Dr Aneta Pluta-Zaremba

Szkoła Główna Handlowa w Warszawie

# **ANALITYKA *BIG DATA* W *E-COMMERCE* [[1]](#footnote-1)**

Streszczenie: Artykuł omawia istotne zagadnienia związane z wykorzystaniem nowoczesnych rozwiązań cyfrowych do zarządzania przedsiębiorstwami. Celem artykułu jest wyjaśnienie pojęcia dużych zbiorów danych wraz z możliwościami ich analizy oraz pokazanie zastosowania analityki *big data* w handlu detalicznym ze szczególnym uwzględnieniem potencjału dla przedsiębiorstw e-commerce. W artykule przedstawiono wyniki przeglądu literatury wsparte analizą obszarów zastosowania analityki biznesowej do badania popytu i zachowań klientów w celu dostosowywania rozwiązań do potrzeb konsumentów. Podkreślono także znaczenie analityki big data dla możliwości osiągnięcia przewagi konkurencyjnej.

## **WSTĘP**

We współczesnych czasach przedsiębiorstwa doświadczają ogromnego napływu niezliczonych ilości danych cyfrowych, pochodzących z różnych źródeł. Jednym z największym wyzwań stawianych przed przedsiębiorstwami i analitykami nie jest jednak sam proces zbierania danych, lecz jak najlepsze wykorzystanie ogromnej ilości danych tworzonych w czasie rzeczywistym. Coraz więcej przedsiębiorstw korzysta z analityki *big data*, aby wykrywać prawidłowości, które następnie wykorzystywane są w prognozowaniu przyszłych zachowań i potrzeb klientów czy zdarzeń. Badacze przedmiotu określają termin *big data* mianem „rewolucji w zarządzaniu” wskazując na kluczową rolę w tworzeniu przewagi konkurencyjnej.

Artykuł ma na celu wyjaśnienie pojęcia dużych zbiorów danych wraz z możliwościami ich analizy oraz pokazanie zastosowania analityki *big data* w handlu detalicznym ze szczególnym uwzględnieniem potencjału dla przedsiębiorstw *e-commerce*.

## **Definicja i cechy *big data***

Pojęcie *big data* doczekało się wielu definicji, które podkreślają różne cechy i atrybuty wielkich zbiorów danych. Wynika to po części ze złożoności pojęcia oraz z jego ciągłego ewoluowania. Firma Gartner wielkie zbiory danych określiła jako „zasoby informacyjne o dużej objętości, dużej szybkości i/lub dużej różnorodności, wymagające opłacalnych, innowacyjnych form przetwarzania informacji, które umożliwiają lepszy wgląd, podejmowanie decyzji, i automatyzację procesów”[[2]](#footnote-2). Definicja ta zwraca uwagę na podstawowe atrybuty i wymagania względem danych określanych mianem Big Data. Do trzech głównych wymiarów (tzw. 3V) charakteryzujących wielkie zbiory danych D. Laney zaliczył[[3]](#footnote-3):

* objętość (ang. *volume*) związaną z ogromnym zasobem potencjalnych danych, ciągle wytwarzanych i przesyłanych przez użytkowników lub same urządzenia (dzięki Internetowi rzeczy), liczonych w terabajtach (TB) lub w petabajtach (PB) i przyrastających w ogromnym tempie przypominającym tsunami. Im więcej działań jest digitalizowanych tym pojawia się więcej nowych źródeł informacji, w tym dane których wcześniej nie można było wykorzystać w analizach (np. dane z portali społecznościowych i blogów, nagrania audio i wideo);
* różnorodność (ang. *variety*) wynikającą z szerokich możliwości pozyskiwania i analizowania danych w różnej postaci – ustrukturyzowanej, nieustrukturyzowanej lub częściowo ustrukturyzowanej;
* i prędkość (ang. *velocity*) nawiązującą do możliwości technologicznych przesyłania powstałych danych niemal w czasie rzeczywistym, oraz ich analizowania bez konieczności zbierania i przechowywania w hurtowni danych. Pozwala to na analizowanie danych na etapie ich powstawania, dzięki czemu przedsiębiorstwa mogą szybko reagować na bieżące wydarzenia czy zakłócenia.

W następnych latach do tych trzech podstawowych, o największym znaczeniu wymiarów *big data* badacze dołączali nowe atrybuty. Jako czwarty atrybut najczęściej wymieniana jest wartość (ang. *value*), która wynika z umiejętności analizy danych w celu opracowania praktycznych informacji na potrzeby predykcji zdarzeń i zarządzania nimi[[4]](#footnote-4). Natomiast P. Taori i H.K. Dasararaju zamiast wartości zaproponowali inną cechę – prawdziwość (ang. *veracity*) związaną z niepewnością danych, wynikającą z poprawności i spójności danych. Po dołączeniu kolejnego wymiaru zaczęto określać *big data* mianem idei czterech ,,V”.

Dane mogą występować w trzech formatach jako: ustrukturyzowane, nieustrukturyzowane i częściowo ustrukturyzowane[[5]](#footnote-5). Dane ustrukturyzowane to wysoce zorganizowane informacje, które można łatwo przechowywać w arkuszu kalkulacyjnym lub tabeli za pomocą wierszy i kolumn, np. dane o sprzedaży, raporty finansowe, rekordy klientów (np. historia zakupów). Dane nieustrukturyzowane mogą mieć własną strukturę wewnętrzną, np. pliki wideo i audio (tj. filmy i dźwięki), obrazy i tekst z portali społecznościowych, opinie i recenzje użytkowników na stronach internetowych. Pomiędzy tymi dwoma kategoriami są dane częściowo ustrukturyzowane, do których należą m.in. pliki dziennika logowania i dane Internetu rzeczy (IoT) generowane z szerokiej gamy czujników i urządzeń (np. dziennik strumienia kliknięć ze strony e-commerce, który podaje szczegółowe informacje o dacie i godzinie działań, adres IP użytkownika, z którego wykonuje transakcję).

Do przetwarzania dużych zbiorów danych do przydatnych (nadających się do użytku) informacji, które przedsiębiorstwo może zaabsorbować, zrozumieć i efektywnie wykorzystać do podejmowania bardziej świadomych decyzji w czasie zbliżonym do rzeczywistego, potrzebne jest zastosowanie nowych rozwiązań technologicznych.

Wątpliwości użytkowników – indywidualnych i biznesowych - budzi bezpieczeństwo danych, których gromadzenie, analizowanie i wykorzystanie jest obarczone ryzykiem zagrożenia cyberprzestępczością. Tym bardziej, że liczba gromadzonych danych w tym big data rośnie w postępie geometrycznym, a wraz z nimi wspomniane ryzyko. Są one zbierane niemalże w każdej sferze naszego życia w trakcie transakcji internetowych, płatności elektronicznych, korzystania z mediów społecznościowych, serwisów informacyjnych oraz z Internetu rzeczy. Przezwyciężeni tych barier wymaga zbudowania zasobów do zarządzania ryzykiem i jego redukowania[[6]](#footnote-6).

## **Analityka biznesowa dużych zbiorów danych**

W zarządzaniu przedsiębiorstwem istotne znaczenie odgrywają umiejętności pozyskiwania oraz analizowania dużych zbiorów danych zewnętrznych i wewnętrznych. Big data, nawet jeśli napływają strumieniowo w czasie rzeczywistym pokazują zdarzenia historyczne. Do analizy w zależności od rodzaju danych wykorzystuje się różne techniki, w tym m.in. eksplorację danych, analitykę deskryptywną (ang. *descriptive analytics*), interaktywną analizę i planowanie (ang. *interactive analysis and planning*), analitykę wbudowaną (ang. *embedded analytics*) oraz analizę strumieniową (ang. *stream analytics*). Analityka deskryptywna analizuje wydarzenia z przeszłości. Big data mogą być automatycznie analizowane bez udziału specjalistów z zastosowaniem m.in. sztucznej inteligencji z maszynowym uczeniem. Zidentyfikowane zależności mogą być wykorzystywane w kolejnym etapie do przewidywania przyszłych zdarzeń (np. rozwoju rynku, zachowań konsumentów, zmian w popycie) czym zajmuje się analityka predyktywna. Z kolei analityka preskryptywna dostarcza wiedzy, jak te zachowania modelować oraz wpływać na kształtowanie pożądanych przez firmę postaw. Te trzy rodzaje analiz wzajemnie się uzupełniają. Na przykład odkrycie trendów lub mikrotrendów w zachowaniach zakupowych klientów pozwala na opracowanie dokładniejszej prognozy popytu a następnie na przygotowanie się przedsiębiorstwa na przyszłe zmiany w popycie. Analityka wielkich zbiorów danych zyskuje szybko na znaczeniu w ostatnich latach, ponieważ umożliwia osiągnięcie przewagi konkurencyjnej.

## **Zastosowanie analityki Big Data w handlu elektronicznym**

Klient i jego potrzeby stanowią centralny element zarządzania współczesnymi przedsiębiorstwami. Analityka dużych zbiorów danych tworzy nowe możliwości poznania potrzeb, zachowań i sposobów podejmowania decyzji zakupowych przez klientów, a następnie planowania działań mających na celu dostosowanie się do uświadomionych lub nieuświadomionych potrzeb lub ich wyprzedzanie. Firmy e-commerce od Amazona do Netflixa przechwytują różne typy danych (np. zamówienia, koszyki, wizyty, użytkownicy, odsyłacze, słowa kluczowe, przeglądanie katalogów, dane społeczne), które można ogólnie podzielić na cztery kategorie: a) dane dotyczące transakcji lub działalności gospodarczej b) dane strumienia kliknięć, (c) dane wideo i (d) dane głosowe. Zastosowanie znajdują zarówno ustrukturyzowane dane (które koncentrują się na danych demograficznych, w tym na nazwisku, wieku, płci, dacie urodzenia, adresie i preferencjach), jak i nieustrukturyzowane dane obejmujące kliknięcia, polubienia, linki, tweety, głosy, wypowiedzi, filmiki itp. Wyzwaniem jest poradzenie sobie z analizą obu typów danych w celu zwiększenia konwersji[[7]](#footnote-7). W e-commerce dane są kluczem do śledzenia zachowań zakupowych konsumentów i personalizowania ofert. Analityka *big data* pozwala na bieżąco „obserwować” klientów, konkurencję i dostawców oraz pomaga identyfikować przyszłe trendy i zachodzące zmiany tak, aby przedsiębiorstwa mogły się przygotować na ich nadejście. Dzięki temu odgrywa istotną rolę w budowaniu przewagi konkurencyjnej wśród przedsiębiorstw e-commerce.

Analityka Big Data stwarza wiele możliwości dla branży e-commerce, głównie w obszarach związanych z klientami i jego zachowaniami zakupowymi, personalizacją oferty i produktów, dynamicznym dostosowywaniem cen, przewidywaniem popytu, pilnowaniem bezpieczeństwa transakcji, a także analizą przyszłych trendów oraz optymalizacją wydatków i obniżeniem kosztów własnych. Przykłady zastosowania analityki big data zawiera tabela 1.

Najlepszym przykładem wyprzedzania zamówień klientów są działania firmy Amazon, która wykorzystuje dane z mediów społecznościowych do analizy zachowań klientów i predykcji popytu[[8]](#footnote-8). Na podstawie analizy dyskusji w mediach społecznościowych na temat określonych produktów i pozytywnych komentarzy internautów są podejmowane decyzje o prze­pływach produktów do magazynów i sklepów, w pobliżu których przewiduje się, że wystąpi popyt na produkty, zanim konsumenci złożą zamówienie. Wszystko po to, aby Amazon mógł realizować dla zarejestrowanych klientów szybką dostawę w ciągu 2 godzin, zgodnie z usługą Amazon Prime Now dostępną w wybranych miastach na terenie m.in. Stanów Zjednoczonych, Wielkiej Brytanii, Francji, Niemiec, Hiszpanii, Włoch, Indii i Japonii.

Tabela 1. Przykłady zastosowania analityki *big data* w *e-commerce*

|  |  |
| --- | --- |
| **Obszar** | **Charakterystyka zastosowania analityki *big data*** |
| Analiza klienta i jego zachowań | Stosowanie zaawansowanych analiz dużych zbiorów danych pozwala m.in. na: - przeprowadzanie analiz koszykowych klientów, - segmentację nabywców, - wykonanie analiz predykcyjnych dotyczących możliwości odejść klientów,- analizę porzuconych koszyków i rekomendacje dotyczące działań zachęcających do zakupu (np. dzięki zniżkom, rekomendacji produktów, doboru kanału kontaktu – email, chat) - zwiększanie satysfakcji klientów z zakupów dzięki personalizacji oferty- ocenę satysfakcji klienta i prawdopodobieństwa zakupu - zindywidualizowanie ofert wysyłanych do konkretnych klientów dzięki silnikowi rekomendacyjnemu - sprawdzenie, co sprawia, że klient zostaje na stronie e‑sklepu, np. jaki język (poważny, zabawny, profesjonalny) i które elementy graficzne (kształty, kolory, zdjęcia, grafiki) wpływają na dłuższe przeglądanie oferty, co zwiększa prawdopodobieństwo sprzedaży |
| Personalizacja oferty produktowej i obsługi  | Wykorzystując analitykę Big Data e-detaliści mogą analizować zachowania konsumentów korzystając z zasobów danych (np. z przeglądarki, profilu na Facebooku, zapisanych ciasteczek, kodów remarketingowych, wcześniej przeglądanych stron internetowych) w celu dostosowania oferty do poszczególnych klientów. Im więcej danych dotyczących konkretnego klienta jest zebranych tym oferta będzie bardziej spersonalizowana, e-sklep może zaoferować programy lojalnościowe czy pozyskać wiedzę o pozakupowym zachowaniu klientów. Dzięki analizom Big Data detalista może dowiedzieć się, jakich produktów poszukują klienci z grupy docelowej. W przypadku analizy w czasie rzeczywistym, może zapewnić personalizację ofert oraz bardziej skuteczny *cross- i up‑selling*. |
| Rekomendowanie produktów  | Personalizacja oferty zwiększa wskaźnik zaangażowania klientów i sprzyja większym zakupom. Na podstawie historii zakupów klienta oraz opinii w mediach społecznościowych system oparty o big data i sztuczną inteligencję może zarekomendować produkty, które z dużym prawdopodobieństwem doda do koszyka lub takie przynoszące detaliście największe zyski. |
| Dostosowywanie oferty dzięki dynamicznym cenom | Automatyczna analityka *big data* umożliwia prowadzenie aktywnej polityki cenowej w czasie rzeczywistym – zarówno nakierowanej na testowanie najwyższej możliwej ceny, jak i udostępnienie specjalnych ofert i kodów rabatowych dla wybranych klientów rezygnujących z zakupu. Dynamiczne ceny wprowadzane są dzięki analizie big data wraz ze sztuczną inteligencją i uczeniem maszynowym. Zastosowanie tej metody umożliwia oferowanie klientom różnych cen za ten sam produkt. Wykorzystywane jest przez e-detalistów w celu rywalizacji pomiędzy nimi oraz zapobieganiu odejściu konsumentów do konkurencji, która ma niższe ceny, lub dostosowywanie popytu do podaży. Wymaga korzystania z wielu źródeł: analiz bieżących cen u konkurencji, historycznych danych, zachowań i preferencji klientów, wysokości marży, zapasów oraz wielu innych czynników. Dynamiczne ceny stosuje m.in. Amazon, który aktualizuje ceny 2,5 miliona razy dziennie w wyniku czego cena zmienia się średnio co 10 minut.  |
| Analiza przyszłych trendów | Analiza danych nieustrukturyzowanych pochodzących z social-media, blogów, profili i opinii użytkowników, hashtagów, popularnych słów kluczowych itp. pozwala wykryć nadchodzące trendy i popyt na określone kategorie oraz zmniejszyć ryzyko decyzji o produkcji lub zakupie określonych modeli czy wdrożeniu nowych produktów. Prognozy uwzględniają wiele czynników wpływających na popyt, np. dane o sprzedaży, wiadomości w mediach społecznościowych, zapytania w wyszukiwarce, sytuację gospodarczą w kraju, a nawet warunki pogodowe. Pozwala to detalistom na wprowadzenie nowych produktów przed wzrostem popytu.Niektóre algorytmy są w stanie rozpoznać kto ze znanych osób tzw. celebrytów, jaki model ubrania założył, co może zwiększyć popyt na model określonej marki albo typ produktu. Te informacje, można wykorzystać do wyeksponowania w e-sklepie podobnych produktów, w celu zwiększenia sprzedaży. |
| Wspieranie działań logistycznych | Analityka *big data* wspomaga planowanie tras kurierów: godzin doręczenia paczki, kolejności, warunków na drodze i innych uwarunkowań (np. Samochody UPS w USA maja ograniczać skręcanie w lewo).Na podstawie analizy historycznych danych zakupowych, świąt, wydarzeń itp. można wyliczać wielkość zapasów poszczególnych produktów w magazynie, aby zapewnić wysoką dostępność produktów ograniczając koszty związane zapasami i magazynowaniem produktów. Takie analizy wykonuje np. Amazon.Istnieją systemy, które mogą na podstawie danych historycznych oszacować ryzyko zwrotu produktu na podstawie zamówień historycznych (np. częste zwroty ze względu na rozmiar) czy wykryć osoby z wysokim prawdopodobieństwem zwrotu. |
| Monitorowanie bezpieczeństwa transakcji | Na podstawie danych z różnorodnych źródeł, m.in.: informacji o podmiotach i relacjach między nimi, transakcji, wniosków kredytowych oraz czarnych / białych list opracowywane mogą być wnioski dotyczące bezpieczeństwa transakcji w celu blokowania tych uważanych za niebezpieczne. Tym samym dzięki analizie zachowań i powiązań można ograniczać przestępstwa finansowe. |

## **PODSUMOWANIE**

Analityka wielkich zbiorów danych zyskuje szybko na znaczeniu w ostatnich latach, ponieważ umożliwia osiągnięcie przewagi konkurencyjnej. Możliwość śledzenia trendów, popytu, potrzeb klientów oraz aktywności konkurencji w czasie rzeczywistym zapewnia współczesnym detalistom internetowym informacje, które pozwalają na dostosowywanie działań do warunków rynkowych i wyprzedzanie niektórych zdarzeń. Dlatego obecnie w zarządzaniu przedsiębiorstwem istotne znaczenie odgrywają umiejętności pozyskiwania oraz analizowania dużych zbiorów danych zewnętrznych i wewnętrznych. Analityka biznesowa i big data są stosowane przez wiele przedsiębiorstw e-commerce, ze względu na potrzebę szczególnego dbania o klientów i ich satysfakcję przy ograniczaniu wysokich kosztów i dużej niepewności prowadzenia działalności.

## **Bibliografia**

Gartner, INC. (2017) *Big Data. IT Glossary,* <http://www.gartner.com/it-glossary/big-data/> (dostęp 20.10.2021).

[**https://www.conversion.pl/blog/wspolczynnik-konwersji/**](https://www.conversion.pl/blog/wspolczynnik-konwersji/) (dostęp 28.10.2021).

Kaisler S.H., Armour F., Espinosa J.A., Money W., *Big Data: Issues and Challenges Moving Forward*", *46th. Hawaii International Conference on System Sciences*, Maui, Hawaii 2013.

Laney D., *The importance of 'Big Data’: A definition*, Gartner 2001, Tom 21, s. 2014-2018.

OECD, *Going Digital: Shaping Policies, Improving Lives*,OECD Publishing, Paris 2019, https://doi.org/10.1787/9789264312012-en.

Pluta-Zaremba A., *Innowacje cyfrowe w branży e-commerce*. W: K. Nowicka (red.) *Biznes cyfrowy. Perspektywa innowacji cyfrowych*, Oficyna Wydawnicza SGH, Warszawa 2019.

Taori P., Dasararaju H.K., *Big Data Management*. W: Bhimasankaram Pochiraju B., Seshadri S. (red.), *Essentials of Business Analytics An Introduction to the Methodology and its Applications,* (s. 71-109), Springer International Publishing, Cham 2019.

1. Sfinansowano ze środków projektu „Nowoczesny model współpracy szkół zawodowych ze szkołami wyższymi i pracodawcami w zakresie kształcenia w zawodach z grupy branżowej teleinformatycznej (technik telekomunikacji, technik informatyk)”, akronim: MEN-IT nr POWR.02.15.00-00-2009/18 [↑](#footnote-ref-1)
2. Gartner, INC. (2017) *Big Data. IT Glossary*, <http://www.gartner.com/it-glossary/big-data/> (dostęp 20.10.2021). [↑](#footnote-ref-2)
3. D. Laney, *The importance of 'Big Data’: A definition*, Gartner 2001, Tom 21, s. 2014-2018. [↑](#footnote-ref-3)
4. S.H. Kaisler, F. Armour, J.A. Espinosa, and W. Money, *Big Data: Issues and Challenges Moving Forward*", *46th. Hawaii International Conference on System Sciences*, Maui, Hawaii 2013, s. 5. [↑](#footnote-ref-4)
5. Taori P., Dasararaju H.K., *Big Data Management*. W: Bhimasankaram Pochiraju B., Seshadri S. (red.), *Essentials of Business Analytics An Introduction to the Methodology and its Applications,* (s. 71-109), Springer International Publishing, Cham 2019, s. 73. [↑](#footnote-ref-5)
6. OECD, *Going Digital: Shaping Policies, Improving Lives*,OECD Publishing, Paris 2019, https://doi.org/10.1787/9789264312012-en. [↑](#footnote-ref-6)
7. Współczynniki konwersji w e-commerce pokazuje odsetek zawartych transakcji względem sesji użytkowników, przy czym sesja odnosi się do wizyty osoby na witrynie e-sklepu przez określony czas, zwykle 30 minut. [**https://www.conversion.pl/blog/wspolczynnik-konwersji/**](https://www.conversion.pl/blog/wspolczynnik-konwersji/) (dostęp 28.10.2021). [↑](#footnote-ref-7)
8. A. Pluta-Zaremba, *Innowacje cyfrowe w branży e-commerce*. W: K. Nowicka (red.) *Biznes cyfrowy. Perspektywa innowacji cyfrowych*, Oficyna Wydawnicza SGH, Warszawa 2019, s. 90. [↑](#footnote-ref-8)