

# ОГЛАВЛЕНИЕ

|  |           |
|--|-----------|
| <b>Об авторе .....</b>                                   | <b>8</b>  |
| <b>Благодарности .....</b>                               | <b>10</b> |
| От издательства.....                                     | 11        |
| <b>Введение .....</b>                                    | <b>12</b> |
| Для кого эта книга .....                                 | 14        |
| Как читать эту книгу.....                                | 15        |
| <b>Глава 1. Как мы принимаем решения .....</b>           | <b>17</b> |
| Четыреста сравнительно честных способов .....            | 20        |
| Чему можно научиться у Amazon?.....                      | 21        |
| Аналитический паралич.....                               | 23        |
| Погрешности — правило штангенциркуля.....                | 25        |
| Принцип Парето .....                                     | 26        |
| Можно ли принимать решения только на основе данных?..... | 27        |
| <b>Глава 2. Делаем анализ данных .....</b>               | <b>31</b> |
| Артефакты анализа данных .....                           | 33        |
| Бизнес-анализ данных .....                               | 34        |
| Гипотезы и инсайты .....                                 | 34        |
| Отчеты, дашборды и метрики.....                          | 37        |
| Артефакты машинного обучения.....                        | 41        |
| Артефакты инженерии.....                                 | 43        |
| Кто анализирует данные .....                             | 45        |
| Идеальная кнопка .....                                   | 46        |

|  |            |
|--|------------|
| Продать аналитику внутри компании.....                             | 48         |
| Конфликт исследователя и бизнеса .....                             | 49         |
| Недостатки статистического подхода в аналитике .....               | 51         |
| <b>Глава 3. Строим аналитику с нуля .....</b>                      | <b>55</b>  |
| Первый шаг .....   | 56         |
| Выбираем технологии .....  | 57         |
| Поговорим об аутсорсе .....  | 61         |
| Наем и увольнения .....  | 64         |
| Кому подчиняются аналитики .....                                   | 67         |
| Должен ли руководитель аналитики писать код .....                  | 68         |
| Управление задачами.....   | 71         |
| Как управлять романтиками .....                                    | 76         |
| <b>Глава 4. Делаем аналитические задачи .....</b>                  | <b>79</b>  |
| Как ставить задачи аналитикам .....                                | 80         |
| Как проверять задачи.....  | 83         |
| Как тестировать и выкладывать изменения<br>в рабочую систему ..... | 87         |
| Как защищать задачу перед инициатором .....                        | 87         |
| Нужно ли уметь программировать? .....                              | 88         |
| Датасет.....   | 90         |
| Описательная статистика .....                                      | 91         |
| Графики .....  | 94         |
| Общий подход к визуализации данных .....                           | 98         |
| Парный анализ данных.....  | 100        |
| Технический долг.....  | 101        |
| <b>Глава 5. Данные .....</b>                                       | <b>103</b> |
| Как собираются данные.....   | 104        |
| Big Data.....  | 105        |
| Связность данных .....   | 106        |
| Много данных не бывает .....                                       | 107        |
| Доступ к данным .....  | 108        |

|  |            |
|--|------------|
| Качество данных.....                                   | 110        |
| Как проверяется и контролируется качество данных ..... | 112        |
| Типы данных .....                                      | 114        |
| Форматы хранения данных .....                          | 116        |
| Способы получения данных.....                          | 120        |
| <b>Глава 6. Хранилища данных.....</b>                  | <b>121</b> |
| Зачем нужны хранилища данных.....                      | 122        |
| Слои хранилища данных.....                             | 125        |
| Какие бывают хранилища.....                            | 126        |
| Как данные попадают в хранилища .....                  | 130        |
| Hadoop и MapReduce.....                                | 132        |
| Spark .....  | 136        |
| Оптимизация скорости работы .....                      | 140        |
| Архивация данных и устаревание .....                   | 141        |
| Мониторинг хранилищ данных .....                       | 143        |
| Личный опыт.....                                       | 144        |
| <b>Глава 7. Инструменты анализа данных.....</b>        | <b>147</b> |
| Электронные таблицы.....                               | 148        |
| Сервисы блокнотов .....                                | 149        |
| Инструменты визуального анализа .....                  | 151        |
| Пакеты статистического анализа данных.....             | 153        |
| Работа с данными в облаках.....                        | 154        |
| Что такое хорошая отчетная система .....               | 156        |
| Сводные таблицы .....                                  | 159        |
| OLAP-кубы.....   | 164        |
| Корпоративные и персональные BI-системы.....           | 168        |
| Мой опыт .....   | 170        |
| <b>Глава 8. Алгоритмы машинного обучения .....</b>     | <b>171</b> |
| Типы ML-задач.....                                     | 174        |
| Метрики ML-задач.....                                  | 178        |

|   |            |
|---|------------|
| ML изнутри .....                                    | 182        |
| Линейная регрессия .....                            | 183        |
| Логистическая регрессия .....                       | 184        |
| Деревья решений.....                                | 186        |
| Ошибки обучения .....                               | 188        |
| Как бороться с переобучением .....                  | 190        |
| Ансамбли.....                                       | 194        |
| <b>Глава 9. Машинное обучение на практике .....</b> | <b>197</b> |
| Как изучать машинное обучение .....                 | 198        |
| Соревнования по ML .....                            | 200        |
| Искусственный интеллект.....                        | 202        |
| Необходимые преобразования данных .....             | 204        |
| Точность и стоимость ML-решения .....               | 207        |
| Простота решения .....                              | 208        |
| Трудоемкость проверки результата .....              | 210        |
| Mechanical Turk / Yandex Toloka .....               | 210        |
| ML и большие данные .....                           | 212        |
| Recency, Frequency и Monetary.....                  | 212        |
| Последний совет.....                                | 215        |
| <b>Глава 10. Внедрение ML в жизнь:</b>              |            |
| <b>гипотезы и эксперименты.....</b>                 | <b>217</b> |
| Гипотезы .....                                      | 218        |
| Планируем тест гипотезы .....                       | 220        |
| Что такое гипотеза в статистике .....               | 222        |
| Статистическая значимость гипотез.....              | 226        |
| Статистические критерии для р-значений .....        | 229        |
| Бутстрэп .....                                      | 232        |
| Байесовская статистика .....                        | 234        |
| А/Б-тесты в реальности .....                        | 237        |
| А/А-тесты .....                                     | 240        |
| Еще несколько слов о А/Б-тестах.....                | 242        |

|   |            |
|---|------------|
| Что делать перед А/Б-тестом .....           | 244        |
| Конвейер экспериментов .....                | 245        |
| <b>Глава 11. Этика данных .....</b>         | <b>247</b> |
| Как за нами следят .....                    | 248        |
| Хорошее и плохое использование данных ..... | 255        |
| Проблема утечки данных .....                | 258        |
| Этика использования данных .....            | 260        |
| Как защищают пользовательские данные .....  | 263        |
| <b>Глава 12. Задачи и стартапы .....</b>    | <b>267</b> |
| Веб-аналитика в рекламе .....               | 268        |
| Внутренняя веб-аналитика .....              | 271        |
| Маркетинг на основе баз данных .....        | 276        |
| Стартапы .....                              | 280        |
| Личный опыт .....                           | 284        |
| <b>Глава 13. Строим карьеру .....</b>       | <b>291</b> |
| Старт карьеры .....                         | 292        |
| Как искать работу .....                     | 294        |
| Требования к кандидатам .....               | 296        |
| Вы приняли оффер .....                      | 298        |
| Как развиваться и работать .....            | 298        |
| Когда менять место работы .....             | 302        |
| Нужно ли все знать? .....                   | 304        |
| <b>Эпилог .....</b>                         | <b>308</b> |
| <b>Список литературы .....</b>              | <b>309</b> |