, Wydawnictwo Helion, Gliwice.
- Wieczorek-Kosmala M., Staszal A. (2014), *Znaczenie nietradycyjnych kanałów dystrybucji ubezpieczeń w Polsce w latach 2007–2012*, „Studia Ekonomiczne” nr 198/14, s. 236–246.
- Wieczorkowski J. (2014), *Wykorzystanie koncepcji big data w administracji publicznej*, „Roczniki Kolegium Analiz Ekonomicznych” 33, s. 567–579.
- World Bank (2016), *The Little Data Book on Information and Communication Technology 2015*, Washington, DC: World Bank, DOI:10.1596/978-1-4648-0558-5.