

O'REILLY®

Umiejętności analityczne w pracy z danymi i sztuczną inteligencją

Wykorzystywanie najnowszych
technologii w rozwijaniu
przedsiębiorstwa



Helion 

Daniel Vaughan

Tytuł oryginału: Analytical Skills for AI and Data Science: Building Skills for an AI-Driven Enterprise

Tłumaczenie: Maksymilian Gutowski

ISBN: 978-83-283-7346-4

© 2021 Helion SA

Authorized Polish translation of the English edition of Analytical Skills for AI and Data Science ISBN 9781492060949 © 2020 Daniel Vaughan

This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls all rights to publish and sell the same.

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/umanda>

Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Printed in Poland.

- [Kup książkę](#)
- [Poleć książkę](#)
- [Oceń książkę](#)

- [Księgarnia internetowa](#)
- [Lubię to! » Nasza społeczność](#)

Spis treści

Wprowadzenie	5
1. Myślenie analityczne i przedsiębiorstwa sterowane sztuczną inteligencją	11
Czym jest sztuczna inteligencja?	12
Dlaczego współczesna AI nie spełni pokładanych w niej nadziei?	13
Jak się tutaj znaleźliśmy?	13
Historia niespełnionych oczekiwań	17
Umiejętności analityczne w nowoczesnym, sterowanym sztuczną inteligencją przedsiębiorstwie	18
Główne wnioski	18
Dodatkowe materiały	19
2. Wprowadzenie do myślenia analitycznego	21
Pytania deskryptywne, predykcyjne i normatywne	21
Pytania biznesowe i KPI	25
Anatomia decyzji: prosty rozkład	27
Wprowadzenie do przyczynowości	30
Niepewność	37
Główne wnioski	41
Dodatkowe materiały	42
3. Zadawanie właściwych pytań biznesowych	43
Od celów do pytań biznesowych	43
Pytania deskryptywne, predykcyjne i normatywne	45
Zawsze zaczynaj od pytania biznesowego i działaj wstecz	45
Dalsza dekonstrukcja pytań biznesowych	46
Nauka zadawania pytań biznesowych: przykłady typowych przypadków użycia	49
Główne wnioski	60
Dodatkowe materiały	61

4. Działania, dźwignie i decyzje	63
Co jest przekładalne na działania?	63
Dźwignie fizyczne	64
Dźwignie ludzkie	65
Przypadki użycia — powtórka	76
Główne wnioski	79
Dodatkowe materiały	80
5. Od działań do konsekwencji: nauka upraszczania	81
Dlaczego musimy upraszczać?	82
Ćwiczenie zmysłu analitycznego: powitajmy Fermiego	83
Powtórka przykładów z rozdziału 3.	92
Główne wnioski	99
Dodatkowe materiały	100
6. Niepewność	101
Z czego bierze się niepewność?	102
Kwantyfikacja niepewności	102
Podejmowanie decyzji w warunkach pewności	107
Podejmowanie prostych decyzji w warunkach niepewności	109
Decyzje w warunkach niepewności	111
Normatywne i deskryptywne teorie podejmowania decyzji	115
Przykładowe paradoksy towarzyszące podejmowaniu decyzji w warunkach niepewności	116
Teoria w praktyce	120
Przypadki użycia — powtórka	125
Główne wnioski	133
Dodatkowe materiały	133
7. Optymalizacja	137
Czym jest optymalizacja?	137
Optymalizacja bez niepewności	143
Optymalizacja w warunkach niepewności	157
Główne wnioski	164
Dodatkowe materiały	164
8. Podsumowanie	167
Umiejętności analityczne	167
Sterowane AI przedsiębiorstwo przyszłości	173
Uwagi końcowe	176
Krótkie wprowadzenie do uczenia maszynowego	177

Zadawanie właściwych pytań biznesowych

W rozdziale 2. omówiłem wstępnie ogólne ramy działania, jakie opracujemy w kolejnych rozdziałach. Ponieważ naszym ostatecznym celem jest przełożenie problemów biznesowych na rozwiązania normatywne, powinniśmy zorientować się, jak zadawać *właściwe* pytania. Mam nadzieję, że nie dziwi Cię, że nauka zadawania pytań może okazać się równie wartościowa, jak wykorzystywanie innych omówionych tutaj technik.

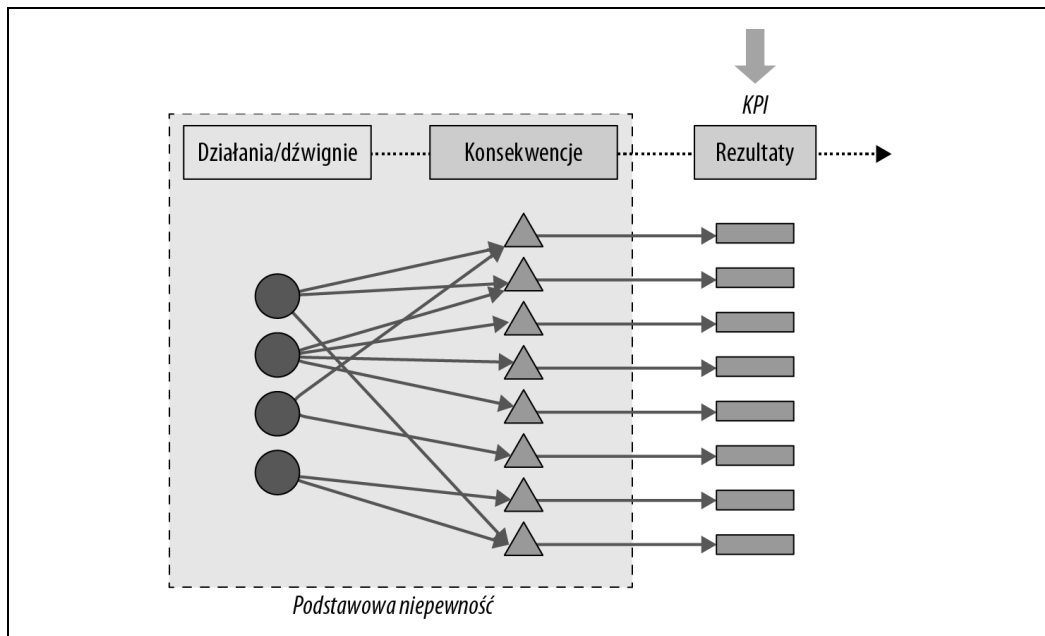
Zapoznałeś się także z bardzo prostą techniką, którą w toku własnych doświadczeń uznałem za dość pomocną w rozpracowywaniu tego, co rzeczywiście chcemy osiągnąć: *sekwencję* pytań „dlaczego?”. Na początek poddaje się pod wątpliwość to, co sądzisz, że chcesz osiągnąć, a następnie przechodzi się na wyższy poziom abstrakcji lub kończy się tę sekwencję, jeśli udało się wyłonić wiarygodny cel biznesowy. W naszych poszukiwaniach rozwiązań normatywnych absolutnie kluczowe jest upewnienie się, że zajmujemy się właściwymi celami. Jednym ze skutków ubocznych stosowania tej techniki, który okaże się całkiem przydatny w rozdziale 4., jest to, że w ten sposób zazwyczaj poszerza się zakres możliwych działań lub dźwigni. Tak zazwyczaj jest, kiedy zaczynamy od podważenia danego działania, a w wyniku tej procedury docieramy do wskaźników, na które rzeczywiście chcemy wpłynąć. Oczywiście wydaje się zatem zastanowienie się, czy istnieją inne działania, które można wykorzystać, aby ułatwić sobie osiągnięcie tego samego celu.

W tym rozdziale przyjrzymy się bliżej niektórym z najlepszych praktyk dotyczących zadawania właściwych pytań biznesowych; omówimy różnicę między pytaniami deskryptywnymi, predykcyjnymi a normatywnymi; na koniec przedstawimy kilka przykładów wziętych z częstych przypadków użycia. Przykłady te zaczerpnąłem z własnych doświadczeń, z rozmów ze studentami i kolegami po fachu, a kierowałem się tym, aby były łatwe do zaprezentowania i pomagały w objaśnieniu metod. Najpierw jednak powinniśmy lepiej zrozumieć, skąd biorą się pytania biznesowe (rysunek 3.1).

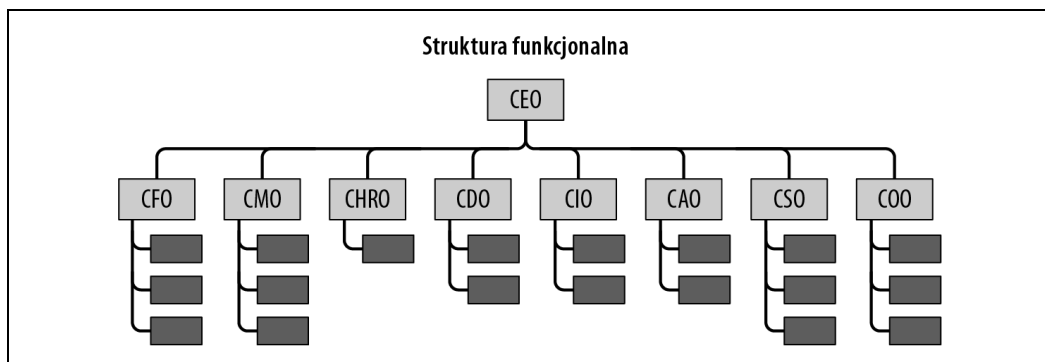
Od celów do pytań biznesowych

Większość firm zorganizowanych jest pionowo poprzez jasne oddzielenie obowiązków i celów w poszczególnych obszarach działania (rysunek 3.2). W ciągu ostatnich paru lat ruch **Agile** pomógł jednak wielu firmom przełamać zjawisko siloizacji i tworzyć zespoły interdyscyplinarne. Dzięki temu każdy zespół ma bardzo jasno określone cele biznesowe i wskaźniki, którymi ma się zajmować¹.

¹ O różnych strukturach organizacyjnych przeczytasz na <https://oreil.ly/T1MzW> i https://oreil.ly/EOa_D.



Rysunek 3.1. Zaczynj od działalności biznesowej



Rysunek 3.2. Przykład przedsiębiorstwa zorganizowanego w działy funkcjonalne — jeśli spojrzeć od lewej do prawej, akronimy oznaczają dyrektorów działów finansów, marketingu, kadr, danych, informacji, analiz, sprzedaży i działalności operacyjnej. Takich skrótowców może być znacznie więcej

To dla nas dobra wiadomość, ponieważ określone ogólnie cele biznesowe są zazwyczaj dobrze sformułowane i stosunkowo prosto można ocenić postępy w ich osiągnięciu na podstawie odpowiednich KPI. Naszym zadaniem jednak jest zadawanie istotnych pytań biznesowych, aby umożliwić sobie osiągnięcie tych celów. Z reguły na każdy cel biznesowy przypada wiele pytań, które można zadać, a z nimi z kolei wiążą się różne działania i dźwignie.

W jaki sposób formułuje się dobre pytania biznesowe? Ponieważ na nasze potrzeby musi istnieć możliwość *podjmowania działań* na podstawie takich pytań, przede wszystkim konieczne jest zrozumienie celów biznesowych, które chcemy osiągnąć, wskaźników używanych do oceny wyników, a także zorientowanie się w ewentualnych dźwigniach. Jeżeli nie można wskazać działań do

podjęcia, oznacza to, że albo pytanie nie przekłada się na takie możliwości, albo że nie przemyślałeś dostatecznie problemu. W przeciwnym wypadku jesteśmy na dobrej drodze. Musimy teraz odróżnić od siebie pytania deskryptywne, predykcyjne i normatywne.

Twarde i miękkie KPI

Często słyszy się o *twardych* i *miękkich* KPI, choć pojęcia te nie mają ustalonych definicji. Twarde wskaźniki uważa się za stosunkowo łatwe do obiektywnego zmierzenia, tak jak w przypadku KPI finansowych. Z drugiej strony, wskaźniki miękkie, takie jak świadomość marki, zadowolenie klienta lub jakość usług, trudniej zmierzyć w sposób obiektywny i dokładny.

To rozróżnienie nie jest oczywiste i zawsze będzie pole do dyskusji na ten temat, ale ogólnie panuje poczucie, że wskaźniki finansowe są silniej ugruntowane w danych i możliwe jest dokładniejsze ich zmierzenie. Być może pamiętasz z rozdziału 2., że wymierność jest jedną z właściwości dobrych KPI.

Pytania deskryptywne, predykcyjne i normatywne

W artykule *What is the question?* (<https://oreil.ly/xcPM7>) Jeff Leek i Roger Peng opisują sześć rodzajów pytań, na które można odpowiedzieć na podstawie danych: deskryptywne, eksploracyjne, dedukcyjne, predykcyjne, przyczynowe i mechanistyczne. Analiza danych na ogół odzwierciedla nasze procesy analityczne, więc odzwierciedlają one dość celnie ujętą tutaj trzejelementową klasyfikację z analizami deskryptywnymi, predykcyjnymi i normatywnymi.

W rozdziale 2. opisałem wszystkie trzy rodzaje analiz, więc tutaj przypomnę jedynie, że analiza *deskryptywna* odnosi się do przeszłości, *predykcyjna* do przyszłości, a *normatywna* pozwala wskazać najlepsze działania, jakie można podjąć dziś, aby wpłynąć na przyszłość.

Do napisania tej książki skłoniła mnie m.in. luźna obserwacja, że większość ludzi zadaje pytania deskryptywne, a ma trudności ze wskazywaniem, gdzie należy korzystać z analiz predykcyjnych i normatywnych. W dalszej części tego rozdziału przedstawię wystarczająco dużo przykładów, aby wyeliminować wszelką niepewność w odniesieniu do tych pojęć.

Zawsze zaczynaj od pytania biznesowego i działaj wstecz

W świecie pracy z danymi mawia się, że praktycy generują wartość dodaną poprzez wynajdowanie *praktycznych wniosków*. Choć nie jest to błędna ocena, istnieje ryzyko poświęcenia godzin, dni lub tygodni na poszukiwania spostrzeżeń wartych miliony.

Na pewnym etapie swojej kariery robiłem coś bardzo podobnego: przekonałem się, że stosunkowo łatwo przyglądać się ogonom różnych rozkładów — tych o niższym prawdopodobieństwie wystąpienia — i wyłaniać niedostrzeżone wcześniej szanse biznesowe z takich mikrosegmentów. Ponieważ większość modeli skupia się na przeciętnym kliencie (tym samym lekceważąc ogony), był to stosunkowo prosty sposób, aby pomóc mojemu pracodawcy zarobić co nieco. To jest niemalże definicja nagrody znajdującej się na wyciągnięcie ręki. Wiązały się z tym jednak dwa problemy: takie podejście nie było skalowalne, a było wykonywanym ręcznie, czasochłonnym procesem.

Lepszą praktyką na ogół jest obranie pytania biznesowego jako punktu wyjścia, aby następnie cofnąć się do samych danych. Ten proces prowadzi do szybszego uzyskiwania praktycznych wniosków, ponieważ wskazujesz je już na samym początku². Opisany w tej książce proces pomoże Ci uściślić i usprawnić analizę, więc być może uda Ci się uniknąć marnowania swojego czasu i cennego czasu zespołu w poszukiwaniu obiecanych spostrzeżeń.

Dalsza dekonstrukcja pytań biznesowych

Sekwencja pytań „dlaczego?” pomaga nam przejść od konkretniejszych do ogólniejszych pytań w ramach poszukiwania tych wskaźników, na które rzeczywiście chcemy wywierać wpływ. Istnieje jednak ryzyko, że taki ostateczny wskaźnik okaże się zbyt ogólny, aby można było podejmować działania z myślą o nim (na najwyższym poziomie uogólnienia celem niemal zawsze jest coś w stylu „zwiększyć dochód”). Powinniśmy jednak pamiętać, że nasze własne cele biznesowe są naturalnym ograniczeniem, więc zazwyczaj istnieje jakiś punkt, w którym sekwencję pytań należy zakończyć. Istnieją ponadto techniki umożliwiające nam zrobienie czegoś zupełnie przeciwnego, czyli rozkładanie pytań tak, aby znaleźć odpowiedni poziom, na którym można jasno wskazać te pośrednie cele, względem których również możliwe jest podjęcie działań.

Rozważ na przykład zagadnienie wskazania najlepszych działań pozwalających na uzyskanie najwyższego możliwego wskaźnika konwersji kampanii marketingowej. Zauważ, że celowo ująłem już tę kwestię w formie problemu normatywnego: wskaźnik biznesowy jest dobrze określony (wskaźnik konwersji), a wskazanie odpowiednich działań (teoretycznie) pozwoli nam wybrać te, które najbardziej korzystnie wpłyną na naszą zdolność do osiągnięcia wyznaczonych celów.

Rozkład wskaźników konwersji

Każdą proporcję można rozłożyć, mnożąc i dzieląc ją przez różne wskaźniki. Tutaj zaczynamy od stosunku transakcji do leadów — wskaźnika konwersji — i najpierw mnożymy i dzielimy go przez liczbę klientów, do których udało się dotrzeć. Powtarzamy to z liczbą klientów, do których rzeczywiście zadzwoniliśmy. Na koniec przeformułujemy równanie tak, aby każdy jego składnik przedstawiał istotny wskaźnik.

$$\frac{\text{Transakcje}}{\text{Leady}} = \frac{\text{Transakcje}}{\text{Dotarcia}} \cdot \frac{\text{Dotarcia}}{\text{Kontakty}} \cdot \frac{\text{Kontakty}}{\text{Leady}}$$

$\frac{WK}{A} \cdot B \cdot C$

Wskaźniki konwersji mają to do siebie, że można je łatwo rozłożyć, co pozwala na uzyskanie pytań umożliwiających podejmowanie bardziej bezpośrednich działań. W tym przypadku wskaźnik konwersji (WK) jest iloczynem trzech różnych współczynników, z czego każdy ma własne dźwignie, niekiedy o różnej sile oddziaływania³. Zaczynając od współczynnika na prawym końcu (C), jeżeli ze stu leadów spróbowałeś skontaktować się telefonicznie tylko z piętnastoma, może to

² Cassie Kozyrkov, główna badaczka decyzji w Google, przedstawiła podobny pogląd w *Never Start with a Hypothesis* (<https://oreil.ly/iMeVy>) oraz *The First Thing Great Decision Makers Do* (<https://oreil.ly/uw7bO>).

³ Jeżeli nie zauważyłeś, mnożę i dzielę przez te same wskaźniki tak, aby zachowana była równość.

oznaczać, że Twój zespół telemarketerów odznacza się niską produktywnością, wobec czego lepiej porozmawiać z jego kierownikiem, aby rozeznać się w możliwościach działania lub chociaż zorientować się, co się dzieje.

Jeżeli z kolei zadzwoniłeś na wszystkie numery i byłeś w stanie dotrzeć jedynie do ułamka klientów (B), dobrym pomysłem może być zbadanie zmiennych, które pomogą Ci przewidzieć, o jakiej porze najlepiej się kontaktować z klientami: to zadanie można powierzyć firmowym analitykom danych.

Wreszcie, jeśli Twój zespół sprzedażowy jest w stanie dokonać konwersji zaledwie małego odsetka tych klientów, do których udało się dotrzeć (A), konieczne może być ulepszenie modeli predykcyjnych w celu generowania wyższej jakości leadów, dostosowanie modelu systemu rekomensat, albo być może dopasowanie produktu do rynku nie jest jeszcze dostateczne.

Zauważ, jak rozkład pozwala nam od razu odnaleźć pośrednie wskaźniki lub pytania wraz z odpowiadającymi im działaniami, które umożliwiają nam podniesienie wskaźnika konwersji. Tę sztuczkę można z łatwością odnieść do większości lejków konwersji. Przyjrzyjmy się przykładowi archetypicznej *platformy dwustronnej*.

Przykład platformy dwustronnej

Platformy, czy też rynki dwustronne, na ogół starają się dopasować do siebie użytkowników z obu stron. Facebook na przykład dopasowuje firmy, które są zainteresowane opublikowaniem reklam (w celu zwiększenia sprzedaży), do właściwych klientów (użytkowników sieci społecznościowej). Amazon kojarzy ze sobą dystrybutorów lub sprzedawców towarów z właściwymi nabywcami, Uber kierowców z pasażerami itd.

Wyobraź sobie, że zakładasz własną platformę randkową. W tym przypadku dwie strony to użytkownicy, którzy chcą znaleźć dopasowanego partnera. Większość takich aplikacji randkowych daje użytkownikom możliwość komunikowania się ze sobą. Dla uproszczenia przyjmijmy, że każdy użytkownik może napisać tylko jedną wiadomość; ten przypadek w bardziej uogólnionym wydaniu jedynie wydłużyłby rozkład.

Jeśli użytkownicy po wymianie wiadomości przypadną sobie nawzajem do gustu, mogą przenieść się gdzieś indziej (do kawiarni, baru, restauracji). Twój zespół analityków chce usprawnić algorytm dopasowujący aplikacji, którego efektywność mierzy się współczynnikiem dopasowań przechodzących w konwersję. Na potrzeby przykładu założmy, że użytkownicy zawsze przekazują aplikacji informacje zwrotne, dzięki czemu wiemy, czy obie strony się spotkały⁴.

Dysponujemy zbiorem danych o wszystkich użytkownikach, ich interakcjach (Wiadomość 1 i Wiadomość 2) oraz o wyniku końcowym (Spotkanie lub Brak spotkania). Wydajność dopasowania (WD) można rozłożyć następująco:

⁴ Niektóre aplikacje randkowe zachęcają użytkowników do przekazywania takich formalnych informacji zwrotnych, ale czasami możliwe jest zmierzenie wydajności dopasowania w mniej bezpośredni sposób.

Rozkład wydajności dopasowania aplikacji randkowej

Wyświetlamy w aplikacji randkowej zdjęcia użytkowników w nadziei, że zachęcą innych użytkowników do skontaktowania się. Każdy użytkownik może postanowić rozpocząć dyskusję i wysłać pierwszą wiadomość (Wiadomość 1), na którą drugi użytkownik może odpowiedzieć (Wiadomość 2). Następnie obydwoje mogą zdecydować, czy chcą spotkać się gdzieś indziej, czy zakończyć rozmowę.

$$\frac{\text{Spotkania}}{\text{Wyświetlenia}} = \frac{\text{Spotkania}}{\text{Wiadomość 2}} \cdot \frac{\text{Wiadomość 2}}{\text{Wiadomość 1}} \cdot \frac{\text{Wiadomość 1}}{\text{Wyświetlenia}}$$

$WD \qquad \qquad A \qquad \qquad B \qquad \qquad C$

W tym równaniu każda terma oznacza liczbę wystąpień danego zdarzenia. Na przykład *Spotkania* oznacza liczbę ludzi, którzy udali się na spotkanie, a *Wiadomość 1* i *Wiadomość 2* liczbę wysłanych pierwszych wiadomości i odpowiedzi. Każdy współczynnik powinien mieć wartość niższą od jednego, ponieważ wartość licznika określa podzbiór dla zdarzenia wskazanego w mianowniku. Tak jest zawsze przy rozkładzie lejków konwersji.

Zwróć uwagę na to, co mówi nam rozkład. Jeżeli chcemy dopasować użytkowników, muszą oni wymienić między sobą wiadomości, co można przedstawić w formie trzech współczynników: kiedy użytkownik trafia na kogoś wyświetlonego w aplikacji, może wysłać pierwszą wiadomość (Wiadomość 1). Ten współczynnik (C) wskazuje, czy algorytm jest wydajny z perspektywy pierwszego użytkownika: jeżeli aplikacja wyświetliła dziesięciu kandydatów i wszyscy okazali się dobrze dobrać, to użytkownik napisałby wiadomość do każdego z nich⁵. Użytkownik drugi może odpowiedzieć lub nie: jeśli postanowi tak zrobić, może to być wskazaniem, że algorytm również sprawdza się u niego (B). Wreszcie po dostarczeniu drugiej wiadomości, obydwoje postanawiają się spotkać lub rezygnują z tego (A).

Nie wszystko jednak zależy od dokładności algorytmu. Decyzja o rozpoczęciu konwersacji (Wiadomość 1) lub napisaniu odpowiedzi (Wiadomość 2) zależy również od uwagi każdego z użytkowników, choćby ze względu na opóźnienia w komunikacji: aplikacje randkowe są wulkanami aktywności, więc jeśli jeden użytkownik ociąga się z odpowiedzią, drugi może stracić zainteresowanie i wznowić poszukiwania potencjalnego partnera. Możemy wobec tego opracować metody na zachęcenie użytkowników do szybszego komunikowania się (e-maile, powiadomienia, wyskakujące okienka z przypomnieniem, że ktoś czeka na odpowiedź). Tak na przykład robi Bumble: pierwszy kontakt musi być nawiązany przez obie strony w ciągu 24 godzin, a w przeciwnym wypadku dopasowanie zostaje utracone.

Wniosek jest z tego taki, że niektóre pytania biznesowe można dodatkowo rozłożyć w celu wskazania właściwych działań, wobec czego konieczne może być rozważenie wywarcia wpływu na pośrednie KPI, aby ułatwić osiągnięcie celów. Omówimy teraz kilka często występujących rzeczywistych przypadków użycia.

⁵ Strategie są oczywiście różne: użytkownik może odezwać się jedynie do najlepszego kandydata, jaki się u niego pojawi i zdać się na to. Z tego względu być może warto ograniczyć analizę do rozkładu pomijającego ostatni współczynnik.

Nauka zadawania pytań biznesowych: przykłady typowych przypadków użycia

Omówimy teraz szereg przykładów, zaczynając od sposobu, w jaki w moim odczuciu standardowo zadaje się pytania biznesowe, aby następnie przedstawić ich deskryptywne, predykcyjne i normatywne odpowiedniki. Przypomnij sobie, że dobre pytanie normatywne zawsze powinno wskazywać sposoby na pociągnięcie za parę dźwigni w celu uzyskania możliwie najlepszego rezultatu odnośnie do wybranego celu biznesowego. Wybrane przykłady dodatkowo rozwinę w kolejnych rozdziałach, aby dojść do moim zdaniem zadowalających rozwiązań normatywnych; pokażę Ci też, jak dodatkowo takie rozwiązania ulepszać. Na razie powinieneś jedynie zapamiętać, że naszym celem w tym rozdziale jest nauczenie się przekładania pytań biznesowych na język analityczny.

Zmniejszanie odpływu klientów

Każde przedsiębiorstwo potrzebuje klientów, aby generować przychód. Zaczynamy od *pozyskania* klientów, a następnie dokładamy starań, aby jak najdłużej pozostawali lojalni. Tempo, w jakim klienci rozstają się z przedsiębiorstwem — wskaźnik odpływu — jest proporcją liczby klientów utraconych w określonych ramach czasowych do ogólnej liczebności bazy klientów w tym samym okresie. Ponieważ koszty pozyskania klientów bywają stosunkowo wysokie w porównaniu z kosztami retencji, większość przedsiębiorstw prowadzi specjalistyczne jednostki, których celem jest zabezpieczenie jak największej części bazy klientów.

Na rysunku 3.3 widnieje standardowy dla większości firm przypadek, który jest świetnym punktem wyjścia, aby rozpocząć korzystanie z omawianych technik.

Zmniejszanie odpływu klientów		Potencjalna wartość
Pytanie biznesowe Chcę obniżyć wskaźnik odpływu naszej firmy	Dlaczego? Na jakie KPI staram się wpłynąć?	
Pytanie deskryptywne Ilu klientów tracimy? Kim oni są?	Czy wszyscy klienci są tacy sami?	\$
Pytanie predykcyjne Czy mogę dowiedzieć się z wyprzedzeniem, którzy klienci odejda?	Jakie cechy sprawiają, że niektórzy klienci są bardziej skłonni do odchodzenia?	\$\$
Pytanie normatywne Jeśli mogę przewidzieć, którzy klienci odejda, to jaką powinienem przyjąć politykę utrzymania klientów i w jaki sposób ją wdrożyć?	Na jakie KPI wywrę wpływ? Jakie są krótko- i długoterminowe korzyści? Jaki jest koszt każdego działania na rzecz utrzymania klientów?	\$\$\$\$

Rysunek 3.3. Różne pytania związane z przypadkiem odpływu klientów

Definiowanie pytania biznesowego

Zacznijmy od pytania biznesowego, z którym mierzy się większość przedsiębiorstw: w jaki sposób możemy obniżyć wskaźnik odpływu? W tym przykładzie zaczynamy od działania zamiast od celu biznesowego, więc zadajemy sekwencję pytań „dlaczego?”, co prawdopodobnie doprowadzi nas do prostego wniosku, że klienci są naszym głównym źródłem stałego przychodu. Wydaje się to mało skomplikowane, ale ten prosty wniosek wskazuje nam główne KPI, które chcemy zmaksymalizować: wcale nie zależy nam na tym, aby wskaźnik odpływu był jak najniższy, lecz przychód jak najwyższy. Czy aby na pewno? Zawsze możesz przecież rzucić wszystko, żeby zatrzymać przy sobie klientów, podnosząc tym samym koszty. Wynika z tego, że przychód *nie jest* właściwym wskaźnikiem, lecz zysk rozumiany jako przychód po odjęciu kosztów retencji.

Pytania deskryptywne

Na poziomie czysto opisowym chcemy dokonać kilku rzeczy. Rzecz jasna zaczynamy od zadania sobie pytania, czy wskaźnik odpływu jest niestandardowo wysoki i jak zmienił się w przeszłości. Możemy zacząć od najbardziej zbiorczego poziomu, przyglądając się trendom czasowym oraz prawidłowościom sezonowym, co pozwoli nam się zorientować w stanie bieżącym. Dane pozwalają nam jednak uszczegółowić poszukiwania i zorientować się, *kim* są klienci, którzy już nas opuścili. Czy ich wartość była wysoka, czy niska? Jak długo korzystali z naszej oferty? Czy skontaktowali się z nami, aby dać wyraz swojemu niezadowoleniu? Czy są zlokalizowani w konkretnych miejscach? Jakie są niektóre ich cechy socjodemograficzne, takie jak wiek i płeć? Jak wyglądają ich wzorce konsumpcji lub użytkowania?

Uszczegółowiamy kwerendę na tyle, na ile pozwala nam czas i dane. Ogólnie pewnie jednak rozumiesz, co tu się dzieje: to jedynie migawka stanu faktycznego i mam nadzieję, że udało mi się przekonać Cię, że niezależnie od poziomu jej szczegółowości trudno wyciągnąć z niej większą wartość dodaną. Na tym etapie służyła ona głównie celom informacyjnym. Rzeczywistą wartością tej analizy opisowej jest to, że okaże się pomocna w naszych poszukiwaniach jak najlepszych decyzji, jakie możemy podjąć, aby osiągnąć nasze ostateczne cele.

Pytania predykcyjne

AI i uczenie maszynowe może pomóc nam znaleźć odpowiedź na pytanie predykcyjne: czy możemy przewidzieć, jacy klienci najczęściej odchodzą? Dzięki szczegółowości naszej analizy opisowej możemy teraz wykręć główne czynniki odpowiadające za nasz bieżący wskaźnik odpływu. Same dane jednak nie zawiodą nas daleko. Najlepsi analitycy to tacy, którzy *rozumieją* i potrafią *hipotetyzować*, dlaczego klienci odchodzą. W ten sposób mogą oni wskazywać konkretniejsze czynniki prognostyczne w procesie zwanym **inżynierią cech**, będącym najlepszym sposobem na uzyskanie naprawdę dużej mocy predykcyjnej. Wiedza o tym, co należy włączać do modeli, a co z nich wykluczać, jest Świętym Graalem konstrukcji sprawnych modeli w stopniu nawet większym niż dobór coraz potężniejszych algorytmów.

Jaką wartość można wygenerować na etapie predykcyjnym? Na rysunku 3.3 sugeruję, że większą niż na etapie deskryptywnym, ale — jak wspominałem w rozdziale 2. — może równie dobrze być zerowa, a nawet ujemna.

Pytania normatywne

Dotarliśmy w końcu do pytania normatywnego: za jakie dźwignie należy pociągnąć, aby zmaksymalizować zysk z kampanii retencyjnych? Czy myślimy o jednak o zysku *krótkoterminowym*? Czy klienci zorientują się w naszej strategii i zaczną ogrywać nasz system retencyjny, zwiększając tym samym koszty na dłuższą metę? Większość dojrzałych przedsiębiorstw posługuje się wskaźnikiem wartości życiowej klienta (CLV), którą wprowadziliśmy w rozdziale 2., i zgadzam się, że pozwala on na lepsze uchwycenie długofalowej wartości netto klientów. Ten wybór wskaźnika wiąże się jednak z własnym zbiorem utrudnień. Parafrazując Yogiego Berre: trudno cokolwiek prognozować, zwłaszcza jeśli chodzi o przyszłość, a jeszcze trudniej zrozumieć długofalowe efekty naszych działań.

Temat dźwigni poruszymy w rozdziale 4., ale w tym miejscu wystarczy powiedzieć, że w przypadku retencji klientów zawsze możemy przynajmniej zaoferować rabat. Jakie zatem rabaty są odpowiednie dla wszystkich klientów? CLV wskazuje górną granicę tego, ile można rozsądnie zaproponować, ale zawsze powinno nam zależeć na rozpoznaniu działania o najniższym koszcie, które jednocześnie zapewni retencję. Przybliży nas to do personalizacji dźwigni.

Idealnie powinniśmy wybrać *właściwe* działanie we *właściwym* czasie z myślą o *właściwym* kliencie. Dużo tutaj właściwych rzeczy: analiza normatywna jest trudna, więc na ogół staramy się uprościć problem. Korzyści płynące z upraszczania omówię w rozdziale 5. Tymczasem już przynajmniej ujęliśmy pytanie w formie potencjalnie umożliwiającej wygenerowanie możliwie największej osiągalnej wartości. Pamiętaj, że w tym rozdziale zajmujemy się nauką zadawania pytań. W rozdziale 7. przybliżę szczegóły związane z jednym możliwym rozwiązaniem tego przypadku użycia.

Sprzedaż krzyżowa: następna najlepsza oferta

Większość firm sprzedaje więcej niż jeden produkt lub oferuje więcej niż jedną usługę. Naturalną przewagą, jaką przedsiębiorstwo ma, kiedy oferuje produkty powstające w wyniku podobnych procesów produkcyjnych, ekonomiści nazywają **korzyściami zakresu**. Większości z nas logiczne wydaje się zatem rozglądanie się za różnymi sposobami pogłębienia naszej relacji z klientami za pomocą sprzedaży krzyżowej. W żargonie doradczym nazywa się ją obecnie **następną najlepszą ofertą** (ang. *next-best offer*), co sprowadza nas na grunt normatywności. Pomijając już terminologię, przekonasz się poniżej, że wcale nie jest oczywiste, pod kątem jakiego celu biznesowego prowadzi się tutaj optymalizację.

Definiowanie pytania biznesowego

Pytanie biznesowe jest tutaj proste (rysunek 3.4): *co powinienem teraz zaoferować moim klientom?* Jeżeli zastanawiasz się, dlaczego w ogóle miałbyś tutaj korzystać z sekwencji pytań „dlaczego?”, odpowiedź nie jest tak jasna, jak w przypadku odpływu klientów. Różnica polega na tym, że sprzedaż krzyżowa przynosi dwa rezultaty. Rezultatem bezpośrednim jest standardowo uzyskanie dopływu wyższego dochodu. Ciekawszy i bardziej złożony jest jednak rezultat pośredni: klienci, którzy kupują od nas więcej, są przeważnie bardziej lojalni, co wydłuża czas, przez jaki pozostają

Sprzedaż krzyżowa: następna najlepsza oferta		Potencjalna wartość
Pytanie biznesowe Co mam zaoferować każdemu z moich klientów?	Dlaczego? Na jakie KPI staram się wpłynąć?	
Pytanie deskryptywne Jakie są wzorce konsumpcji naszych klientów? Czy niektóre z tych wzorców wyłaniają się bardziej spontanicznie? U których klientów?	Czy wszyscy klienci są tacy sami? Czy wszystkie sekwencje produktów są równe?	\$
Pytanie predykcyjne Mając na uwadze dotychczasowe zakupy każdego klienta, co następnego prawdopodobnie kupią	Jakie cechy sprawiają, że niektórzy klienci są bardziej skłonni do kupowania poszczególnych produktów?	\$\$
Pytanie normatywne Jeśli chcę uzyskać największą wartość od każdego klienta, jaką najlepszą ofertę mogę za każdym razem złożyć?	Na jakie KPI wywrę wpływ? Czy powinienem uwzględnić pojedyncze produkty, czy całe ich sekwencje?	\$\$\$\$

Rysunek 3.4. Różne pytania związane z przypadkiem sprzedaży krzyżowej

przy nas. Z tego względu często warto rozważyć sprzedaż krzyżową z rabatem zwyczajnie dlatego, że długofalowy, ogólny zysk dzięki temu wzrasta, nawet jeśli sprzedaż pojedynczego produktu przynosi firmie straty. Znowu wygląda na to, że optymalizować należy CLV.

Pytania deskryptywne

W obszarze opisu standardowo należy zadawać pytania dotyczące wzorców konsumpcji różnych rodzajów klientów. Oczywistym rozwiązaniem jest sprawdzenie, czy pewne *sekwencje* produktów wyłaniają się bardziej spontanicznie w przypadku niektórych klientów. Rozważ przykład banku: większość klientów zaczyna korzystać z jego usług w dość młodym wieku, posługując się prostym produktem w rodzaju karty debetowej. Z czasem, kiedy ich dochody wzrastają z biegiem kariery, decydują się skorzystać z bardziej zaawansowanych ofert kredytowych i inwestycyjnych: najpierw można zaciągnąć kredyt mieszkaniowy, następnie wykupić ubezpieczenie na życie itd. W przypadku sekwencji ważna jest kolejność, w jakiej kupuje się każdy produkt, więc standardem jest rozglądanie się za takimi wzorcami w danych.

Pytania predykcyjne

Ponieważ każdy klient już coś zakupił, oczywiste wydaje się zadanie sobie pytania, czy można *przewidzieć* na podstawie dotychczasowych wzorców konsumpcji klientów, co będą kupować w przyszłości. Możemy następnie aktywnie zadziałać, zamiast czekać, aby przekonać się, czy dokonają zakupu u nas, czy u konkurencji. Czy powinniśmy jednak zaoferować produkt, którego sprzedaż przyniesie nam największy zysk, nawet jeśli jego zakup wydaje się bardzo mało prawdopodobny? Wracając do przykładu z bankiem, może zależeć Ci na udzielaniu kredytów hipotecznych (ze względu na to, że przynoszą znaczny zwrot), ale mało prawdopodobne jest, by studenci i młodzi pracownicy byli

zainteresowani taką ofertą. Prowadzi nas to do najciekawszego kompromisu towarzyszącego analizie sprzedaży krzyżowej: między prawdopodobieństwem zakupu a wzrostem wartości. Prowadzi nas to z kolei do pytania normatywnego.

Pytania normatywne

Ponieważ możemy każdemu klientowi zaoferować kilka artykułów, który z nich powinniśmy wybrać, aby uzyskać jak największą wartość? Jak wspomniałem, ponieważ mamy do czynienia z sekwencjami i upływem czasu, właściwym wskaźnikiem prawdopodobnie jest CLV. Kiedy jesteśmy rzeczywiście zorientowani na klienta, pytania normatywne powinny nam umożliwić wskazanie *właściwego* produktu dla *właściwego* klienta we *właściwej* cenie we *właściwym* czasie. Później przyjrzymy się, jak rozwiązać ten niezwykle złożony problem.

Optymalizacja CAPEX

Przemysł motoryzacyjny, paliwowy, telekomunikacyjny i lotniczy należą do branż kapitałochłonnych: do działania potrzebują dużej ilości zasobów na potrzeby budowy i utrzymywania fabryk, zakładów, wież, samolotów oraz wszelkich innych fizycznych aktywów, które z biegiem czasu tracą na wartości. Ten rodzaj inwestycji nazywa się **nakładami kapitałowymi** (ang. *capital expenses* — CAPEX), a jest on obecny we wszystkich branżach, nie tylko w tych wymienionych powyżej⁶.

Jedno z pytań, które dyrektorzy finansowi i inni kierownicy wysokiego szczebla wszelkich przedsiębiorstw oczywiście zadają, dotyczy tego, jak przydzielać CAPEX na przestrzeni różnych obszarów funkcjonalnych lub lokalizacji geograficznych (rysunek 3.5). Ponieważ CAPEX składa się na znaczną część przepływu środków pieniężnych przedsiębiorstwa, posługujemy się nawet specjalnymi KPI do pomiaru jego wpływu, takimi jak zwrot z inwestycji (ang. *return on investment* — ROI) lub zwrot z zaangażowanego kapitału (ang. *return on capital employed* — ROCE). Niemniej jednak zawsze powinniśmy podważać zasadność przydziału CAPEX i zastanawiać się nad tym, co konkretnie chcemy osiągnąć: na przykład skąd się bierze przychód figurujący w liczniku ROI⁷?

Na poziomie opisowym moglibyśmy zacząć od znalezienia korelacji pomiędzy różnymi przydziałami CAPEX i przychodem w różnych lokalizacjach. Wykorzystuje się przy tym zmienność poprzednich inwestycji w połączeniu z głównym wskaźnikiem pomiarowym, na jaki sądzimy, że wywarła ona wpływ. Inną możliwością jest wykorzystanie zróżnicowania czasowego i zagregowanych szeregów, aby znaleźć jakiegokolwiek wstępne wskazówki dotyczące zależności między przydziałem CAPEX a przychodem. Jeżeli niczego takiego nie znajdziemy, możemy zmienić swój cel i zastanowić się nad zróżnicowaniem przychodu, jako że CAPEX można także regulować poprzez obniżanie kosztów.

⁶ Porównaj to z wydatkami operacyjnymi lub OPEX, które obejmują m.in. wynagrodzenie wypłacane pracownikom.

⁷ Pamiętaj, że $ROI = \frac{\text{Przychód z inwestycji} - \text{Koszt inwestycji}}{\text{Koszt inwestycji}}$

Optymalizacja CAPEX		Potencjalna wartość
Pytanie biznesowe Jak mam przydzielić nakłady kapitałowe w tym kwartale?	Dlaczego? Na jakie KPI staram się wpłynąć?	
Pytanie deskryptywne Co stało się w przeszłości, kiedy w niektórych lokalizacjach geograficznych zainwestowałem więcej?	Czy te korelacje są pewne?	\$
Pytanie predykcyjne Czy jeśli zainwestuję w lokalizacji X, to wzrośnie sprzedaż lub poziom zadowolenia klientów? Jak bardzo?	Czy to są związki przyczynowe?	\$\$
Pytanie normatywne Jeżeli chcę zmaksymalizować zwrot z inwestycji, w jaki sposób powinienem przydzielić całość mojego CAPEX w różnych lokalizacjach?	Czy to rzeczywiście najlepszy rezultat, jaki mogę osiągnąć? Czy może to być zaledwie jedna z kilku lepszych alokacji? Co możemy zrobić lepiej?	\$\$\$\$

Rysunek 3.5. Różne pytania związane z przypadkiem optymalizacji CAPEX

Głównym problemem związanym z rozważaniem dowolnej inwestycji jest fakt, że *nie wiemy*, jakiego zwrotu się spodziewać, więc miło byłoby móc go dokładnie oszacować. Optymalna alokacja może być w takim przypadku kwestią rankingu: kiedy masz do zainwestowania złotówkę i znasz zwroty z wszystkich potencjalnych alokacji, wystarczy zainwestować ją w tą, która przyniesie największy zwrot. Czy można jednak ufać korelacjom pojawiającym się w analizie opisowej? Czy wykryty związek rzeczywiście jest przyczynowy? Jak zwykle trudność polega na uzyskaniu wiarygodnych predykcji przyczynowo-skutkowych, a właśnie tym zajmują się nasi badacze za pomocą narzędzi do uczenia maszynowego.

Zakładając jednak, że udało nam się uzyskać pewne i dokładne oszacowania, etap normatywny przechodzimy błyskawicznie: wystarczy przydzielić budżet różnym lokalizacjom w kolejności odpowiadającej wysokości zwrotów z nich. Później pokażę Ci przykład tego, jak można to zrobić, ale póki co interesuje nas głównie sposób ujęcia tej kwestii.

Lokalizacja sklepu

Jednym z moich ulubionych przypadków użycia jest określanie lokalizacji sklepu. Ponieważ właśnie omówiliśmy kwestię optymalizacji CAPEX, od razu widzimy, że jest to przykład tego problemu w praktyce. Dysponujemy budżetem na umocnienie naszej obecności handlowej i najchętniej otworzylibyśmy sklep tam, gdzie przyniesie nam to możliwie największy zwrot (rysunek 3.6). Oczywiście KPI wydaje się wartość bieżąca netto (ang. *net present value* — NPV) zysków sklepu. Czy rzeczywiście nim jest?

Aby uzmysłwić sobie złożoność tego problemu, rozważ otworenie sklepu w bliskim sąsiedztwie istniejącego (czy zastanawiałeś się kiedyś, dlaczego na danym osiedlu znajduje się tak wiele lokali

Lokalizacja sklepu		Potencjalna wartość
Pytanie biznesowe Gdzie powinienem otworzyć kolejne dziesięć sklepów? Czy powinienem otworzyć ich więcej przy obecnym budżecie?	Dlaczego? Na jakie KPI staram się wpłynąć?	
Pytanie deskryptywne Jakie są wzorce geograficznego zróżnicowania rentowności?	Czy te dane korelują z innymi cechami okolicy?	\$
Pytanie predykcyjne Gdybym otworzył sklep pod danym adresem, jakie przyniosłoby to korzyści?	Co z zyskami pobliskich sklepów?	\$\$
Pytanie normatywne Ile sklepów powinienem otworzyć i gdzie, mając na uwadze, że tegoroczny budżet wynosi x PLN i zależy mi na maksymalizacji zysków?	Czy to zyski przypadające na poszczególne sklepy, czy powinienem uwzględnić łączny zysk?	\$\$\$\$

Rysunek 3.6. Różne pytania związane z przypadkiem wyboru lokalizacji sklepu

jednej sieci handlowej?). Powinieneś uzyskać dodatkowy przychód i zysk, lecz jedynie kosztem zysków w pozostałych sklepach. Pewniejszym KPI jest zatem zagregowany zysk, przynajmniej na poziomie lokalnym (osiedla, ulicy lub nawet miasta).

Analizę opisową zacząłbym od rozejrzenia się za prawidłowościami w zróżnicowaniu zysków na przestrzeni różnych lokalizacji: czy występują jakieś w naszych własnych sklepach w okolicy? Co z konkurencją? Czy dysponujemy danymi umożliwiającymi określenie w przybliżeniu liczby potencjalnych klientów, którzy wchodzi do różnych sklepów? Co z przeciętnym dochodem w tej okolicy? Czy to przestrzeń mieszkaniowa? Możemy zadać wiele pytań, aby wyłonić prawidłowości tłumaczące zróżnicowanie zysków.

Podobnie jak w przypadku alokacji nakładów kapitałowych możliwość idealnego oszacowania NPV łącznych zysków pozwala od razu rozwiązać problem i przeznaczyć cały budżet na otwarcie sklepów według ich planowanego poziomu KPI. Oczywiście zakładam, że dysponujesz skończonym budżetem i że nie inwestujesz w otwieranie sklepów, z których zwrot będzie ujemny.

Kogo zatrudnić?

Dość jasne jest, że to pracownicy decydują o tym, jak dobrze przedsiębiorstwo sobie radzi. Jedną z najważniejszych decyzji, jakie stale podejmujemy, dotyczy zatem wyboru tego, kogo należy zatrudnić. Jednostki zajmujące się zasobami ludzkimi wkładają znaczny wysiłek w opracowywanie solidnych i pewnych procesów rekrutacyjnych (rysunek 3.7). Przy zatrudnianiu głównym problemem jest to, że niektóre z KPI może być trudno zmierzyć. Rozważmy na przykład produktywność. Jeżeli jesteś sprzedawcą, to możemy z łatwością zmierzyć Twoją produktywność, kierując się liczbą transakcji we wskazanym okresie. Dokonanie jej pomiaru czy nawet określenie wpływu pracownika na przychody firmy jest jednak w przypadku wielu innych stanowisk pracy znacznie trudniejsze.

Kogo zatrudnić?		Potencjalna wartość
Pytanie biznesowe Muszę obsadzić wolne stanowisko; kogo powinienem zatrudnić?	Dlaczego? Na jakie KPI staram się wpłynąć?	
Pytanie deskryptywne Jakie były wyniki poprzednich zatrudnionych pracowników w odniesieniu do ich cech?	Czy wyniki testów psychometrycznych są skorelowane z sukcesem? Co z innymi umiejętnościami twardymi i miękkimi?	\$
Pytanie predykcyjne Czy mogę przewidzieć sprawność każdego z kandydatów na podstawie CV, przebiegu rozmów kwalifikacyjnych i innych testów?	Czy dane są w jakiś sposób tendencyjne?	\$\$
Pytanie normatywne Jeżeli chcę uzyskać możliwie najlepszy rezultat (staż/produktywność) i wyniki budżetowe, to kogo powinienem zatrudnić?	Co konkretnie chcemy zoptymalizować? Czy możemy zmierzyć produktywność pracowników?	\$\$\$\$

Rysunek 3.7. Różne pytania związane z przypadkiem zatrudniania pracowników

Przyjmijmy, że możemy niezawodnie mierzyć wydajność tak jak w przypadku pracowników działu sprzedaży. Czy to jedyne ważne KPI? Co ze średnim stażem? Raczej nie chcielibyśmy zatrudnić gwiazdora, który osiąga świetne wyniki, ale potrafi zmienić pracę po miesiącu, skoro jego produktywność może nie pokryć kosztów zatrudnienia i przeszkolenia go. W warunkach idealnych chcielibyśmy zmierzyć coś w rodzaju CLV, więc posłużmy się analogicznym pojęciem życiowej wartości pracownika: wartości bieżącej netto indywidualnego wkładu w osiągnięty zysk. W ten sposób możemy uwzględnić szacowany czas pracy i wpływ na finanse.

Wyobraźmy sobie, że dysponujemy zbiorem danych o wszystkich naszych sprzedawcach z ostatnich 24 miesięcy. Podobnie jak w przypadku wskaźnika odpływu klientów, zbiór musi obejmować zarówno aktualnych pracowników o różnym stażu, jak i tych, którzy już od nas odeszli. To zróżnicowanie pozwoli nam na wykrycie prawidłowości. Idealnie byłoby też uwzględnić niektóre z ich twardych wskaźników wydajności (miesięczna sprzedaż), ale również bardziej miękkie dane, takie jak wyniki oceny 360 stopni. Wreszcie świetnie byłoby sięgnąć po dane, które uzyskujemy przy zatrudnianiu: CV, wykształcenie, doświadczenie zawodowe, wyniki badań psychometrycznych, płeć, wiek itd. Możemy następnie rozejrzeć się za korelacjami w danych, aby zorientować się, jakie zmienne mogą pomóc w prognozowaniu wydajności.

Pytanie predykcyjne jest dość proste: czy na podstawie informacji uzyskanych z wyprzedzeniem — wszystkich, które zebraliśmy przed podjęciem decyzji o zatrudnieniu — możemy przewidzieć wydajność kandydatów? Gdybyśmy mogli, problem byłby przynajmniej rozwiązywalny, choć podobnie jak wcześniej musielibyśmy się liczyć z pewnymi komplikacjami. Najważniejsza decyzja, z jaką mamy tutaj do czynienia, odnosi się do ogółu problemów związanych z poszukiwaniem (tak jak w przypadku rozglądania się za partnerem): czy powinniśmy rozglądać się dalej za nowymi kandydatami? Czy trafiliśmy na najlepszy wybór? Jeśli będziemy prowadzić poszukiwania przez kolejny miesiąc, to czy trafimy na kogoś lepszego? O obecnym w wielu problemach dotyczących poszukiwań kompromisie między *eksploracją* a *eksploatacją* jeszcze powiemy sobie bliżej, ale na razie zwróć jedynie uwagę

na to, że możemy albo zatrudnić nowego pracownika i wyciągnąć z niego jak najwięcej — to aspekt eksploatacji, chociaż oczywiście nie należy tego utożsamiać z wyczyszczeniem — albo dalej rozglądać się na rynku pracy za lepszymi kandydatami.

Na koniec muszę ostrzec, na wypadek gdybyś posiłkował się AI przy rozwiązywaniu tego typu problemów, że algorytmy predykcyjne są silnie wyczułone na tendencyjność danych, a badacze powinni dołożyć wszelkich starań, aby ją wykryć i skorygować. Wyobraź sobie, że według Twojego zbioru danych pracownice działu sprzedaży są bardzo produktywne, ale rezygnują z pracy po miesiącu. Model predykcyjny może sugerować, że życiowa wartość pracownika w przypadku kobiet jest wyraźnie niższa, co może z kolei skłonić Cię do zatrudniania głównie mężczyzn. Dlaczego jednak kobiety odchodzą tak szybko? Może w rzeczywistości kierownik działu jest pełnym mizoginem? W takim przypadku to właśnie tę osobę należy zwolnić, aby zatrudnić następnie więcej kobiet. Morał płynie z tego taki, że bez rzetelnej weryfikacji tendencyjności danych modele predykcyjne są bardzo ułomne. Jak to się mawia w świecie pracy z danymi: „śmieci na wejściu — śmieci na wyjściu”.

Wskaźnik zaległości płatniczych

Wiele firm oferuje klientom różne możliwości finansowania ich zakupów, np. w formie sprzedaży ratalnej. Lepiej powierzyć finansowanie zakupów na kredyt specjalistycznym przedsiębiorstwom (czyli bankom), ale często musimy zajmować się tym sami. Pytanie biznesowe dotyczy tego, jak zapewnić finansowanie bez podnoszenia wskaźnika zaległości płatniczych (rysunek 3.8).

Pytania deskryptywne są podobne do tych z poprzednich przypadków, które omówiliśmy, ale chciałbym podkreślić dwie sprawy: jeśli poprawnie zdefiniowałeś pytanie biznesowe i ująłeś je jako normatywne, powinieneś rozglądać się za takimi prawidłowościami w danych, które determinują cel, zamiast się nim kierować. Należy także wykorzystywać zróżnicowanie bazowych właściwości problemu, aby przewidzieć rezultat, na jakim nam zależy. Przyjrzałbym się zatem zróżnicowaniu geograficznemu, które skorelowałbym z wskaźnikiem zaległości płatniczych, aby sformułować problem predykcyjny: czy możemy przewidzieć, czy klient nie spłaci kredytu? Jeżeli z kolei coś takiego rzeczywiście nastąpi, to czy możemy odzyskać chociaż ułamek użytej kwoty, być może w drodze intensywnej windykacji?

Do kwestii etycznych jeszcze wrócimy, ale należy zauważyć, że tendencyjność danych może znacząco wpłynąć na wyniki, które chcemy uzyskać. W przypadku udzielania pożyczek musimy uważać, aby nie godzić w interesy grup zmarginalizowanych poprzez kierowanie się sposobem, w jaki odbywało się to w przeszłości.

Aby sformułować pytanie normatywne, zacznijmy od wskaźnika, na który chcemy wpłynąć: nie chodzi o to, czy klient nie spłaci pożyczki (czyli o prawdopodobieństwo niewywiązania się z zobowiązań), lecz o oczekiwane korzyści netto wyniesione z jej udzielenia. Temat wartości oczekiwanych jeszcze poruszymy, ale póki co wystarczy powiedzieć, że do sformułowania i udzielenia odpowiedzi na wszystkie pytania normatywne musimy uzyskać jasne rozeznanie w kosztach i korzyściach płynących z naszych działań. Jakimi dźwigniami dysponujemy? Z pewnością jesteśmy w stanie określić wysokość pożyczki i, rzecz jasna, czy należy jej udzielić. Banki mają możliwość określania oprocentowania i zmieniania go, ale ze względu na przepisy prawne ta dźwignia może nie być dostępna dla przedsiębiorstw innego typu.

Wskaźnik zaległości płatniczych		Potencjalna wartość
Pytanie biznesowe W jaki sposób zwiększyć ilość sprzedaży na kredyt bez zwiększania wskaźnika zaległości?	Dlaczego? Na jakie KPI staram się wpłynąć?	
Pytanie deskryptywne Z jakimi prawidłowościami w zaległościach płatniczych mieliśmy do czynienia w przeszłości?	Czy mogę oddzielić efekty indywidualne od zagregowanych?	\$
Pytanie predykcyjne Czy mogę przewidzieć prawdopodobieństwo wystąpienia zaległości dla indywidualnego klienta?	Czy dane są w jakiś sposób tendencyjne? Co z selekcją negatywną?	\$\$
Pytanie normatywne Komu powinienem udzielać kredytów o jakiej wysokości i o jakim oprocentowaniu?	Ile środków można odzyskać w przypadku niewywiązania się klienta z płatności?	\$\$\$\$

Rysunek 3.8. Różne pytania związane z przypadkiem podejmowania decyzji o udzielaniu kredytów

Optymalizacja zapasów magazynowych

Częstym problemem dla wielu firm jest określenie, ile jednostek każdego produktu powinno znaleźć się w magazynie każdego sklepu. Na podobnej zasadzie banki muszą stale decydować o tym, ile gotówki powinny umieszczać w bankomatach. Zaczniemy tym razem od końca, czyli pytania normatywnego (rysunek 3.9).

Optymalizacja zapasów magazynowych		Potencjalna wartość
Pytanie biznesowe Ile jednostek każdego z naszych produktów powinno znaleźć się w każdym sklepie?	Dlaczego? Na jakie KPI staram się wpłynąć?	
Pytanie deskryptywne Jak wygląda czasowe zróżnicowanie sprzedaży?	Czy widoczne są tendencje sezonowe?	\$
Pytanie predykcyjne Czy mogę przewidzieć wysokość sprzedaży każdego artykułu w każdym sklepie każdego dnia?	Jakie czynniki bazowe decydują o sprzedaży? Czy w danych pojawiają się wartości odstające?	\$\$
Pytanie normatywne Ile egzemplarzy każdego produktu powinno znaleźć się w magazynie, aby zmaksymalizować sprzedaż bez podwyższania kosztów?	Jakie są koszty i korzyści płynące z niedoboru lub nadwyżki artykułów w magazynach?	\$\$\$\$

Rysunek 3.9. Różne pytania związane z przypadkiem podejmowania decyzji o optymalizacji zapasów magazynowych

Rozważmy koszty i korzyści związane z nadmiarem lub niedoborem danego artykułu na stanie. Jeżeli egzemplarzy nie wystarczy, sprzedaż danego dnia spada, więc kosztem jest spadek dochodu. Niedobór może również zwiększać koszt transportu i logistyki w wystarczająco dużym stopniu, aby to uwzględnić w analizie. Co z nadmiarem? Istnieje możliwość, że wartość tych artykułów obniży się, czy to w wyniku spadku jakości, słabego zarządzania lub kradzieży, czy też dlatego, że jutro pojawi się nowa, lepsza alternatywa, więc popyt na stary produkt spadnie do zera.

Na etapie normatywnym zależy nam na określeniu właściwej liczby egzemplarzy każdego artykułu umożliwiającej zminimalizowanie *oczekiwanej* sumy kosztów. Później przybliżę jeszcze szczegóły tego, jak można to zrobić. Na czym polega podstawowa niepewność związana z tym problemem? Wspomożemy się AI, aby uporać się z nią na etapie predykcyjnym.

Po pierwsze, nie wiemy, jak wiele jednostek każdego produktu zostanie sprzedanych. Gdybyśmy codziennie sprzedawali jednakową liczbę stu jednostek, musielibyśmy zawsze mieć ich na stanie co najmniej tyle. Niezadowolony klient reprezentuje nie tylko zaprzepaszczoną na bieżąco, pojedynczą możliwość zbytu, ale potencjalnie wiele takich możliwości w przyszłości: jego znajomi mogą postanowić nie przyjść do sklepu, więc miejmy nadzieję, że nie był influencerem! Co za tym idzie, powinniśmy zacząć od stworzenia prognozy popytu w określonym przedziale czasowym. Jaki powinien to być jednak okres? Dzień? Tydzień? Zależy to od wszystkich innych kosztów: jeśli transport jest tani w porównaniu z kosztem innych zagrożeń w rodzaju kradzieży lub deprecjacji, możesz bez problemu uzupełnić zapasy następnego dnia (tak jak w przypadku bankomatów). W przeciwnym wypadku konieczne może być przewidzenie prawdopodobieństwa tego, że takie zdarzenia mogą rzeczywiście nastąpić. Także tutaj poruszamy się po wartościach oczekiwanych, które omówimy w jednym z późniejszych rozdziałów, ale jak sam widzisz, optymalizacja w takich warunkach może być niezwykle trudna.

Możemy teraz pokierować analizą deskryptywną: w jaki sposób sprzedaż różni się w zależności od czasu i lokalizacji geograficznej? Czy pojawiają się tendencje sezonowe? Co z kradzieżami i napadami? Jeżeli Twoje produkty są dobrami trwałymi (takimi jak samochody, łódki, telefony, laptopy itp.), w jaki sposób ich wartość spada z biegiem czasu? Teraz masz możliwość uzyskania oglądu.

Personel sklepu

Ostatni przykład dotyczy problemu wyboru liczby sprzedawców pracujących w sklepie (rysunek 3.10). Przypomina to w pewnym sensie problem zarządzania zasobami magazynowymi: jakie koszty i korzyści płyną z niedostatecznego lub nadmiernego obsadzenia stanowisk w sklepie?

Jeżeli nie wystarczy nam ludzi, z całą pewnością uzyskamy niższe przychody: klienci zmuszeni do długiego czekania wyjdą ze sklepu i udadzą się do konkurencji. Mogą też zostać, poczekać i dokonać zakupu, ale ich niższy poziom zadowolenia prawdopodobnie przełoży się na wyższy wskaźnik odpływu, co wpłynie na przyszłe przychody. Z drugiej strony nadmiar liczby personelu prowadzi do niepotrzebnych, a przewidywalnych kosztów. Co za tym idzie, spodziewane zyski — przychód ze sprzedaży minus koszt zatrudnienia pracowników — wydają się KPI, które sensownie byłoby zoptymalizować. Efekt odpływu klientów może być na dobry początek zbyt rozległym zjawiskiem, więc zacznijmy od próby wskazania takiej liczby sprzedawców w sklepach, która pozwoli na uzyskanie

Personel sklepu		Potencjalna wartość
Pytanie biznesowe Czy mogę lepiej obsadzić stanowiska w każdym sklepie?	Dlaczego? Na jakie KPI staram się wpłynąć?	
Pytanie deskryptywne Jak wygląda korelacja między zadowoleniem klientów i sprzedażą a naszymi przeszłymi decyzjami personalnymi?	Co z kosztami? Czy różnią się w zależności od sklepu?	\$
Pytanie predykcyjne Jeżeli przydzielę X sprzedawców do sklepu Y, jak przełożą się to na zadowolenie klientów, sprzedaż i koszty?	Jak zmierzyć zadowolenie klientów?	\$\$
Pytanie normatywne Jak wielu sprzedawców powinienem codziennie zatrudniać, aby zwiększyć zysk i zadowolenie klientów?	Czy nie upraszczamy? Czy będziemy w stanie ocenić wyniki po podjęciu nowych decyzji? Co jest operacyjnie wykonalne?	\$\$\$\$

Rysunek 3.10. Różne pytania związane z przypadkiem podejmowania decyzji o obsadzie personelu

najwyższego możliwego zysku w ciągu dnia. Jeżeli uda nam się uporać z tym trudnym problemem, będziemy mogli zabrać się za optymalizację wizji długofalowej (przypomnij sobie tutaj o wartości płynącej z uproszczeń).

Co musimy wiedzieć, aby móc rozwiązać ten problem? Jak postąpilibyśmy, gdyby żadnej podstawowej niepewności nie było? Idealnie byłoby znać liczbę klientów przychodzących do sklepu o danej porze dnia, np. w każdej godzinie lub przedziale czasowym w rodzaju 30 minut. Musimy przewidzieć przepływ klientów przez każdy z naszych sklepów. Zaprowadzi nas to oczywiście do czasu oczekiwania względem wielkości personelu sprzedażowego. Konieczne może być, abyśmy wskazali akceptowalny czas oczekiwania, jako że brak czasu oczekiwania może być zbyt kosztowny, ponieważ mamy zarówno godziny szczytu, w których klientów jest ogrom, jak i okresy przestoju, kiedy wydaje się, że w sklepie jest za dużo pracowników. Czas oczekiwania wpływa na nasz zysk już teraz.

Czego powinniśmy zatem szukać w danych? Musimy wykorzystać zróżnicowanie popytu i obsady sklepów, ale powinniśmy też uwzględnić rezultaty, które chcemy przewidzieć. Do głowy od razu przychodzą cztery: sprzedaż, zysk, czas oczekiwania i zadowolenie klienta. Etap deskryptywny należy przygotować z myślą o wyszukiwaniu takich korelacji.

Główne wnioski

- *Cele biznesowe zazwyczaj są już zdefiniowane:* musimy jednak nauczyć się zadawać właściwe pytania biznesowe, aby te cele osiągnąć.
- *Zawsze zaczynaj od celu biznesowego i działaj wstecz:* do każdej planowanej lub już podjętej decyzji należy odnieść cel biznesowy, który chce się osiągnąć. Można wówczas cofnąć się i wskazać zbiór możliwych dźwigni wraz ze sposobami, w jakie mogą doprowadzić do rezultatów wpływających na działalność przedsiębiorstwa.

- *Sekwencja pytań „dlaczego?” może pomóc w zdefiniowaniu właściwego celu biznesowego do osiągnięcia:* takie oddolne podejście ogólnie pomaga we wskazywaniu celów biznesowych i poszerzaniu zbiorów działań, jakie możemy podjąć. Niekiedy jednak można też używać podejścia odgórnego, podobnego do rozkładu wskaźnika konwersji.
- *Pytania deskryptywne, predykcyjne i normatywne:* pytania deskryptywne odnoszą się do bieżącego stanu celu biznesowego; pytania predykcyjne dotyczą jego przyszłego stanu; pytania normatywne pomagają nam wybrać odpowiednie dźwignie, których pociągnięcie umożliwi uzyskanie możliwie najlepszego przyszłego scenariusza.

Dodatkowe materiały

Nie znalazłem żadnych książek traktujących o zadawaniu dobrych pytań biznesowych w kontekście procesu decyzyjnego. Nie oznacza to bynajmniej, że jako pierwszy zwracam uwagę na to, że sposób, w jaki ujmujemy nasze pytania biznesowe, może istotnie wpłynąć na rezultaty. Niemal w każdej książce o metodologii nauki o danych pojawia się przynajmniej wzmianka o tym zagadnieniu. Możesz ponownie rzucić okiem na materiały podane w rozdziale 1. lub sięgnąć po książkę *Analiza danych w biznesie. Sztuka podejmowania skutecznych decyzji* (Helion, 2014) Fostera Provosta i Toma Fawcetta.

Moim zdaniem w dostępnej literaturze można dostrzec co najmniej dwa niedopatrzenia: analityków rzadko interesuje rozwiązywanie problemów normatywnych, a zamiast tego wolą skupiać się na dostarczaniu wysokiej jakości rozwiązań predykcyjnych. Ponadto literatura skierowana do ludzi biznesu jak dotąd nie była w stanie dostarczyć całościowego oglądu problemów decyzyjnych, które można rozwiązać za pomocą AI i myślenia analitycznego.

Kilka z omówionych tutaj przypadków użycia można znaleźć w innych źródłach, ale nie jestem pewien, jak szczegółowo są opisane. Możesz poszukać w sieci białych ksiąg napisanych przez firmy consultingowe, z których być może wyciągniesz ciekawe spostrzeżenia (tego typu firmy oczywiście jednak zarabiają na rozwiązywaniu takich przypadków, więc nie oczekuj zbyt wielu szczegółów).

Na temat platform dwustronnych przyjemnie mi się czytało *Platform Revolution: How Networked Markets Are Transforming the Economy and How to Make Them Work for You* Geoffreya G. Parkera, Marshalla W. Van Alstynę'a i Sangeeta Paula Choudary'ego, a także *Matchmakers: The New Economics of Multisided Platforms* Davida S. Evansa i Richarda Schmalenseego.

Etyczne aspekty uczenia maszynowego są ważnym zagadnieniem, do którego odniosę się w rozdziale 6. i 8.

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 

Opieraj swoje decyzje na technikach analizy danych!

Czy technologie sztucznej inteligencji spełnią obietnicę wielkiego sukcesu? Dotychczas stosunkowo niewielu przedsiębiorstwom udało się dokonać przełomu dzięki modelom biznesowym opartym na nowych technologiach. Niepowodzenia te można różnie tłumaczyć. Często problemem jest brak umiejętności analitycznych, a przecież rozwój przedsiębiorstwa zależy od jakości decyzji, a nie samego wykorzystywania danych lub technologii predykcyjnych. Okazuje się jednak, że lepsze decyzje można podejmować dzięki wykorzystaniu sztucznej inteligencji i dostępnych danych.

Oto praktyczny przewodnik po wypróbowanej kompleksowej metodzie ułatwiającej przekładanie decyzji biznesowych na łatwe w realizacji, normatywne rozwiązania, oparte na danych i sztucznej inteligencji. Badacze danych, analitycy i menedżerowie znajdą tu techniki ułatwiające zadawanie właściwych pytań i generowanie wartości z wykorzystaniem nowoczesnych technologii AI i reguł decyzyjnych. Przedstawiono szczegóły procesu decyzyjnego, począwszy od pożądaných konsekwencji lub wyników, aby następnie cofnąć się do kwestii działań, które można podjąć, i skończyć na omówieniu problemów oraz szans pojawiających się w wyniku wpłynięcia na niepewność i przyczynowość. Opisano też, jak formułować i rozwiązywać problemy normatywne.

Z książki dowiesz się, jak:

- przekształcać pytania biznesowe w normatywne rozwiązania
- rozkładać decyzje biznesowe na etapy i stosować do nich różne metody analityczne
- zrozumieć i zaakceptować niepewność w procesie decyzyjnym
- optymalizować decyzje za pomocą predykcyjnych i normatywnych metod analizy
- generować znaczne wartości za pomocą technologii opartych na AI i danych

Dr Daniel Vaughan od piętnastu lat zajmuje się rozwiązywaniem problemów przy użyciu metod predykcyjnych i normatywnych. Obecnie prowadzi dział *data science* w Airbnb w Ameryce Łacińskiej. Wcześniej był dyrektorem do spraw danych i kierownikiem działu *data science* w Telefónica México.

Helion 

 helion.pl

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

Sprawdź nasze szkolenia!

SZKOLENIA



AKADEMIA IT & BUSINESS

HELIONSZKOLENIA.PL

KOD KORZYŚCI
Sięgnij po więcej! ▶



ISBN 978-83-283-7346-4



INFORMATYKA W NAJLEPSZYM WYDANIU

Cena: 59,00 zł