

BIG DATA, NAUKA O DANYCH I AI BEZ TAJEMNIC



**Podjmuj
lepsze decyzje
i rozwijaj swój biznes!**

DAVID STEPHENSON

Tytuł oryginału: Big Data Demystified: How to use big data, data science and AI to make better business decisions and gain competitive advantage

Tłumaczenie: Wojciech Bombik

ISBN: 978-83-283-5796-9

© Pearson Education Ltd 2018 (print and electronic)

This translation of Big Data Demystified is published by arrangement with Pearson Education Limited.

Polish language edition published by HELION SA. Copyright © 2019.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from the Publisher.

Wszelkie prawa zastrzeżone. Nieautoryzowane rozpowszechnianie całości lub fragmentu niniejszej publikacji w jakiegokolwiek postaci jest zabronione. Wykonywanie kopii metodą kserograficzną, fotograficzną, a także kopiowanie książki na nośniku filmowym, magnetycznym lub innym powoduje naruszenie praw autorskich niniejszej publikacji.

Wszystkie znaki występujące w tekście są zastrzeżonymi znakami firmowymi bądź towarowymi ich właścicieli.

Autor oraz Helion SA dołożyli wszelkich starań, by zawarte w tej książce informacje były kompletne i rzetelne. Nie biorą jednak żadnej odpowiedzialności ani za ich wykorzystanie, ani za związane z tym ewentualne naruszenie praw patentowych lub autorskich. Autor oraz Helion SA nie ponoszą również żadnej odpowiedzialności za ewentualne szkody wynikłe z wykorzystania informacji zawartych w książce.

Helion SA

ul. Kościuszki 1c, 44-100 Gliwice

tel. 32 231 22 19, 32 230 98 63

e-mail: helion@helion.pl

WWW: <http://helion.pl> (księgarnia internetowa, katalog książek)

Drogi Czytelniku!

Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres

<http://helion.pl/user/opinie/bigdata>

Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Printed in Poland.

- Kup książkę
- Poleć książkę
- Oceń książkę

- Księgarnia internetowa
- Lubię to! » Nasza społeczność

Spis treści

O autorze 7

Podziękowania 8

Wprowadzenie 9

Część 1 Wprowadzenie do koncepcji big data 15

1 Historia big data 19

2 Sztuczna inteligencja, uczenie maszynowe i big data 37

3 Dlaczego technologie big data są użyteczne? 49

4 Przykłady zastosowań big data w analizie danych 61

5 Zrozumieć ekosystem big data 81

Część 2 Jak wykorzystać ekosystem big data w swojej organizacji? 93

6 W jaki sposób big data może pomóc w kierowaniu strategią? 97

7 Strategia wykorzystywania big data oraz data science 117

8 Wykorzystanie data science — analityka, algorytmy
i uczenie maszynowe 135

9 Wybór technologii 159

10 Budowanie zespołu 173

11 Zarządzanie danymi i kwestie prawne 193

12 Skuteczne realizowanie projektów 205

Przypisy 215

Słowniczek 223

Skorowidz 231

6

Rozdział

W jaki sposób big data może pomóc w kierowaniu strategią?

Oceniając strategię firmy, a być może nawet biorąc pod uwagę dużą zmianę w dalszym kierunku jej rozwoju, dobrze będzie zebrać i wykorzystać wszystkie dostępne dane w celu dogłębnego zrozumienia swoich **klientów**, **konkurencji** i **zewnętrznych czynników**, które wpływają na firmę, a nawet swojego własnego **produktu**. Ekosystem *big data* odgrywa ważną rolę w tym procesie, umożliwiając uzyskiwanie wiedzy i kierowanie działaniami w sposób dotychczas niemożliwy.

Twoi klienci

Dane dotyczące klientów to jeden z największych atutów, jakie posiadasz. W dzisiejszych czasach mamy dostęp do większej ilości informacji niż kiedykolwiek wcześniej i możesz z nich dowiedzieć się więcej niż byś się spodziewał o swoich obecnych i potencjalnych klientach. Kim są, co ich motywuje, jakie są ich preferencje i zwyczaje. Im więcej danych zbierzesz, tym pełniejszy obraz klientów możesz uzyskać, więc zbieranie danych z tylu źródeł, z ilu tylko się da, powinno stać się Twoim celem.

Zdobywanie danych

Rozpocznij od zidentyfikowania wszystkich punktów interakcji z klientami:

- Wizyty na platformach cyfrowych, takich jak strony internetowe, aplikacje itp.
- Kontakt ze wsparciem klienta przez telefon, e-mail, czat itp.
- Media społecznościowe, w tym przesłane wiadomości, tweety, posty opublikowane na Twoim koncie lub kontaktach klientów.

- Wszystkie zapisy ruchu, uzyskane przy wykorzystaniu nagrań wideo i czujników ruchu ze sklepów. Istnieje wiele technologii, które mogą pomóc przy monitorowaniu ruchu, takich jak czujniki, Wi-Fi, Bluetooth, beacons, a nawet aplikacje na smartfona wykorzystujące zmiany częstotliwości światła.
- W niektórych branżach możesz mieć dostęp do dodatkowych danych z czujników, identyfikatorów RFID, aplikacji fitnessowych itp., które mogą dostarczyć takich danych jak odczyty biomedyczne, dane z akcelerometrów, termometrów itp.

Dla każdego punktu interakcji zastanów się:

- jakie dane możesz zbierać,
- jakie są potencjalne możliwości użycia tych danych,
- jakie ograniczenia prawne dotyczące prywatności należy wziąć pod uwagę (patrz rozdział 11.).

Dla wielu organizacji na takie punkty interakcji będą składały się sklepy, strony sieciowe, aplikacje na Androida i iOS, kanały w mediach społecznościowych, personel obsługujący klientów w sklepach, call center i czaty. Zilustruję je kilkoma przykładami.

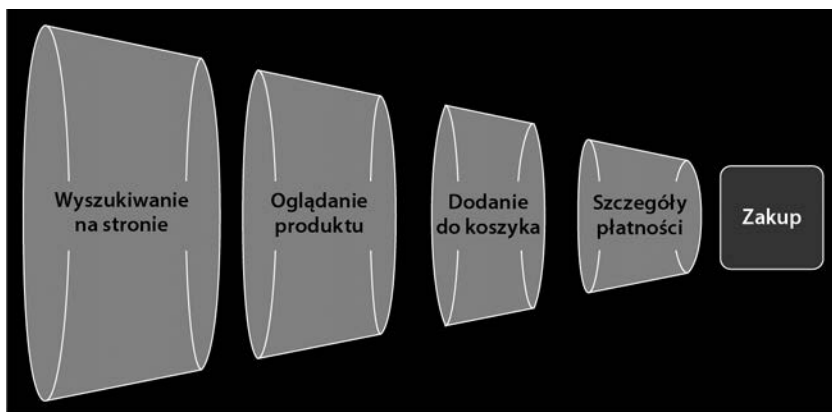
Platformy cyfrowe

Zacznijmy od platform cyfrowych. Na początek od podstaw, pomijając na razie technologie *big data*.

Zapewne już wykorzystujesz na swojej stronie jakieś narzędzia analityczne, które zapisują wysokopoziomowe zdarzenia. Upewnij się, że zapisujesz też najważniejsze punkty ścieżek zakupowych, takie jak wyszukiwania, używane filtry, odwiedzanie konkretnych podstron, ściągane pliki, odtworzone filmy lub towary umieszczane w koszyku. Zapisuj akcje wykonywane myszką, takie jak scrollowanie lub umieszczenie kursora nad danym obiektem. Upewnij się, że te zdarzenia są zapisywane wraz z etykietami zawierającymi dodatkowe szczegóły, które będą potrzebne przy późniejszych analizach, takie jak kategoria produktu, zakres cen czy ID produktu. Umożliwi to szybkie analizowanie najbardziej interesujących zagadnień, takich jak np. sprawdzenie, jak często produkty z danej kategorii były oglądane lub jak skuteczna była kampania reklamowa pod kątem konkretnych działań klientów. Na koniec prawdopodobnie uzyskasz kilkadziesiąt lub nawet kilkaset wymiarów, które będziesz mógł wykorzystać w gotowych narzędziach analitycznych. To jeszcze nie jest *big data*.

Wykorzystanie przy zapisie etykiet z dodatkowymi szczegółami umożliwi dalszą analizę i zrozumienie wielu aspektów ścieżek zakupowych klientów, dając wgląd w to, jak różne typy klientów reagują na prezentowane im produkty. Poniżej pokażę to na konkretnym przykładzie.

Jeżeli jeszcze tego nie zrobiłeś, zdefiniuj lejki konwersji, składające się z kolejnych zdarzeń, które prowadzą do ważnego celu konwersji, takiego jak na przykład dokonanie zakupu. Rysunek 6.1 pokazuje, jak może wyglądać taki lejek konwersji.

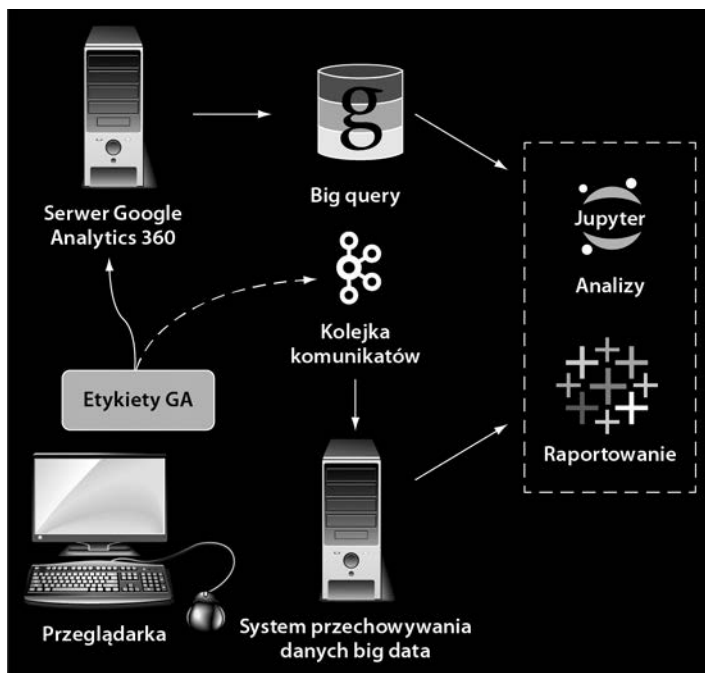


Rysunek 6.1. Przykładowy lejek konwersji

Wszystkie pośrednie cele w lejku konwersji, tak zwane mikrokonwersje, prowadzą razem do makrokonwersji (w tym przypadku jest to zakup). Wybieraj mikrokonwersje w sposób, który odzwierciedla rosnące zaangażowanie oraz prawdopodobieństwo końcowego zdarzenia. Zdefiniowane lejki pozwalają na analizę współczynnika rezygnacji na każdym etapie, umożliwiając zauważenie potencjalnych wrażliwych punktów i zwiększenie prawdopodobieństwa przechodzenia użytkowników do kolejnych etapów lejka, aż do końcowego zdarzenia. W zależności od rodzaju produktu przejście przez cały lejek może trwać kilka dni, tygodni lub nawet miesięcy, więc należy się dobrze zastanowić, co można uznać za rezygnację.

Pamiętaj o zachowaniu zgodności z przepisami dotyczącymi używania plików *cookie*. Przechowuj informacje dotyczące tego, jakie dane powiązane z konkretnymi użytkownikami zapisujesz (np. adres IP) i w jaki sposób je później wykorzystujesz. Na przykład, jeżeli personalizujesz zawartość strony lub działania marketingowe na podstawie działań użytkownika w sieci, powinieneś rozważyć też stronę etyczną oraz prawną takiego zagadnienia. Przypomnij sobie opisany przykład firmy Target.

Teraz możemy przejść do części związanej z technologiami *big data*. Jak na razie mamy narzędzia analityczne zapisujące najbardziej istotne dane ze strony internetowej. Teraz można podłączyć stronę pod system *big data*, który będzie zapisywał wszystkie działania każdego odwiedzającego stronę. Potrzebny będzie odpowiedni system do przechowywania danych, taki jak HDFS, oraz skrypty zbierające zdarzenia na stronie (zazwyczaj używa się do tego skryptów JavaScript). Jeżeli interesuje Cię rozwiązanie wymagające najmniej pracy, można rozważyć konto premium w Google Analytics (GA360) i aktywować w nim integrację z usługą BigQuery. Dzięki temu dane ze strony będą wysyłane do chmury Google, umożliwiając szczegółowe analizy w ciągu kilku godzin. Jeżeli potrzebujesz rozwiązania działającego w czasie rzeczywistym, wystarczy zmienić w Google Analytics metodę `sendHitTask` (JavaScript) tak, by przesyłała te same dane zarówno do chmury, jak i do Twojego własnego systemu. Taka architektura jest przedstawiona na rysunku 6.2. Zwróć uwagę na to, że regulamin Google Analytics nie pozwala na przesyłanie danych osobowych (określanych jako **PII** — ang. *personally identifiable information*). Omówimy tę kwestię szerzej w rozdziale 11.



Rysunek 6.2. Przykładowa architektura do strumieniowania danych big data (źródło: ikony BigQuery, Apache Kafka, Jupyter notebooks i Tableau)

Masz już więc surowe dane potrzebne do bardzo szczegółowego zrozumienia swoich klientów, co opiszemy w dalszej części rozdziału.

Wsparcie klienta

Zastanówmy się nad zapisywaniem i analizowaniem wszystkich interakcji między klientami i sprzedawcami oraz wsparciem klienta: rozmów telefonicznych, czatów, maili, a nawet zapisów wideo ze sklepów. Większość z tych danych jest łatwa do przeanalizowania w pojedynczych fragmentach, ale żeby zrobić to samo w większej skali, potrzebne będą zaawansowane narzędzia. Przy zapisie tego typu informacji pracownicy wsparcia klienta powinni dostarczać dodatkowych informacji, takich jak ID klienta czy też pora dnia, i oznaczyć je kategoriami, na przykład „zapytanie o zamówienie”, „nowy zakup”, „anulowanie” lub „reklamacja”. Następnie całość informacji można zapisać w bazie przeznaczonej do danych *big data* (na przykład MongoDB lub HDFS). W dalszej części rozdziału pokażemy, w jaki sposób używać takich danych.

Fizyczne położenie klientów

Istnieje kilka technologii przeznaczonych do monitorowania przemieszczania się klientów w sklepie. Oprócz tradycyjnych kamer oraz laserowych czujników ruchu przy wejściach niektóre technologie umożliwiają śledzenie smartfonów przy użyciu sygnału z telefonu, Wi-Fi lub Bluetooth. Rozwiązania tego typu oferują firmy ShopperTrak lub Walkbase. Tego typu monitoring pomaga zrozumieć wzorce zachowań klientów, na przykład jakie kategorie produktów są rozważane przez tych samych klientów lub ile czasu zajmuje podjęcie decyzji o zakupie. Pomaga to kierować personel sklepu w miejsca, gdzie jest on najbardziej potrzebny. Tego typu dane są wartościowe, nawet jeżeli przy ich zbieraniu zachowujemy anonimowość klientów.

Gdy klient podchodzi do kasy, by dokonać zakupu, być może nawet przy użyciu karty, która jest z nim skojarzona, możesz zobaczyć nie tylko, co zamierza kupić, ale również jakie rejony sklepu odwiedził. Takie informacje mogą być pomocne w przyszłych działaniach marketingowych lub zmianie układu sklepu, jeżeli zauważysz, że obecny może niekorzystnie wpływać na możliwości sprzedaży krzyżowej.

Oto zaledwie kilka przykładów. Dobrze jest zacząć zbierać i przechowywać tak szczegółowe dane, jak to tylko możliwe, biorąc pod uwagę ich wartość biznesową, poszanowanie prywatności klientów i zachowanie zgodności

z prawem podczas zbierania, przechowywania oraz używania tych danych. Uważaj też, by nie przekroczyć cienkiej linii między „pomocny” a „nachalny”. Miej na uwadze najlepszy interes swoich klientów i pamiętaj, że prędzej czy później dowiedzą się oni o wszystkich stosowanych przez Ciebie technikach.

Zapamiętaj

Za każdy kawałek prywatności, który zabierasz swoim klientom, staraj się dać im jakąś użyteczną usługę. Na przykład, jeżeli Twoja aplikacja śledzi fizyczne położenie klienta, upewnij się, że dostaje on dzięki temu wartościowe informacje związane z jego położeniem. Zapewnij również lepszą, spersonalizowaną obsługę klientom, którzy logują się na Twojej stronie. W ten sposób interesy Twoje i Twoich klientów będą częściowo się pokrywać.

Łączenie danych o klientach

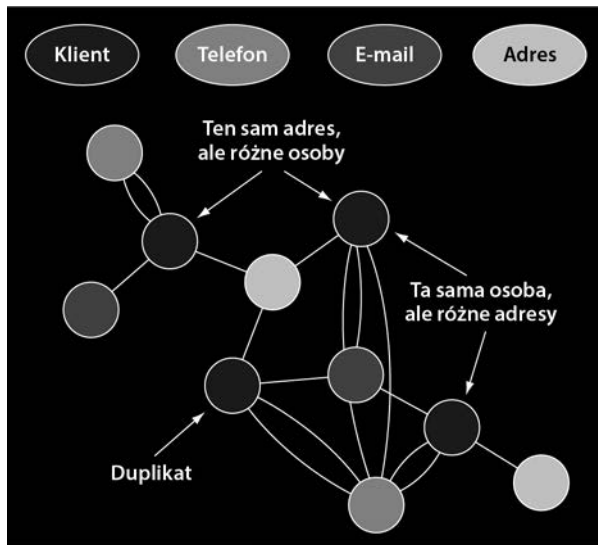
Połącz dane pochodzące ze wszystkich punktów interakcji z klientem, aby uzyskać całościowy obraz ścieżek zakupowych. Jeżeli klient dzwoni do call center po sprawdzeniu na stronie warunków rozwiązania umowy, powinieneś móc połączyć te dwa zdarzenia w swoim systemie. By tego dokonać, potrzebujesz pola identyfikującego konkretnego klienta (takiego jak telefon lub nazwa użytkownika).

Jeżeli monitorujesz drogę, jaką porusza się klient w sklepie, połącz ją z informacją z kasy o dokonanych zakupach (do rozważenia pozostaje tutaj kwestia prywatności). Możesz to zrobić, łącząc dokładny czas zakupu w kasie z trasą przejścia przez sklep. Połącz te dane w centralnym systemie, by uzyskać pełny obraz.

Czasami możesz wykorzystać anonimowe dane klientów, na przykład gdy analizujesz przepływ całkowitego ruchu przez sklep. Czasami jednak przydają się dane połączone z konkretnym klientem, na przykład przy analizie wszystkich jego działań. W tym drugim przypadku dobrze będzie zdeduplikować dane o klientach. Jest to trudne zadanie i prawdopodobnie osiągniesz na tym polu umiarkowany sukces. Najlepsza sytuacja jest wtedy, gdy klient zawsze używa unikatowego identyfikatora, korzystając z Twoich usług. W sklepie sieciowym może to być trwały, unikatowy login (tak jak w przypadku np. Facebooka). W życiu rzeczywistym takim unikatowym identyfikatorem jest zazwyczaj dokument ze zdjęciem. W większości przypadków nie będziesz miał tego luksusu, więc dołóż wszelkich starań, by dobrze połączyć interakcje z klientami.

Zazwyczaj przy identyfikacji klientów możesz spotkać się z takimi problemami:

- **Problem:** Klient się nie identyfikuje (na przykład nie loguje się na stronie).
- **Możliwe rozwiązania:** Użyj plików *cookie* oraz adresu IP do połączenia wszystkich wizyt tego samego klienta, by stworzyć całościowy obraz ścieżek zakupowych osobnego, anonimowego klienta. Użyj szczegółów płatności, by połączyć zakupy z konkretnym klientem. Aplikacje zainstalowane na smartfonie mogą dostarczyć dodatkowych informacji; porozmawiaj na ten temat z twórcami swojej aplikacji.
- **Problem:** Klienci używają wielu loginów.
- **Możliwe rozwiązania:** Przeszukaj bazę klientów pod kątem użytkowników z takimi samymi nazwiskami, mailami, adresami, datami urodzin lub adresami IP. Grafowa baza danych, taka jak Neo4j, może pomóc w tym procesie, tak jak jest to pokazane na rysunku 6.3. Opracuj razem ze stroną biznesową logikę opisującą, które konta można scalać, a które powiązać, używając specjalnego pola (na przykład „mąż/żona”). Zmień proces tworzenia nowych kont w taki sposób, aby wykryć i ograniczyć możliwość tworzenia zduplikowanych kont, na przykład oznaczając maile powiązane z już istniejącymi kontami.



Rysunek 6.3. Grafowa baza danych może pomóc w deduplikacji klientów

Używanie danych

Większość danych Twoich klientów będzie przydatna, nawet w zagregowanej i zanonimizowanej formie, w jakiej dostarczają je standardowe narzędzia do analityki sieciowej. Będziesz dzięki nim widział, jak wielu klientów pojawia się w danych godzinach, uzyskasz też przydatne informacje, takie jak średni czas spędzony na stronie, liczba odwiedzonych stron, jak wielu klientów trafia na Twoją stronę przez inne strony lub dzięki kampaniom marketingowym itp. Dowiesz się również, jaki jest całkowity obrót wygenerowany przez konkretne grupy klientów w podziale np. względem położenia lub sposobu pozyskania. To da Ci obraz tego, w jaki sposób i kiedy Twoje produkty są używane, w szczególności, gdy połączysz te informacje z danymi o kampaniach marketingowych, dniach wolnych, przerwach w działaniu usług lub nowych inicjatywach.

Ścieżka zakupowa

Dane przechowywane w zbiorach *big data* stają się znacznie bardziej użyteczne, gdy używasz ich do właściwego zrozumienia zamiarów, preferencji i zwyczajów swoich klientów. Powinieneś już móc podzielić swoich klientów na tak zwane **persony** na podstawie cech statycznych, takich jak adres, płeć, język, wiek oraz prawdopodobny poziom dochodów. Jeżeli włączysz odpowiednie ustawienia, Google Analytics może dostarczyć Ci dodatkowych danych demograficznych (przy użyciu plików *cookie* stworzonych przez firmę DoubleClick).

Możesz rozszerzyć kryteria segmentacji klientów tak, by wykorzystać dane ze ścieżek zakupowych, takich jak wymienione poniżej:

- **Jakich filtrów używają najczęściej? Czy jest to sortowanie po cenie od najwyższej? Od najniższej? Najlepiej oceniane produkty? Najnowsze produkty?** Klienci, którzy sortują od najniższej ceny, prawdopodobnie dobrze orientują się w cenach. Dla tych, którzy sortują od najwyższej ceny lub po ocenach produktu, prawdopodobnie większą rolę odgrywa jakość. Ci, którzy sortują po najnowszych produktach, również skupiają się na jakości lub są entuzjastami nowinek. Użytkownicy sortujący po ocenach mogą też być tzw. *late adopters*, czyli klientami, którzy wolą poczekać, aż produkt trochę się „uleży” na rynku. Wszystkie te informacje wpływają na to, w jaki sposób podejmujesz działania wobec takich klientów. Jeżeli dla danego klienta jakość jest ważniejsza od ceny, powinieneś w wynikach wyszukiwania oraz w mailach marketingowych pokazywać mu

produkty wysokiej jakości, natomiast nie powinniśmy kierować do niego informacji o wyprzedażach lub produktach, które dopiero co się pojawiły. Wobec klientów stawiających na pierwszym miejscu cenę należy postępować dokładnie na odwrót.

- **Ile przedmiotów zazwyczaj przeglądają, zanim dokonają zakupu?** Ta informacja pomoże Ci zdecydować, kiedy należy zadziałać podczas procesu sprzedaży, na przykład oferując zniżkę, gdy podejrzewamy, że klient zamierza zrezygnować z dokonania zakupu.
- **Jakie kategorie odwiedzają najczęściej?** Może być to pomocne przy segmentacji klientów oraz przy zwracaniu najlepszych wyników wyszukiwania niejednoznacznych fraz (na przykład „jaguar” jako samochód lub zwierzę albo „Panama” jako miasto na Florydzie lub kraj). Informacja ta będzie też dobrym wskaźnikiem, jakie towary reklamować danemu klientowi.
- **Czy zmieniają sposób wysyłki, aby zaoszczędzić?** Ponownie jest to sygnał, że mamy do czynienia z klientami przykładającymi dużą wagę do ceny.
- **Czy czytają lub piszą opinie o produktach?** Jeżeli klient za każdym razem czyta opinie, nie należy w materiałach reklamowych, wynikach wyszukiwania albo podczas sprzedaży krzyżowej prezentować mu produktów ze słabymi ocenami. Jeżeli często pisze opinie lub posiada w mediach społecznościowych konto z wyjątkowo dużą liczbą śledzących, upewnij się, że otrzyma obsługę z najwyższej półki.
- **Na jaki rodzaj działań marketingowych najlepiej odpowiadają? Czy czytają newslettery? Czy robią zakupy podczas błyskawicznych wyprzedaży (tzw. *flash sales*)?** Nie wysyłaj klientom materiałów reklamowych, jeżeli na nie nigdy nie odpowiadają. Wybierz najodpowiedniejszy sposób kontaktu, a zwiększysz szansę na to, że zareagują lepiej, niż klikając „wypisz się z newslettera”.
- **Jeżeli są aktywni w mediach społecznościowych, jakich hashtagów najczęściej używają? Czy możesz wykorzystać tę wiedzę do skuteczniejszych działań marketingowych?** Ponownie, dzielisz klientów na osoby przy użyciu wszystkich istotnych informacji. Spróbuj wykorzystać ich konta w mediach społecznościowych. Tak jak powiedzieliśmy w rozdziale 4., działania podejmowane w sieci mogą sporo powiedzieć o klientach.

Grupy klientów (persony)

Używając zebranych danych, powinieneś określić, które czynniki są najistotniejsze przy dzieleniu klientów na segmenty, tak zwane persony. Przykładem mogą być „mężczyźni w wieku od 20 do 30 lat, zwracający uwagę na cenę” lub „wydający dużo pieniędzy miłośnicy nowinek, którzy podejmują szybkie decyzje podczas zakupów” albo „klienci, którzy codziennie przeglądają konkretne kategorie, ale kupują tylko podczas zniżek”. Możesz tworzyć takie segmenty w sposób jakościowy, przy użyciu intuicji ekspertów od marketingu, lub możesz robić to w sposób ilościowy, przy użyciu narzędzi analitycznych, takich jak **klasteryzacja** oraz **analiza głównych składowych**. Obydwa podejścia są poprawne, ale jeżeli masz dostęp do dużej ilości danych, które mogą zostać pogrupowane na wiele sposobów, podejście ilościowe prawdopodobnie da lepsze efekty.

Towary

Dane ze ścieżek zakupowych mogą dać Ci lepszy wgląd w to, jak towary wpływają na doświadczenia klientów. Jeżeli rozważasz usunięcie towarów z niskim poziomem sprzedaży, ścieżki zakupowe mogą pokazać wysoki odsetek zyskownych transakcji dokonanych przez klientów, którzy trafili na Twoją stronę, szukając tych mało opłacalnych towarów. Dobrze jest przyjrzeć się lepiej ścieżkom zakupowym takich klientów — czego szukali i dlaczego podjęli takie, a nie inne decyzje — zanim postanowisz zrezygnować z towarów, które mogą przyciągać pewnych klientów do sklepu.

Z drugiej strony, możesz sprzedawać produkt, który często pojawia się w wynikach wyszukiwania, ale nie cieszy się zainteresowaniem klientów, co pokazują ich ścieżki zakupowe. W takim przypadku powinieneś albo zmienić sposób wyświetlania wyników wyszukiwania, albo zupełnie usunąć dany towar, żeby nie zajmował miejsca w wynikach wyszukiwania.

Krytyczne interwencje

Wykorzystuj zarówno podstawową analitykę, jak i zaawansowane uczenie maszynowe do badania danych swoich klientów, a prawdopodobnie znajdziesz sposoby na zmniejszenie ich migracji i zwiększenie sprzedaży. Podstawowa analiza tego, gdzie i kiedy klienci są aktywni, pomoże w ustaleniu odpowiedniego harmonogramu dla swoich pracowników. Może też zasygnalizować prawdopodobne kolejne kroki klienta (na przykład klient odwiedzający sklep z AGD dwa razy w ciągu poprzedniego tygodnia może planować większy zakup).

Przy nieco dokładniejszych analizach możesz wykrywać subtelne, ale ważne sygnały. Jedna z firm telekomunikacyjnych przeanalizowała ostatnio dane dotyczące klientów rezygnujących z umów i odkryła, że znaczna ich część przed rezygnacją przeszła przez te same trzy lub cztery kroki: sprawdzenie szczegółów swojej umowy na stronie, telefon do wsparcia klienta, następnie wizyta w salonie i kwestionowanie rachunku, a na końcu rozwiązanie umowy. Łącząc te zdarzenia, firma zidentyfikowała sygnały nadchodzącej migracji klienta i mogła podjąć odpowiednie działania.

Na jeszcze wyższym poziomie techniki uczenia maszynowego mogą wykryć zbliżające się usunięcie konta lub prawdopodobieństwo sprzedaży na podstawie analizy tekstu albo nagrań audio lub wideo. Taki system może być znaczącą inwestycją czasu i zasobów, ale warto poznać jakiś przypadek biznesowy, który pomoże to właściwie ocenić lub znaleźć dostawcę, który już rozwinął technologię pasującą do Twoich potrzeb, tak jak jest to opisane w poniższym studium przypadku.

Studium przypadku

Istotne informacje o klientach ze sklepowych nagrań wideo

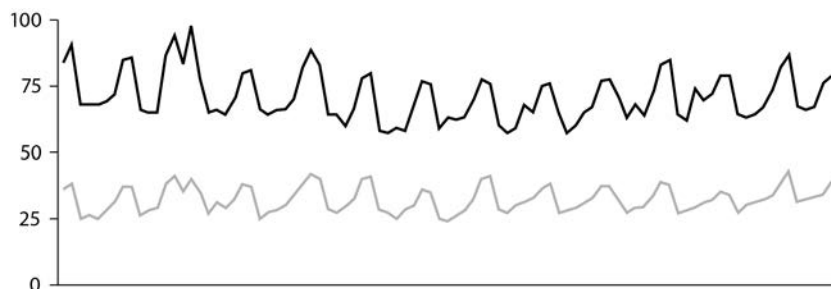
Magazyn „The Economist” opisał ostatnio dwie innowacje w dziedzinie wideoanalizy. Realeyes, firma z Londynu zajmująca się wykrywaniem emocji, odkryła, że kupujący, którzy uśmiechają się, wchodząc do sklepu, wydają trzy razy więcej niż inni. W innym programie pilotażowym sieć księgarni zaczęła używać oprogramowania firmy Angus.ai do monitorowania klientów, którzy doszli do końca sklepu, a następnie wracając, mieli zmarszczone brwi. W takich przypadkach była wysyłana dyskretna informacja do obsługi sklepu, aby ta mogła odpowiednio zareagować. W rezultacie sklep odnotował 10-procentowy wzrost sprzedaży³⁸.

Jak zawsze w takich przypadkach, należy skonsultować się z prawnikiem, by nie naruszyć granic prywatności i działać w zgodzie z najlepszym interesem klientów, a także by nie zrobić niczego, co mogłoby zaszkodzić wizerunkowi sklepu po ujawnieniu takich działań klientom.

Twoja konkurencja

Szczególne wyzwanie stanowi uzyskanie informacji na temat konkurencji. Brokerzy informacji, tacy jak Nielsen, Comscore lub SimilarWeb, mogą na zamówienie oszacować ruch na stronach i w aplikacjach konkurencyjnych firm, być może dołączając nawet informacje o stronach, z których następowały

przekierowania. Na stronie <https://trends.google.com/trends/> można zobaczyć wykres liczby wyszukiwań danej frazy, co umożliwia porównanie się z konkurencyjnymi firmami pod względem wyszukiwań (patrz rysunek 6.4).



Rysunek 6.4. Wyszukiwania frazy „McDonalds” (górną linią) oraz „Burger King” (dolną linią) w Brazylii w drugim kwartale 2017 (Google Trends)

Informacje o towarach, usługach lub fizycznych lokalizacjach konkurencji można uzyskać przez tak zwany *web scraping*, czyli zautomatyzowane wyciąganie informacji bezpośrednio ze stron internetowych. Można użyć do tego narzędzi, takich jak na przykład Selenium. Jeżeli starasz się konkurować ceną, będziesz chciał dopasowywać ceny, kierując się bliskością konkurencji. W przypadku lokalizacji fizycznych można ją określić przy użyciu adresów oraz dróg przesyłania towaru. Dla sklepów sieciowych częściowy wpływ na kształtowanie cen będzie miała informacja o tym, skąd klienci trafiają na Twoją stronę. Jeżeli ze stron porównujących ceny, to powinno być to sygnałem dużego ryzyka tego, że kupią u konkurencji.

Pracuj nad zwiększeniem swojego udziału w rynku, czyli procenta wydawanych przez klientów pieniędzy, które zasilają Twój biznes, a nie konkurencję. Rozpocznij, wykorzystując szczegółowe dane o klientach, które zebrałeś, i przeanalizuj, jakie kategorie oraz konkretne produkty są zazwyczaj kupowane przez te same segmenty klientów. Będziesz mógł zobaczyć, którzy klienci już dokonują takich krzyżowych zakupów, którzy tylko przeglądają dane kategorie, ale nie kupują, a którzy pozostają aktywni tylko w obrębie jednej kategorii.

Zidentyfikuj produkty, które Twoi klienci kupują gdzie indziej, by zobaczyć, gdzie tracisz udziały w rynku. Jeżeli sprzedajesz produkty spożywcze, a klient kupuje tylko owoce i warzywa, to możesz zgadywać, że mleko i jajka kupuje w innym sklepie. Jeżeli sprzedajesz elektronikę, a klient kupuje tylko smartfony, to znaczy, że komputery kupuje zapewne gdzie indziej. W ten sposób możesz poznać obszary, w których powinienś mocniej konkurować. Wyko-

rzystując utworzone segmenty klientów, możesz zobaczyć, czy konkurencja przemawia bardziej do klientów skupiających się na jakości, tych reagujących na akcje marketingowe, takich, którzy wydają dużo pieniędzy itp.

Przeglądaj sieciowe strony z ogłoszeniami o pracy, by wykryć różne posunięcia konkurencji. Znaczny wzrost liczby ogłoszeń dotyczących danej lokalizacji lub stanowiska może oznaczać działania podejmowane w konkretnym obszarze. Utwórz listę profili pracowników konkurencji na LinkedIn i monitoruj je pod względem nietypowych aktualizacji. Jeżeli wyjątkowo duża liczba pracowników aktualizuje swoje profile, może to być sygnałem zamieszania lub nadchodzących zwolnień w firmie. Podobnie analiza wykorzystująca przetwarzanie języka naturalnego, badająca publiczne oświadczenia firmy, może pomóc odkryć nietypowe działania. Ta technika została skutecznie wykorzystana do sygnalizowania zbliżających się pierwszych ofert publicznych przy sprzedaży papierów wartościowych.

Czynniki zewnętrzne

Na Twoją strategię ma wpływ szeroki zakres czynników, od regulacji rządowych po lokalną pogodę. Jeżeli zajmujesz się branżą turystyczną, dni wolne od pracy będą wpływały na rezerwacje robione z dużym wyprzedzeniem, a pogoda na te dokonywane w ostatnim momencie. Ceny surowców wpłyną na koszty produkcji oraz transportu, a kursy wymiany walut lub zawirowania polityczne dotkną działań podejmowanych za granicą.

Większość takich czynników można wychwycić, analizując tradycyjne źródła danych, ale zasoby *big data* mogą dostarczyć dodatkowych, wartościowych sygnałów. Uważnie obserwuj zmiany w aktywności klientów w sieci, gdyż mogą one sygnalizować pojawienie się niespodziewanych czynników, które będą wymagać Twojej uwagi. Przykładowo, dzięki Google Maps lub Waze możesz wykryć nowe place budowy lub zamknięcia dróg, analizując natężenie ruchu drogowego.

Innym przykładem może być to, że pojawienie się nowego, innowacyjnego produktu możesz zauważyć, wyszukując go w sieci, ale możesz to też zrobić, obserwując wpływ, jaki wywiera na sprzedaż Twoich innych produktów. Jeżeli prowadzisz sieć hoteli i masz jeden z nich w Scranton w Pensylwanii, możesz nie wiedzieć o tym, że w drugim tygodniu lutego ma tam miejsce duża konferencja na jakiś niejasny temat. Jeżeli przygotowujesz prognozę rezerwacji na luty, zauważysz niespodziewany wzrost aktywności użytkowników strony z rezerwacjami i call center już w październiku, zanim w ogóle

dowiesz się o tej konferencji. Monitorując zachowania klientów, możesz podnieść stawki za pokój jeszcze w październiku, zanim wynajmiesz niedoszacowane pokoje na luty kilka tygodni później.

W tym celu powinieneś stworzyć regularne prognozy kluczowych metryk, takich jak liczba odwiedzin, oraz prognozę sprzedaży. Można to zrobić, łącząc dane historyczne, przewidywania stóp wzrostu oraz wiedzę biznesową o wszystkich nietypowych czynnikach (dni wolne, ważniejsze wydarzenia, zmiany prawne lub gospodarcze itp.). Takie prognozy powinny być podzielone na segmenty aż do poziomu, na którym możesz sterować odpowiednimi działaniami (na przykład do poziomu produktu lub regionu), i najlepiej, aby były aktualizowane codziennie. Jeżeli automatycznie monitorujesz prognozowane wartości w cyklach dziennych lub tygodniowych, możesz podjąć odpowiednie działania, gdy tylko przekroczą spodziewane wartości, sygnalizując, że jakiś zewnętrzny czynnik wpływa na biznes w nieoczekiwany sposób.

Twój własny produkt

Musisz naprawdę dobrze rozumieć oferowane przez siebie usługi oraz produkty podczas tworzenia strategii. Możesz nie rozumieć ich tak dobrze, jak Ci się wydaje, a Twoi klienci mogą postrzegać je w zupełnie inny sposób, niżbyś się spodziewał. Co działa dobrze, a co nie? Jak klienci reagują na Twoje produkty? Gdzie tracisz pieniądze przez nieefektywne działania?

Jeżeli znaczna część Twojej oferty znajduje się w sieci, dowiedz się, jakie rozwiązania sprawdzają się tam dobrze i co możesz poprawić. Zdefiniuj i śledź mikrokonwersje, by zobaczyć, jak wypadają Twoje produkty jeszcze przed zakupem. Mogą one dostarczyć cennych informacji, nawet jeżeli nie są częścią analizy lejków konwersji.

Śledź zaangażowanie klientów na swoich platformach cyfrowych.

- Jaki jest **współczynnik uznania** (ang. *applause rate*) w Twoich mediach społecznościowych? (Jak często Twoje tweety są udostępniane, ile mają polubień? Ile komentarzy pojawia się pod Twoimi postami na Facebooku?)
- Ile osób ściąga Twoje materiały?
- Ile osób zapisuje się do newslettera?

Przetestuj zmiany w swoich produktach przy pomocy testów A/B. Możesz to zrobić w następujący sposób:

1. Zaproponuj jedną małą zmianę, która może ulepszyć Twoją ofertę. Zmień jedną ramkę, jedną frazę lub jeden banner. Upewnij się ze swoim zespołem inżynierów, że jest to zmiana łatwa we wprowadzeniu.
2. Zdecyduj, jakie kluczowe wskaźniki efektywności (**KPI**) najbardziej chciałbyś zwiększyć: dochód, liczbę zakupów, sprzedaż dodatkową, czas spędzony na stronie itp. Badaj także wpływ na inne wskaźniki.
3. Uruchom równocześnie wersję oryginalną i zmienioną (A i B). Dla stron internetowych możesz użyć odpowiednich narzędzi, takich jak Optimizely. Zbadaj wyniki przy użyciu wybranego narzędzia lub umieść identyfikator wersji testowej w tagach na stronie i porównaj wersję, badając na przykład długości dróg prowadzących do konwersji.
4. Przy pomocy testu istotności różnic dla dwóch prób niezależnych sprawdź, czy rezultaty są statystycznie istotne. Jeżeli nie masz analityka, który może się tym zająć, możesz skorzystać z kalkulatora dostępnego na stronie <https://abtestguide.com/calc/>.
5. Użyj systemu *big data* do głębszej analizy:
 - a) Czy były istotne zmiany w ścieżkach zakupowych, takie jak liczba odwiedzonych kategorii lub użyte filtry?
 - b) Czy jest jakiś kluczowy produkt lub grupa klientów, które powinienś potraktować inaczej?
 - c) Czy jakieś zewnętrzne czynniki wpływają na rezultaty?
 - d) Czy wskaźniki efektywności przesuwają się w innych kierunkach?

Dostosuj swoje założenia dotyczące produktów do tych nowych spostrzeżeń. Na przykład:

- Czy próbujesz konkurować ceną, podczas gdy większość Twoich dochodów pochodzi od klientów, którzy na pierwszym miejscu stawiają jakość?
- Czy poświęcasz za mało czasu na recenzje klientów, podczas gdy większość Twoich klientów sortuje po ocenach i czyta te oceny?

Jeżeli Twoje założenia na temat produktów nie pasują do tego, czego dowiedziałeś się o preferencjach swoich klientów i ich zwyczajach, być może jest to czas na zmianę strategii.

Wykorzystaj swoje dane oraz techniki *data science* do ich analizy, by uzyskać wiedzę, która pomoże Ci określić oraz udoskonalić strategię. Selektywnie wybieraj obszary, które mógłbyś udoskonalić dzięki technologii *big data* oraz *data science*, a następnie dobieraj odpowiednie narzędzia, ludzi oraz procesy.

W następnym rozdziale opowiem, jak określać i priorytetyzować działania w obszarze danych.

Podsumowanie

- Źródła danych *big data* mogą pomóc w uzyskaniu informacji potrzebnych do tworzenia strategii, dostarczając nowej wiedzy na temat klientów, konkurencji, środowiska biznesowego oraz produktów.
- Istnieje wiele nowych, nietradycyjnych źródeł danych. Stwórz listę tego, co jest dostępne i co może być najbardziej użyteczne.
- Zazwyczaj napotkasz problemy, próbując połączyć akcje podejmowane przez klientów w różnych punktach kontaktu.
- Twoja strona oraz inne platformy cyfrowe mogą dostarczyć szczegółowych informacji o intencjach, preferencjach oraz zwyczajach klientów, sygnalizując potrzebę zmiany strategii.
- Stosowanie testów A/B w kombinacji z systemem *big data* umożliwia przeprowadzanie znacznie dokładniejszych analiz.

Zastanów się

- Utwórz listę punktów kontaktu z klientem. Dla każdego punktu zaznacz, czy zapisujesz jakiegokolwiek dane, a jeżeli tak, to czy tylko niektóre, czy też wszystkie dostępne. Określ każdy z tych punktów w skali od 1 do 10, biorąc pod uwagę (a) jaką wartość mają te dane, (b) jak trudno byłoby zapisywać i analizować dodatkowe dane z tego punktu. Pomnóż obydwie liczby przez siebie. Punkty kontaktu z najwyższymi wynikami to najbardziej obiecujące źródła danych o klientach.
- Jakie źródła danych potrzebowałbyś dołączyć, by uzyskać pełny obraz interakcji klientów? Czy coś powstrzymuje Cię przed użyciem tych danych?
- W jaki sposób różnią się zachowania i preferencje Twoich klientów i które z nich mogą wpływać na sposób, w jaki się do nich odnosisz, oraz produkty, które im oferujesz?
- Jaka była najbardziej udana zmiana w produkcie, której dokonałeś po przetestowaniu różnych możliwości?

- Jakie zewnętrzne dane są istotne z punktu widzenia Twojej firmy?
Dane o gospodarce, pogodzie, dniach wolnych itp.?
- Jakie źródła danych mogłyby dać Ci lepszy wgląd w Twoją konkurencję?
Pomyśl o prywatnych i publicznych źródłach informacji, jak również o wykresach i sygnałach udostępnianych przez firmy internetowe, takie jak Google lub LinkedIn.

Skorowidz

A

- agile, *Patrz:* metodyka zwinna
- AI, 20, 37, 121
 - rozwój, 39
 - system wartości, 44
 - wąska, 43, 44
 - zastosowania, 58
 - medycyna, 44, 57, 141, 142
- algorytm, 39, 140, 177
 - CUIImage, 42
 - Flink, 57
 - grafowy, 121
 - prototypowanie, 177
 - skalowanie, 122
- Allchin Jim, 89
- AlphaGo, 37, 38, 49, 146, 147
- Amazon, 37
- Amazon Web Services, *Patrz:* AWS
- analityk biznesowy, *Patrz:* zespół specjalista od biznesu
- analityka
 - diagnostyczna, 135, 137, 139
 - model poziomów, *Patrz:* model poziomów analityki
 - narzędzia, *Patrz:* narzędzia
 - opisowa, 135, 136, 137
 - predykcyjna, 135, 139, 140
 - preskryptywna, 135, 140
 - sieciowa, 67
- analiza
 - dostarczanie wyników użytkownikowi końcowemu, 163, 164
 - głównych składowych, 106
 - języka naturalnego, 109
 - samodzielna, 137
 - sentymentu, 65, 75
 - szlaków sygnałowych, 57
 - ścieżki zakupowej klienta, 30, 54, 56, 62, 67, 71, 104, 111, 179
 - pełne dane, 55
 - segmentacja, 104, 105, 106
 - towar, 106
 - wielkość danych, 23, 56
 - zbieranie danych, 98
 - wykres, *Patrz:* wykres
- Analytics Ascendancy Model, *Patrz:* model poziomów analityki
- Andreasen Alan, 52
- Apache Beam, 85
- Apache Software Foundation, 88
- applause rate, *Patrz:* współczynnik uznania
- Apple, 37
- architektura
 - lambda, 85
 - sieci neuronowej, 143, 146
 - SSN, 40, 41
- artificial intelligence, *Patrz:* AI
- artificial neural network, *Patrz:* SSN
- ASKAP, 25
- atak linkage attack, 197, 198
- Australian Square Kilometre Array Pathfinder, *Patrz:* ASKAP
- AWS, 32
- Azure, 32

B

- Banko Michele, 56
- batch job, *Patrz:* zadanie wsadowe
- baza danych, 124, 148
 - bezschematowa, 148
 - dokumentowa, 149
 - dostępność, 150

baza danych

- grafowe, 103, 150
- klucz-wartość, 149
- kolumnowa, 149
- masowo równoległa, *Patrz:* MPP
- NoSQL, 148
- relacyjna, 31, 148
- silnik wyszukiwania, 149
- spójność, 150
- tolerancja na partycjonowanie, 150
- w środowisku rozproszonym, 150
- wybór, 150, 162

Berkeley Vision Lab, 40

biblioteka

- do tworzenia wykresów, 152
- Keras, 146, 163
- TensorFlow, *Patrz:* TensorFlow

big data, 81, *Patrz też:* AI

- koszty, 167, 168
- narzędzia, 90
- zastosowania, 29, 49, 50, 51
- marketing, 67

BigQuery, 100

Bork Robert, 196

bozon Higgsa, 24, 29

Bratton Bill, 50

Brewer Eric, 150

Brill Eric, 56

C

Caffe, 40

Cassandra, 90

cecha, 42

- inżynieria, *Patrz:* dane inżynieria cech
- przykłady, 71

cena, 68

chmura, 160, *Patrz też:* przetwarzanie w chmurze

- prywatna, 32
- publiczna, 32, 161

CLV, 73, 74

CNN, 146

collaborative filtering, *Patrz:* filtrowanie treści

uwspólnione

content-based filtering, *Patrz:* filtrowanie oparte

na zawartości

convolutional neural network, *Patrz:* CNN

Critical Path Software, 185

CRO, 70

cross-sell, 74

cross-validation, *Patrz:* model sprawdzian krzyżowycustomer lifetime value, *Patrz:* CLV

cykl rozwoju nowych technologii, 29

D

dane

- baza, *Patrz:* baza danych
- cecha, *Patrz:* cecha
- dostęp, 199
- genomiczne, 57
- graficzne, 146
- gromadzenie, 53, 54, 98, 101, 120
- etyka, 99, 107
- konkurencja, 107
- medycyna, 57
- przepisy, 99, 107, 145, 193, 194, 195, 196, 197, 198, 199
- historyczne, 66, 110
- hurtownia, *Patrz:* hurtownia danych
- inżynieria cech, 42, 66, 137
- jezioro, *Patrz:* jezioro danych
- nieustrukturyzowane, 42
- ochrona, 194, 195
- oczyszczanie, 162
- osobowe, 100, 193
- prawo do bycia zapomnianym, 199
- quasi-identyfikator, 194
- prognozy kluczowych metryk, 110
- prywatność, 194, 196
- różnicowa, 198
- przechowywanie, 25, 27, 32, *Patrz też:* pamięć
- rozproszone, 83, 84
- przesyłanie, 162
- przetwarzanie, *Patrz:* przetwarzanie
- różnorodność, 82
- sekwencyjne, 146, 180
- strumieniowane, 85, 86
- surowe, 42, 101
- treningowe, 40, 41, 49, 71, 144, 146
- niewystarczające, 43, 44
- typ, 53, 54
- wideo, 146
- wizualizacja, *Patrz:* wizualizacja
- wolumen, 81
- wrażliwe, 194
- zarządzanie raportowaniem, 200

data mining, 121

data science, 123

data-driven, 50, 51, 53, 64, 72, 117, 195, 207

Deep Blue, 37, 38
 DeepMind, 49, 185
 Deming William Edwards, 50
 drzewo decyzyjne, 147

E

eBay, 30, 185
 Elasticsearch, 149

F

Facebook profil na podstawie polubień, 198
 fast data, 85
 Few Stephen, 180
 filtrowanie

- oparte na zawartości, 65
- treści uwspólnione, 65

 firma

- finansowa, 165
- handlowa, 69, 131
- kontrolowanie problemów, 136
- lornicza, 72
- strategia, *Patrz:* strategia
- telekomunikacyjna, 107
- turystyczna, 69, 109
- ubezpieczeniowa, 69, 143
- wgląd w dane, 136
- wymagania operacyjne, 136

 format

- JSON, 149
- XML, 149

 framework

- Hadoop, *Patrz:* Hadoop
- Spark, *Patrz:* Spark

 Fundacja Apache, 31

G

General Electric, 153, 185
 genom, 57
 Geometric Intelligence, 43
 goodness-of-fit, *Patrz:* model miara dopasowania
 Google, 20, 27, 29, 40

- misja, 30

 Google Analytics, 100
 Google Cloud, 32
 Google Maps, 109
 GPU, 66, 144

gra

go, 37, 147
 szachy, 37
 graf, 103, 150
 graphical processing unit, *Patrz:* GPU

H

Hadoop, 20, 27, 31, 84, 88, 90

- jako relacyjna baza danych, 31

 HDFS, *Patrz:* system HDFS
 Hive, 31
 hurtownia danych, 67, 124, 137, 162, 176

I

IaaS, 32, 91, 160
 ILSVRC, 42
 ImageNet, 42
 Immelt Jeff, 153
 Infrastructure as a Service, *Patrz:* IaaS
 infrastruktura

- jako usługa, *Patrz:* IaaS

 Instacart, 184
 inteligencja sztuczna, *Patrz:* AI
 internet rzeczy, *Patrz:* IoT
 Internet of Things, *Patrz:* IoT
 IoT, 24, 72, 86

J

Jaklevic Mary Chris, 206
 Jeopardy, 49, 205
 jezioro danych, 50
 język

- naturalny, 127
- programowania, 122, 124, 150
 - C++, 151, 177
 - Java, 151, 177
 - Python, 141, 144, 151, 177
 - R, 141, 144, 151, 177
 - Scala, 177

K

Kafka, 31
 kanał społecznościowy, 68
 Kasparow Garri, 37
 klasteryzacja, 106

klient, 97, 120
 adres IP, 99
 aktywność w sieci, 74
 deduplikacja, 102, 150
 doświadczenie, 69
 zakupowe, 55
 identyfikacja, 102, 103
 interakcja, 97, 98, 118
 łączenie danych, 102
 koszyk, *Patrz:* koszyk
 late adopter, 104
 lojalność, 69
 migracja, 69, 71, 74
 ograniczanie, 71, 72, 73, 106
 prognozowanie, 142
 segmentacja, 74, 104, 105, 106, 139, 147
 ilościowa, 106
 jakościowa, 106
 ścieżka zakupowa, *Patrz:* analiza ścieżki
 zakupowej klienta
 wartość, 74
 długookresowa, *Patrz:* CLV
 wizyta na platformie cyfrowej, 97, 98, 110
 wsparcie, 97, 101, 102, 118
 wykrywanie emocji, 107
 zapis ruchu, 98, 101, 102
 konkurencja, 107, 120
 udział w rynku, 108
 kontrola wersji, 124
 konwersja
 lejek, *Patrz:* lejek konwersji
 optymalizacja, *Patrz:* CRO
 prawdopodobieństwo, 70
 korelacja, 69
 koszyk, 69, 98, 127, 140
 KPI, *Patrz:* wskaźnik efektywności kluczowy
 krzywa podaży i popytu, 69
 kultura upraszczania, 153

L

Laney Doug, 81
Large Hadron Collider, *Patrz:* LHC
 Large Scale Visual Recognition Challenge,
 Patrz: ILSVRC
 latencja, 86
 lead scoring, 74, 142
 HR, 75
 lejek konwersji, 99, 180
 LHC, 24

licencja, 87, 88
 LIME, 143
 Local Interpretable Model-Agnostic
 Explanations, *Patrz:* LIME
 LoRaWAN, 87

M

machine learning, *Patrz:* uczenie maszynowe
 MacLaurin Ian, 51
 makrokonwersja, 99
 MapReduce, 27, 85
 Marcus Gary, 43
 massively parallel processing, *Patrz:* MPP
 maszyna wektorów nośnych, 147
 McCarthy John, 39
 media społecznościowe, 68, 69, 74, 97, 105
 filtrowanie treści, 38
 pobieranie danych, 162
 profil na podstawie polubień, 198
 metoda
 dynamicznej wyceny, 69
 fast data, 71
 gradientu prostego, 177
 grupowania, 42
 quasi-Newtona, 177
 sendHitTask, 100
 metodyka
 kaskadowa, 152
 zwinna, 152, 153, 211
 mikrokonwersja, 55, 99, 110
 ze ścieżki zakupowej, 69
 minimum viable product, *Patrz:* MVP
 Minsky Marvin, 39
 moc obliczeniowa
 koszt, 28
 skalowanie, 65
 model
 3V, 81, 83
 analityczny, 45, 121, 140
 projektowanie, 141
 czarnoskrzynkowy, 141, 142, 143, 145
 przejrzystość, 143
 dopasowywanie do danych, 143, 144
 kalibrowanie, 143, 144
 miara dopasowania, 144
 poziomów analityki, 135, 140
 przejrzystość, 141
 przetrenowanie, 144
 RFM, 139

sprawdzian krzyżowy, 144
 struktura, 143, 144
 Treacy'ego i Wiersemy, 118
 trenowanie, 146
 wdrażanie, 144
 MongoDB, 90, 149
 MPP, 30, 83
 MVP, 129, 154, 211
 iteracje, 130

N

nadużyć wykrywanie, 71
 narzędzia, 122, 137, 141, 147, 152
 do prototypowania algorytmów, 177
 do wizualizacji, 152
 ETL, 162, 176
 Neo4j, 90
 Netflix, 197
 Netscape Communications Corporation, 88
 neuron, 40
 nowotwór, 57

O

oprogramowanie
 jako usługa, *Patrz:* SaaS
 open source, 31, 66, 87, 163, 168
 historia, 87, 88
 korzyści, 89
 licencja, *Patrz:* licencja
 w big data, 90
 płatne, 168
 zakup, 163
 optymalizacja
 współczynnika konwersji, *Patrz:* CRO
 z ograniczeniami, 121
 Otto Group, 130, 131
 outsourcing, 186
 overfitting, *Patrz:* model przetrenowanie

P

PaaS, 161
 pamięć
 koszt, 26, 27
 o dostępie swobodnym, *Patrz:* pamięć RAM
 RAM, 25, 27
 stała, 25
 ulotna, 25

persona, 106, 139
 personally identifiable information, *Patrz:* dane osobowe
 PII, *Patrz:* dane osobowe
 piksel niewidzialny, 67
 Platform as a Service, *Patrz:* PaaS
 platforma jako usługa, 161
 plik cookie, 99
 Predix, 185
 procesor
 graficzny, *Patrz:* GPU
 tensorowy, 144
 produkt, 110, 120
 dopasowywanie w czasie rzeczywistym, 70
 historia, 120
 konkurencyjność, 118
 minimalnie wykonany, *Patrz:* MVP
 testowanie, 110, 111
 programowanie
 całkowitoliczbowe, 147, 177
 liniowe, 177
 obiektywne, 177
 przetwarzanie
 krawędziowe, *Patrz:* przetwarzanie mgłowe
 mgłowe, 86
 prędkość, 82
 rozproszone, 84
 równoległe masowe, *Patrz:* MPP
 strumieniowe, 69
 w chmurze, 32, 90, 160
 bezpieczeństwo, 161, 162
 zalety, 91, 161
 w czasie rzeczywistym, 85, 100, 127
 publikowanie, 22
 pulpit, 139

R

rak, *Patrz:* nowotwór
 RankBrain, 41
 RDMS, 148
 Realeyes, 107
 recurrent neural network, *Patrz:* RNN
 regresja logistyczna, 147
 reklama kontekstowa, 70
 relational database management system,
 Patrz: RDMS
 RNN, 146
 równanie finansowe, 147

S

SaaS, 32, 91
 bezpieczeństwo, 161
 Safe Harbour, 196
 Salesforce, 32
 SAS Enterprise Miner, 151, 152, 177
 Sedol Lee, 37
 Selenium, 108
 serwer HTTP Apache, 87, 88
 serwis REST, 151
 sieć neuronowa, 142, 146
 architektura, 143, 146
 konwolucyjna, *Patrz:* CNN
 rekurencyjna, *Patrz:* RNN
 sztuczna, *Patrz:* SSN
 tworzenie, 146
 silnik wyszukiwania, 149
 Siri, 37
 SKA, 25, 82
 słowo kluczowe, 67
 Snowden Edward, 162, 196
 Software as a Service, *Patrz:* SaaS
 Solr, 50, 149
 spam, 38, 41
 Spark, 20, 27, 90
 Splunk, 149
 sprzedaż dodatkowa, 74
 sprzęt, 159
 lokalizacja, 160
 outsourcing, 161
 wybór, 160
 Square Kilometre Array, *Patrz:* SKA
 SSN, 145
 architektura, 40, 41
 historia, 40
 trenowanie, 40, 41
 wielowarstwowa, 41
 zastosowania, 40, 41
 Stallman Richard, 87
 startup, 185
 statystyka, 177
 stos technologiczny, 159, 199
 strategia, 97, 104, 109, 110, 118
 cel biznesowy, 125, 127
 dokumentacja, 127, 128
 doskonałość operacyjna, 118
 konkurencyjność produktu, 118
 przewaga, 118
 ramy czasowe, 126

relacje z klientami, 118
 wdrażanie
 spotkanie inauguracyjne, 125, 127
 zespół, *Patrz:* zespół
 zmiana, 111
 zrywania nisko wiszących owoców, 126
 streaming, 22
 symulacja Monte Carlo, 69, 147
 system
 aukcji, 67
 CompStat, 50
 ekspercki, 39
 HDFS, 84, 85, 100
 kontroli wersji, 89
 operacyjny Linux, *Patrz:* Linux
 rekomendacji, 63, 65, 69
 zarządzania relacyjny, *Patrz:* RDMS
 szereg czasowy, 71
 sztuczna inteligencja, *Patrz:* AI

Ś

śląd cyfrowy, 67

T

Target, 52, 56, 99, 197
 TensorFlow, 40, 146, 163
 teoria CAP, 150
 Teradata, 30, 82
 Tesco, 51, 52, 54
 test
 A/B, 61, 69, 70, 110, 111, 177
 usprawnienia, 62, 63
 istotności różnic, 111
 Theano, 40
 TPU, *Patrz:* procesor tensorowy

U

Uber, 43
 uczenie
 głębokie, 40, 41, 121, 142, 145, 147,
 Patrz też: SSN
 maszynowe, 20, 38, 39
 up-sell, *Patrz:* sprzedaż dodatkowa
 urządzenie
 generujące dane cyfrowe, 21, 22, 23, 24, 54, 98
 przenośne, 23
 utrzymanie według stanu, 72

V

VCS, *Patrz:* system kontroli wersji
 version control system, *Patrz:* system kontroli
 wersji
 Visual Networking Index, 24

W

warstwa, 41
 Washington Post, 64
 Watson, 205, 206
 Waze, 109
 web scraping, 108
 wideoanaliza, 107
 Wielki Zderzacz Hadronów, *Patrz:* LHC
 wizualizacja, 138, 152, *Patrz też:* wykres
 narzędzia, 152
 wskaźnik
 efektywności
 kluczowy, 111, 126, 207
 zmiana, 111, 136
 kluczowy optymalizacja, 135
 przetwarzania klienta, 71
 współczynnik
 konwersji optymalizacja, *Patrz:* CRO
 uznania, 110
 wycena dynamiczna, 69
 wykres, 137, 152, 180

Y

Yahoo, 29
 YouTube, 81

Z

zadanie wsadowe, 65
 zarządzanie łańcuchem dostaw, 73
 zasoby ludzkie, 75, *Patrz też:* zespół
 zdarzenie wysokopoziomowe, 98
 zespół, 118
 budowanie, 173
 lead data scientist, 181, 182
 outsourcing, 186, 187
 przywódca, 181, 182
 specjalista
 data scientist, 173, 174, 211
 inżynier danych, 176
 IT, 124, 125, 127, 144, 165
 nietechniczny, 120
 od algorytmów, 177, 178
 od analizy, 121, 123, 127, 129, 136,
 139, 184
 od biznesu, 119, 120, 126, 129, 136, 179
 od danych, 183
 od ochrony danych osobowych, 145, 165,
 199
 od raportowania, 180
 od strategii, 118
 platform engineer, 176
 sieciowy, 179
 techniczny, 123, 124, 141
 zewnętrzny, 187
 spotkanie inauguracyjne, 125
 wynik, 127
 w małej firmie, 188

NOTATKI

PROGRAM PARTNERSKI

— GRUPY HELION —



1. ZAREJESTRUJ SIĘ
2. PREZENTUJ KSIĄŻKI
3. ZBIERAJ PROWIZJĘ

Zmień swoją stronę WWW w działający bankomat!

Dowiedz się więcej i dołącz już dzisiaj!

<http://program-partnerski.helion.pl>

GRUPA
Helion 

BIG DATA: ŁATWIEJSZE, NIŻ MYŚLISZ, SKUTECZNIEJSZE, NIŻ MARZYSZ!

Koncepcja **big data** zmieniła zasady gry w biznesie. Wiele osób z kadry zarządczej nie rozumie specyfiki tego rodzaju danych: ogromnych, szybko narastających, często niepasujących do tradycyjnej struktury. Są one zasadniczo różne od konwencjonalnych danych, zarówno pod względem wielkości, jak i złożoności. Rzucają nowe wyzwania, stwarzają nowe możliwości, zacierają tradycyjne granice konkurencji i zmuszają do zmiany paradygmatów pozyskiwania wartości z danych. **Big data** i **data science** wraz z uczeniem maszynowym radykalnie zmieniają ekosystem biznesu. Aby przetrwać tę rewolucję, trzeba dostosować się do nowych warunków.

Ta książka jest przystępnym wprowadzeniem do koncepcji **big data** i **data science**. Pozwoli na uzyskanie wiedzy niezbędnej do oceny, czy korzyści z tych technologii są warte kosztów i wysiłku związanych z wdrożeniem w firmie. Poszczególne techniki zostały dokładnie i przejrzysto opisane. Przedstawiono zasady tworzenia odpowiednich strategii. Wyjaśniono, jakich zasobów i jakich ludzi potrzeba do przeprowadzenia transformacji w kierunku zbierania, analizy i wykorzystywania danych, a także omówiono związane z tym ryzyko. Ważnym elementem książki są praktyczne wskazówki i podpowiedzi.

W tej książce:

- podstawy **big data**, **data science** i sztucznej inteligencji
- praktyczne zastosowanie **big data** w technikach analitycznych
- przegląd podstawowych rodzajów analityki i dobór technologii
- przygotowanie firmy do wdrożenia projektów **big data** i **data science**
- wymagania prawne i ochrona danych a korzystanie z narzędzi **big data**



DAVID STEPHENSON jest naukowcem i konsultantem. Specjalizuje się w zagadnieniach data science oraz analityki big data. Od niemal dwudziestu lat tworzy narzędzia i rozwiązania analityczne wspomagające decyzje biznesowe o wartości milionów dolarów. Jest znanym w Europie i w USA ekspertem, z którego wiedzy czerpią najwyższej klasy firmy doradcze. Prowadził programy analityczne dla firm z sześciu kontynentów. Pochodzi z USA, obecnie mieszka w Amsterdamie.

Helion

helion.pl

HELION SA
ul. Kościuszki 1c
44-100 Gliwice
tel.: 32 230 98 63
helion@helion.pl

INFORMATYKA W NAJLEPSZYM WYDANIU

Sprawdź nasze szkolenia

SZKOLENIA

AKADEMIA IT & BUSINESS

WWW.SZKOLENIA.HELION.PL

KOD KORZYŚCI
Sięgnij po więcej!

ISBN 978-83-283-5796-9

9 788328 357969

Cena: 49,00 zł