

УДК 004.6
ББК 32.973.26-018.2
Ц55

Jike Chong, Yue Cathy Chang

How to Lead in Data Science

Original English language edition published by Manning Publications USA.

Copyright © 2021 by Manning Publications.

Russian-language edition copyright © 2024 by Eksmo Publishing House.

All rights reserved.

Цзикэ Чонг.

Ц55 Как быть успешным в Data Science : эффективное управление проектами и развитие профессиональной команды / Цзикэ Чонг, Юэ Кэти Чанг ; [перевод с английского М. А. Райтмана]. — Москва : Эксмо, 2024. — 640 с. — (Data Science. Лучшие книги о науке о данных).

ISBN 978-5-04-184474-5

Практическое руководство предназначено для тех, кто хочет преуспеть в стремительно развивающейся сфере Data Science. Это исчерпывающий обзор ключевых аспектов науки о данных, включая основы управления проектами, создание успешной команды и стратегическое планирование роста.

Книга поможет выявить и решить основные проблемы в организации, прокачать навыки сотрудников, приоритезировать проекты с учетом возможных рисков и разработать эффективные пайплайны и бизнес-стратегии. Вы создадите корпоративную культуру управления данными и приведете свою команду к успеху.

УДК 004.6

ББК 32.973.26-018.2

ISBN 978-5-04-184474-5

© Райтман М.А., перевод на русский язык, 2024

© Оформление. ООО «Издательство «Эксмо», 2024

Краткое содержание

Глава 1. Что делает дата-сайентиста успешным?	33
---	----

ЧАСТЬ I. ТЕХЛИД: РАЗВИВАЕМ КАЧЕСТВА РУКОВОДИТЕЛЯ

59

Глава 2. Способности для руководства проектами	61
--	----

Глава 3. Качества, необходимые для руководства проектами	122
--	-----

ЧАСТЬ II. МЕНЕДЖЕР: ВОСПИТЫВАЮЩИЙ КОМАНДЫ

179

Глава 4. Способности для руководства людьми	181
---	-----

Глава 5. Качества для руководства людьми	234
--	-----

ЧАСТЬ III. ДИРЕКТОР: РУКОВОДЯЩИЙ ФУНКЦИЕЙ

285

Глава 6. Способности для руководства функцией	287
---	-----

Глава 7. Качества для руководства функцией	361
--	-----

ЧАСТЬ IV. РУКОВОДИТЕЛЬ: ВДОХНОВЛЯЮЩИЙ ИНДУСТРИЮ

413

Глава 8. Способности для руководства компанией	415
--	-----

Глава 9. Качества для руководства компанией	472
---	-----

ЧАСТЬ V. LOOP И БУДУЩЕЕ

533

Глава 10. Ландшафт, организация, возможность и практика	534
---	-----

Глава 11. Руководство в data science и перспективы на будущее	599
---	-----

Содержание

Введение	14
Предисловие	16
Примечания	19
Благодарности	20
О книге	22
Об авторах	28
Изображение на обложке	31
Глава 1. Что делает дата-сайентиста успешным?	33
1.1. Ожидания от дата-сайентистов	34
1.1.1. Диаграмма Венна спустя десятилетие	35
1.1.2. Чего не хватает?	36
1.1.3. Понимание возможностей и мотивации: оценка способностей и качеств	38
1.2. Карьерный рост в data science	39
1.2.1. Проблемы на собеседовании и при продвижении	41
1.2.2. Чего ищут менеджеры (по найму)?	46
Резюме	56
Примечания	57

ЧАСТЬ I

ТЕХЛИД: РАЗВИВАЕМ КАЧЕСТВА РУКОВОДИТЕЛЯ

Глава 2. Способности для руководства проектами	61
2.1. Технология: инструменты и навыки	63
2.1.1. Постановка задачи для максимизации влияния на бизнес	63
2.1.2. Выявление закономерностей данных	65
2.1.3. Установка ожиданий успеха	78
2.2. Исполнение: лучшие практические методы	82
2.2.1. Определение и приоритизация проектов на основании нечетких требований	82

2.2.2. Планирование и управление проектами data science	87
2.2.3. Уравновешивание сложных технических компромиссов	96
2.3. Экспертные знания: глубокое понимание предметной области	103
2.3.1. Прояснение бизнес-контекста возможностей	103
2.3.2. Учет особенностей источников данных предметной области	105
2.3.3. Управление организационными структурами	111
2.4. Самооценка и фокус развития	116
2.4.1. Понимание своих интересов и сильных сторон руководства	116
2.4.2. Практикуем с помощью процесса CPR	117
2.4.3. Разработка плана приоритизировать-практиковать- и-выполнять	118
2.4.4. Примечание для менеджеров техлидов DS	119
<i>Резюме</i>	120
<i>Примечания</i>	121
Глава 3. Качества, необходимые для руководства проектами	122
3.1. Этические стандарты поведения	123
3.1.1. Действие в насущных интересах клиентов	124
3.1.2. Адаптация к бизнес-приоритетам в динамичной рыночной среде	126
3.1.3. Уверенная передача знаний	131
3.2. Развитие строгости, повышение стандартов	136
3.2.1. Прояснение основ научной строгости	136
3.2.2. Мониторинг аномалий и принятие на себя ответственности за создание ценности предприятия	143
3.2.3. Принятие ответственности за ценность предприятия	147
3.3. Позитивная установка	163
3.3.1. Позитивный настрой и упорство в преодолении неудач	163
3.3.2. Любознательность и сотрудничество при реагировании на инциденты	164
3.3.3. Уважение различных точек зрения при взаимовыгодном сотрудничестве	169
3.4. Самооценка и фокус развития	172
3.4.1. Понимание своих интересов и сильных сторон руководства	172
3.4.2. Практикуем с помощью процесса CPR	173
3.4.3. Селф-коучинг с помощью модели GROW	174
3.4.4. Примечание для менеджеров техлидов DS	176
<i>Резюме</i>	176
<i>Примечания</i>	177

ЧАСТЬ II

МЕНЕДЖЕР: ВОСПИТЫВАЮЩИЙ КОМАНДЫ

<i>Примечание</i>	180
Глава 4. Способности для руководства людьми	181
4.1. Технология: инструменты и навыки	182
4.1.1. Эффективное делегирование проектов	183
4.1.2. Управление согласованностью между моделями и проектами	186
4.1.3. Предоставление рекомендаций купить-или-создать	190
4.2. Исполнение: лучшие практики	194
4.2.1. Создание эффективных команд под вашим руководством	194
Оказание влияния на команды партнеров для увеличения своего влияния	206
Управление отношениями с руководителем	210
Экспертные знания: глубокое понимание предметной области	215
Расширение знаний в нескольких технических и бизнес-областях	216
4.3.2. Распознавание основных возможностей предметной области	220
4.3.3. Оценка рентабельности инвестиций для приоритизации, несмотря на отсутствующие данные	224
4.4. Самооценка и фокус развития	228
4.4.1. Понимание своих интересов и сильных сторон руководства	229
4.4.2. Практикуем с помощью процесса CPR	230
<i>Резюме</i>	231
<i>Примечания</i>	232
Глава 5. Качества для руководства людьми	234
5.1. Этические стандарты поведения	235
5.1.1. Развитие команды с помощью наставничества, менторства и консультирования	236
5.1.2. Уверенное представление команды на межфункциональных обсуждениях	240
5.1.3. Содействие и взаимное выполнение более широких управленческих обязанностей	245
5.2. Воспитание строгости, повышение стандартов	248
5.2.1. Наблюдение и смягчение антипаттернов в системах ML и DS	249
5.2.2. Эффективное извлечение уроков из инцидентов	255
5.2.3. Управление ясностью за счет перевода сложных проблем в лаконичные нарративы	260

5.3. Позитивная установка	264
5.3.1. Управление графиком творца по отношению к графику менеджера	264
5.3.2. Доверяем исполнение членам команды	270
5.3.3. Формирование культуры институализированного обучения	275
5.4. Самооценка и фокус развития	278
5.4.1. Понимание своих интересов и сильных сторон руководства	278
5.4.2. Практикуем с помощью процесса CPR	280
<i>Резюме</i>	281
<i>Примечания</i>	282

ЧАСТЬ III

ДИРЕКТОР: РУКОВОДЯЩИЙ ФУНКЦИЕЙ

Глава 6. Способности для руководства функцией	287
6.1. Технология: инструменты и навыки	289
6.1.1. Создание технологических дорожных карт	290
6.1.2. Руководство функцией DS для создания нужных свойств продукта для нужных людей в нужное время	298
6.1.3. Финансирование и защита многообещающих проектов	302
6.2. Исполнение: лучшие практические методы	309
6.2.1. Последовательное достижение результата за счет управления людьми, процессами и платформами	310
6.2.2. Выстраивание устойчивой функции с четкими карьерными картами и надежным процессом найма	323
6.2.3. Поддержка руководителей при исполнении главных инициатив компании	335
6.3. Экспертные знания: глубокое понимание предметной области	339
6.3.1. Предвидение бизнес-потребностей на всех этапах разработки продукта	339
6.3.2. Реагирование на срочные проблемы с помощью проверенных первоначальных решений	344
6.3.3. Управление фундаментальными бизнес-воздействиями с глубоким пониманием предметной области	349
6.4. Самооценка и фокус развития	355
6.4.1. Понимание своих интересов и сильных сторон руководства	355
6.4.2. Практикуем с помощью процесса CPR	357
<i>Резюме</i>	358
<i>Примечания</i>	359

Глава 7. Качества для руководства функцией	361
7.1. Этические стандарты поведения	362
7.1.1. Установление формализаций проекта во всей функции	362
7.1.2. Наставничество в качестве социального лидера с помощью интерпретаций, нарративов и требований	368
7.1.3. Организация инициатив для предоставления возможностей карьерного роста	373
7.2. Воспитание строгости, повышение стандартов	378
7.2.1. Управление успешным процессом годового планирования	379
7.2.2. Избегание антипаттернов планирования и исполнения	384
7.2.3. Обеспечение обязательств со стороны партнеров и команд	389
7.3. Позитивная установка	394
7.3.1. Признание и содействие разнообразию внутри вашей команды	395
7.3.2. Практика инклюзивности при принятии решений	401
7.3.3. Воспитание причастности к вашей функции	405
7.4. Самооценка и фокус развития	408
7.4.1. Понимание своих интересов и сильных сторон руководства	408
7.4.2. Практикуем с помощью процесса CPR	410
Резюме	411
Примечания	412

ЧАСТЬ IV

РУКОВОДИТЕЛЬ: ВДОХНОВЛЯЮЩИЙ ИНДУСТРИЮ

Глава 8. Способности для руководства компанией	415
8.1. Технология: инструменты и навыки	416
8.1.1. Разработка бизнес-стратегий и дорожных карт в области данных на срок от одного до трех лет	417
8.1.2. Внедрение управляемой данными культуры во все аспекты бизнес-процессов	423
8.1.3. Структурирование инновационных и продуктивных организаций data science	431
8.2. Исполнение: лучшие практики	437
8.2.1. Внедрение возможностей data science в видение и миссию	438
8.2.2. Создание устойчивого пула талантов в data science	443
8.2.3. Прояснение вашей роли как композитора или дирижера	450
8.3. Экспертные знания: глубокое понимание предметной индустрии	453
8.3.1. Выявление дифференциации и конкурентоспособности среди аналогов в индустрии	453

8.3.2. Руководство бизнесом с помощью точек поворота, когда это необходимо	458
8.3.3. Формулировка бизнес-планов для новых продуктов и сервисов	462
8.4. Самооценка и фокус развития	466
8.4.1. Понимание своих интересов и сильных сторон руководства	467
8.4.2. Практикуем с помощью процесса CPR	469
<i>Резюме</i>	469
<i>Примечания</i>	471
Глава 9. Качества для руководства компанией	472
9.1. Этические стандарты поведения	473
9.1.1. Практика ответственного машинного обучения на основании этических принципов	474
9.1.2. Обеспечение доверия и безопасности клиентов	478
9.1.3. Принятие на себя социальной ответственности за решения	483
9.2. Воспитание строгости, повышение стандартов	488
9.2.1. Создание продуктивной и гармоничной рабочей среды	489
9.2.2. Увеличение скорости и улучшение качества принятия решений	494
9.2.3. Фокусировка на ценности предприятия	498
9.3. Позитивная установка	507
9.3.1. Демонстрация исполнительного присутствия	508
9.3.2. Утверждение командного превосходства в индустрии	516
9.3.3. Изучение и применение лучших практик различных отраслей	521
9.4. Самооценка и фокус развития	525
9.4.1. Понимание своих интересов и сильных сторон руководства	526
9.4.2. Практикуем с помощью процесса CPR	527
<i>Резюме</i>	528
<i>Примечания</i>	529

ЧАСТЬ V LOOP И БУДУЩЕЕ

Глава 10. Ландшафт, организация, возможность и практика	534
10.1. Ландшафт	535
10.1.1. Озера и хранилища данных	536
10.1.2. Поточковая обработка	537
10.1.3. Самостоятельное понимание	538
10.1.4. Автоматизация операций данных и ML	540

10.1.5. Управление данными	541
10.1.6. Периодический анализ основных тенденций в архитектуре	543
10.2. Организация	544
10.2.1. Функциональная организационная структура	544
10.2.2. Дивизиональная организационная структура	548
10.2.3. Матричная организационная структура	550
10.2.4. Альтернативная организационная структура	554
10.2.5. Управление возможностями и сложностями в различных структурах	556
10.3. Возможности	557
10.3.1. Оцениваем отрасли	558
10.3.2. Оцениваем компанию	559
10.3.3. Оцениваем команду	563
10.3.4. Оцениваем роль	567
10.3.5. Адаптация к новой роли	569
10.4. Практика	581
10.4.1. Набор навыков, которые вы можете принять в команду	581
10.4.2. Новые карьерные направления для руководителей DS	588
10.5. Обзор LOOP	593
<i>Резюме</i>	596
<i>Примечания</i>	597
Глава 11. Руководство в data science и перспективы на будущее	599
11.1. Почему, чем и как руководить в DS	600
11.1.1. Почему обучение руководству в DS становится все более важным?	600
11.1.2. Какова структура для руководства в DS?	602
11.1.3. Как мы применяем эту структуру на практике?	608
11.2. Взгляд в будущее	609
11.2.1. Роль — появление менеджеров по продукту данных	610
11.2.2. Возможности — доступность функциональных решений для данных	614
11.2.3. Обязанности: внушение доверия к данным	619
<i>Резюме</i>	622
<i>Примечания</i>	623
Эпилог	625
Предметный указатель	627

*Нашим родителям
за то, что вдохновили нас на упорный труд,
глубокое погружение и самоотдачу*

*Нашим читателям
за потраченное на чтение книги время.
Давайте вместе ускорим процесс понимания
и совершенствования нашего мира*

*Друг другу
за споры, воодушевление и поддержку
на протяжении всего пути*

Введение

В течение последнего десятилетия я побывал председателем и сопредседателем на более чем сорока международных конференциях по данным и искусственному интеллекту. Получил невероятные впечатления, став свидетелем эволюции и влияния на бизнес аналитики, data science и машинного обучения по всему миру. Data science остается одной из самых быстрорастущих рабочих областей в индустрии. Во время моей работы ведущим дата-сайентистом в O'Reilly Media каждое проведенное исследование подтверждало, что компании продолжают инвестировать в инфраструктуры данных, data science и машинное обучение. Мы также обнаружили, что компании, по праву считающиеся лучшими в использовании data science и машинного обучения, инвестировали в фундаментальные технологии и использовали эти инструменты для расширения своих возможностей постепенно, по одному кейсу за раз.

Большая часть того, что мы читаем, относится к инструментам или прорывам в моделях, но в реальности узкими местами в работе большинства компаний остаются организационные моменты. Важнейшим компонентом является распознавание организационного мастерства в людях, культуре и структуре. Если у вас нет необходимых специалистов и организационной структуры, вы уступите конкурентам, у которых они есть.

Поскольку спрос на дата-сайентистов продолжает расти, а количество обучающих программ увеличивается, ко мне часто обращаются за советом.

Новички спрашивают, как вступить в ряды дата-сайентистов, опытные дата-сайентисты интересуются, как подняться на следующий карьерный уровень.

К сожалению, сложно найти четкие рекомендации, как оставаться компетентным и эффективным на протяжении всей карьеры в data science. Большинство посвященной этому литературы сфокусировано на подготовке к работе: где учиться, какие навыки приобрести и как пройти собеседование и получить первую должность. Мало кто описывает, как уже нанятому на работу дата-сайентисту сохранить успех и подняться по карьерной лестнице.

«Как быть успешным в data science. Эффективное управление проектами и развитие профессиональной команды» — это важное руководство для дата-сайентистов на разных этапах их карьеры в качестве линейного дата-сайентиста, такого как техлид (технический руководитель, tech lead), эксперт, ведущий или высокопоставленный дата-сайентист, или в качестве руководителя, такого как менеджер, директор или ответственный за направление data science в компании. Эта книга предназначена для дата-сайентистов, желающих подняться на следующий карьерный уровень. В ней также представлено руководство по инструментам и методам в контексте оказания помощи дата-сайентистам, чтобы увеличить их положительное влияние на бизнес и общество.

Я знаком с авторами, Цзикэ и Кэти, много лет. Они обладают разнообразным опытом работы с широким кругом организаций, включая публичные и частные компании, а также навыками практического консультирования. Я видел, как они преподают материал из этой книги на обучающих курсах для специалистов с разным опытом и из разных отраслей. Их курсы всегда пользовались популярностью, их хорошо принимали на конференциях, на которых мне довелось председательствовать.

Цель этой книги — заполнить информационные пробелы в знаниях дата-сайентистов, ищущих способы продвижения по карьерной лестнице. Читателям, находящимся на разных этапах карьеры, стоит возвращаться к этой книге по мере своего развития и становления. Этот труд я планирую рекомендовать дата-сайентистам. Надеюсь, он вдохновит на новые дискуссии и выпуск новых книг по теме. В ближайшие годы эта книга пригодится дата-сайентистам и работающим с ними партнерам!

БЕН ЛОРИКА

Бен Лорика — главный автор на GradientFlow.com, сопредседатель саммита по НЛП и Ray-саммита; бывший ведущий дата-сайентист и председатель программ в O'Reilly Media; владелец и организатор подкаста The Data Exchange with Ben Lorica (сайт thedataexchange.media); консультировал многие стартапы и организации, включая Databricks, Anyscale и Faculty.ai.

Предисловие

Как руководитель в области data science, вы можете масштабировать свои данные, алгоритмы и команду, но делаете ли вы это? Что такое руководство вообще? Как вы применяете способности для достижения более значимого результата, чем полученный индивидуально? Влияете ли вы на бизнес и окружающих, воспитываете, направляете и вдохновляете проекты и сотрудников?

С этими вопросами сталкиваются многие практикующие дата-сайентисты, пытающиеся продвинуться по карьерной лестнице в этой быстро растущей и стремительно развивающейся индустрии. Большинство специалистов-практиков работают в компаниях с менее чем десятью дата-сайентистами, выполняя широкий спектр обязанностей по руководству проектами, координации специалистов различного профиля, разработке дорожных карт и влиянию на руководство компании. Зачастую их роли не определены и связаны с нереалистичными ожиданиями.

В то же время в мире существуют более 150 000 дата-сайентистов, и их число ежегодно увеличивается на 37% [1]. Компаниям требуются талантливые сотрудники, чтобы руководить проектами, воспитывать команды, управлять функциями и вдохновлять индустрии.

Несмотря на существование посвященных этой области знания блогов, подкастов и платформ, таких как комнаты Meetups и Clubhouse, до настоящего момента не существовало всеобъемлющего практического руководства, рассматривающего развитие карьеры в data science.

По настоянию друзей и коллег, многим из которых мы помогли вырасти от линейных специалистов до руководителей в data science, а затем и руководителей организаций с семьюдесятью дата-сайентистами, мы написали книгу, чтобы поделиться знаниями, обретенными за последнее десятилетие. Идеи, включенные в эту книгу, почерпнуты из нашего собственного опыта создания, развития и консультирования дата-сайентистов в публичных и частных компаниях. Мы также проинтервьюировали десятки успешных руководителей дата-сайентистов и воспользовались их практическим опытом.

Работая над этим руководством, мы с радостью обнаружили, что основы самосовершенствования согласуются с некоторыми известными концепциями. Какова вероятность, что основы формирования набора навыков, принятия ответственности и оказания влияния в мире существуют тысячи лет? В этой книге мы познакомимся с такими этапами руководства, как развитие индивидуального руководства, воспитание команды, руководство функцией (то есть направлением в компании) и вдохновение индустрии, основанными на учении Конфуция [2]. Мы обсудим необходимые на каждом этапе руководства профессиональные навыки, которые называем способностями, и гибкие психологические навыки, которые называем качествами. Качества являются необходимыми чертами характера, позволяющими практическим специалистам обрести счастье и благополучие, вдохновленные греческим философом Аристотелем [3]. Этапы карьеры, а также способности и качества представлены на рис. 1.

В этих проверенных временем концепциях и отображаются преобразующие идеи, индивидуальный опыт и отраслевые примеры руководства в data science. Вы можете использовать их для укрепления уверенности в себе как в руководителе, осознавая свои сильные стороны, выявляя области, в которых пока плохо разбираетесь, открывая возможности для новых практик и используя свою команду и организацию для достижения более существенного влияния на бизнес.

В этой книге мы демонстрируем желаемые цели приложения способностей и качеств в области data science. Вы можете обращаться к ним, определяя профессиональное развитие своей команды, но не стоит использовать их в качестве причин для отсрочки продвижения сотрудников по службе. Член команды, продемонстрировавший способности и качества в одних областях и потенциал в других, может быть готов принять больше обязанностей, а значит, готов и к продвижению по службе.

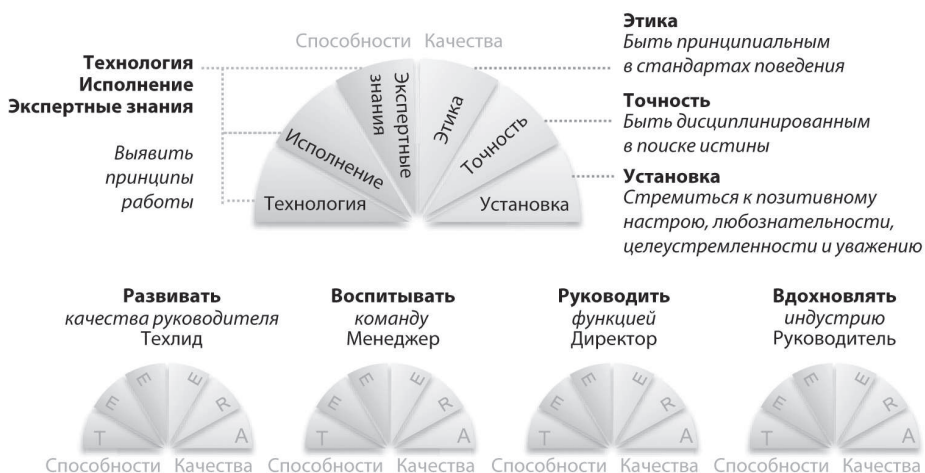


Рис. 1. Индивидуальные способности и качества, необходимые на каждом этапе вашего карьерного роста

Лучшие практические методы, процессы и советы применимы к ситуациям, с которыми сталкиваются техлиды, выполняющие функции линейного сотрудника на уровне штатного (staff), ведущего (principal) и высокопоставленного (distinguished) дата-сайентиста, а также управляющие персоналом на уровне менеджера, директора и руководителя. Это продемонстрировано на внутренней стороне задней обложки книги, поскольку данная информация приобретет больше смысла, когда вы прочитаете об этих уровнях.

Чтобы помочь распознать ситуации, в которых вы можете применить практический опыт, процессы и советы, мы выделили семь реальных сценариев, с которыми сталкивались практикующие специалисты data science, начиная от выпускников вузов и заканчивая опытными руководителями. Для каждого мы описали ситуацию, выявили причины и предложили решения, предоставив вам возможность подумать, как бы вы справились с подобными ситуациями.

Мы задумывали эту книгу как спутник вашего карьерного роста на ближайшие годы. Если она вам поможет в сложных ситуациях, пожалуйста, сообщите нам. И не забудьте поделиться знаниями в социальных сетях!

Для нас честь вдохновить вас на лучшую работу в вашей карьере и извлечь максимальную пользу из своего потенциала, чтобы оказать более существенное положительное влияние на мир с помощью data science!

ЦЗИКЭ ЧОНГ и ЮЭ КЭТИ ЧАНГ