
Оглавление

Предисловие	15
Целевая аудитория.....	15
Почему я написала эту книгу.....	15
Структура издания.....	16
Условные обозначения, используемые в книге.....	17
Использование примеров кода	17
Благодарности.....	18
От издательства	20
О научном редакторе русскоязычного издания	20
Глава 1. Роль бизнес-аналитика и важность аналитики.....	21
Роль бизнес-аналитика.....	21
Навыки	22
Обязанности.....	23
Типы аналитиков	24
Зачем бизнес-аналитику разбираться в аналитике	28
Взрывной рост данных.....	29
Бизнес-контекст.....	30
Аналитика.....	31
Бизнес-аналитик, повышающий ценность аналитики	32
Бизнес-задачи, решаемые аналитикой.....	33
Сотрудничество с другими отделами.....	34
Навыки, используемые в аналитике	35
Python и R	35
Жизненный цикл аналитического проекта.....	36
Резюме	36
Глава 2. Методологии для бизнес-аналитиков и аналитических проектов	37
Понимание бизнес-задачи.....	38
Определите бизнес-задачи	38
Оцените ситуацию	39

Определите цели	40
Определите подход и план.....	40
Оцените инструменты и методики	40
Изучение и подготовка данных	41
Оцените содержание и качество данных	41
Выберите и очистите данные	41
Подготовьте и интегрируйте данные.....	42
Создайте выборки данных для разработки модели	42
Моделирование и оценка	42
Выберите метод аналитики.....	42
Создайте и оцените модель	43
Развертывание	43
Оцените эффективность модели	43
Определите интервалы оценки	43
Эксплуатация модели.....	44
Мониторинг моделей.....	44
Срок службы модели.....	44
Дообучение	44
Резюме	45
Глава 3. Введение в R и Python.....	46
Варианты установки и настройки R и Python.....	47
Зачем изучать R и Python.....	47
Изучение обоих языков. Одновременно или по очереди?	47
Плюсы и минусы различных стратегий обучения	48
Установка R	48
Установка Python	49
Скрипты на R и Python	50
Создание скриптов на языке R.....	50
Создание скриптов на Python	52
Концепции ООП.....	55
Структура ООП.....	55
Принципы ООП	58
Типы данных в R и Python	64
Типы данных в R.....	64
Структуры данных в R.....	65
Типы данных в Python	67
Структуры данных в Python	69
Взаимодействие с реляционными базами данных	76
Почему именно реляционные базы данных	77

Подключение к реляционным базам данных в R.....	77
Примеры использования R и реляционных баз данных	78
SQLite.....	79
Подключение к реляционным базам данных в Python	79
Примеры использования Python и реляционных баз данных	80
Резюме	80
Глава 4. Статистический анализ данных с помощью R и Python.....	81
Примеры аналитических проектов	81
Отток клиентов в сфере телекоммуникаций	82
A/B-тестирование	84
Маркетинговые кампании.....	85
Финансовое прогнозирование	86
Диагностика в здравоохранении	87
Формулирование проблемы или задачи	88
Подход к аналитической задаче	89
Классификация	93
Регрессия.....	93
Что мы хотим измерить.....	93
Подходы к анализу.....	94
EDA.....	95
Обучение без учителя.....	95
Статистический анализ для регрессии	96
Анализ для классификации.....	98
Роль проверки гипотез	100
Визуализация в аналитике	101
Визуализация в R и Python для поддержки EDA.....	101
Визуализация регрессии.....	103
Визуализация классификации	110
Резюме	120
Глава 5. Разведочный анализ данных с помощью R и Python.....	121
Качество данных.....	121
Характеристики качества данных.....	122
Профилирование данных	123
Кластеризация и обучение без учителя	131
Цель обучения без учителя.....	132
Пример влияния кластеризации на обучение с учителем.....	133
Кластеризация методом k-средних.....	135
Иерархическая кластеризация	139
Другие методы обучения без учителя в EDA.....	142

Обнаружение выбросов	143
Выбросы в регрессии.....	143
Выбросы при классификации	145
Подготовка данных для моделирования	147
Выборка, или сэмплинг	148
Обучение и тестирование.....	150
Преобразование данных.....	152
Манипулирование данными.....	160
Выбор и сокращение количества признаков.....	164
Отбор признаков	165
Методы уменьшения размерности	173
Резюме	177
Глава 6. Применение и оценка моделирования с помощью R и Python.....	178
Этапы моделирования.....	178
Выбор модели и ее обучение	178
Оценка модели.....	180
Оптимизация модели	181
Развертывание модели	182
Мониторинг и обслуживание моделей	182
Выбор подходящего алгоритма	183
Регрессия	184
Распространенные примеры использования	185
Уравнение линейной регрессии.....	186
Линейная регрессия в R.....	187
Линейная регрессия в Python.....	189
Пример использования линейной регрессии	194
Другие виды регрессии	198
Сложности, связанные с регрессионными моделями	200
Другие алгоритмы регрессии	201
Деревья решений для регрессии.....	204
Оценка линейной регрессии.....	206
Классификация.....	212
Распространенные варианты использования	214
Алгоритмы классификации.....	216
Классификация в R	217
Классификация в Python	220
Пример классификации: отток клиентов в телекоммуникационной отрасли.....	221
Оценка классификации.....	235
Оценка классификации для конкретной задачи.....	245
Резюме	255

Глава 7. Выбор моделей и алгоритмов	257
Алгоритмы.....	257
Критерии алгоритмов.....	258
Тип задачи.....	259
Интерпретируемые модели.....	261
Точность прогнозирования.....	271
Скорость обучения.....	272
Скорость прогнозирования.....	273
Настройка гиперпараметров.....	275
Работа с небольшими наборами данных.....	280
Работа с большими наборами данных.....	281
Взаимодействие признаков.....	282
Характеристики данных.....	286
Пример: выбор подходящего алгоритма.....	288
Выбор алгоритма для прогнозирования продаж.....	289
Оценка критериев.....	291
Принятие решения и реализация.....	292
Резюме.....	293
Глава 8. Эксплуатация моделей (ModelOps)	294
Обзор ModelOps.....	294
Процессы ModelOps.....	296
Оценка моделей.....	296
Оценка моделей в R: использование Shiny-приложений для оценки в реальном времени.....	298
Оценка моделей на Python: развертывание моделей с помощью Streamlit.....	302
Мониторинг моделей.....	304
Ключевые метрики и индикаторы для мониторинга производительности моделей.....	304
Методы автоматического мониторинга моделей.....	305
Дообучение моделей.....	312
Запуск событий для дообучения модели.....	312
Методы автоматического дообучения модели.....	313
Реализация на R: использование заданий cron для запланированного дообучения.....	313
Реализация на Python: использование Airflow для управления рабочими процессами.....	314
Создание отчетов.....	315
Содержание и структура итоговых отчетов.....	315
Методы автоматического создания отчетов.....	319

Реализация на R: создание отчетов с помощью R Markdown и knitr	320
Реализация на Python: создание отчетов с помощью Jupyter Notebooks и nbconvert.....	322
Контроль версий и воспроизводимость моделей.....	325
Методы совместной работы и документирования	326
Примеры использования ModelOps	326
Прогнозирование розничных продаж: автоматизация оценки и мониторинга.....	326
Обнаружение мошенничества: динамическое дообучение моделей и отчетность	327
Прогнозирование оттока клиентов: запланированное дообучение модели и создание итогового отчета	327
Интеграция с существующими системами и инфраструктурой.....	328
Будущее направление развития MLOps	328
Резюме	331
Глава 9. Продвинутая визуализация	332
Расширенная визуализация с R Shiny	333
Что такое R Shiny	333
Особенности и возможности R Shiny.....	334
Настройка среды.....	338
Создание вашего первого приложения Shiny.....	338
Продвинутая разработка пользовательского интерфейса.....	346
Пример: создание дашборда для мониторинга продаж в реальном времени.....	358
Визуализация в Python	362
Обзор визуализации в Python.....	363
Распространенные библиотеки: Matplotlib, Seaborn, Plotly и Dash	364
Matplotlib: основы визуализации на Python.....	366
Форматирование графиков	367
Статистические графики: диаграммы рассеяния, тепловые карты, скрипичные диаграммы	369
Интерактивные графики с Plotly	373
3D-графики с Matplotlib и Plotly	375
Визуализация геопространственных данных.....	379
Создание дашборда: используйте Plotly Dash.....	382
Пример: использование Python для продвинутого проекта по визуализации данных	384
Выбор между R Shiny и библиотеками визуализации Python	393
Резюме	394

Глава 10. Использование в аналитике современных типов данных	395
Полуструктурированные данные (JSON)	396
Использование Python для работы с данными в формате JSON	396
Использование R для работы с данными в формате JSON	400
Данные из социальных сетей	403
Типы данных социальных сетей	403
Использование Python для анализа данных из социальных сетей	407
Использование R для анализа данных из социальных сетей	430
Данные с изображений	442
Обработка изображений на Python	442
Обработка изображений на R	444
Видеоданные	445
Использование Python для обработки видеоданных	446
Использование R для работы с видеоданными	450
Резюме	454
Глава 11. Измерение пользы аналитики для бизнеса и роль ИИ	455
В чем польза аналитики для бизнеса	456
Стратегическое влияние	456
Операционная эффективность	456
Удовлетворенность и лояльность клиентов	457
Метрики и KPI для измерения пользы для бизнеса	457
Финансовые метрики	458
Операционные метрики	458
Клиентские метрики	459
Согласование метрик с целями и задачами организации	459
Использование метрик для демонстрации ценности	460
Метрики и ключевые показатели эффективности на практике	460
Примеры бизнес-кейсов, демонстрирующих пользу аналитики	461
Шаг 1. Определение проблемы и постановка измеримых целей	461
Шаг 2. Определение метрик для оценки успеха и провала	461
Шаг 3. Внедрение аналитических решений	462
Шаг 4. Измерение и демонстрация ценности	463
Шаг 5. Подготовка отчетности и постоянное совершенствование	464
Генеративный ИИ в бизнес-аналитике	464
Введение в генеративный искусственный интеллект	465
Применение в дизайне товаров	466
Применение в создании контента	466
Применение в маркетинге	466
Улучшение клиентского опыта	467

Повышение операционной эффективности	467
Перспективы и проблемы.....	467
Примеры использования ИИ в бизнес-аналитике.....	468
Пример 1. Аналитика и рекомендации для клиентов на основе ИИ.....	468
Пример 2. Генеративный ИИ в создании контента.....	468
Пример 3. Оптимизация цепочки поставок с помощью ИИ	469
Пример 4. Улучшение процесса принятия решений с помощью ИИ	469
Пример 5. ИИ в аналитике здравоохранения.....	469
Пример 6. Генеративный ИИ для персонализации клиентского опыта	470
Пример 7. ИИ в аналитике розничной торговли.....	470
Устранение фактических несоответствий и сотрудничество человека и ИИ	470
Перспективы на будущее	471
Сложности и рекомендации	471
Проблемы интеграции и масштабируемости при внедрении ИИ-решений.....	472
Устранение предвзятости и обеспечение справедливости при принятии решений на основе ИИ.....	472
Технические и организационные сложности при внедрении ИИ.....	473
Вопросы стоимости и ресурсов	473
Инвестиции в ИИ.....	473
Резюме	474
Об авторе	475
Иллюстрация на обложке	476