

# Оглавление

<b>Предисловие .....</b>	<b>12</b>
От издательства .....	15
<b>Введение .....</b>	<b>16</b>
<b>Почему Data Science? .....</b>	<b>18</b>
<b>Глава 1. Об основах без лишних слов.....</b>	<b>21</b>
1.1. Подготовка данных.....	22
Формат данных.....	23
Типы переменных .....	24
Выбор переменных .....	25
Конструирование признаков .....	25
Неполные данные .....	26
1.2. Выбор алгоритма.....	27
Обучение без учителя.....	28
Обучение с учителем .....	29
Обучение с подкреплением .....	30
Другие факторы.....	31

---

1.3. Настройка параметров .....	31
1.4. Оценка результатов .....	33
Метрики классификации .....	34
Метрика регрессии .....	35
Валидация.....	36
1.5. Краткие итоги .....	38
<b>Глава 2. Кластеризация методом k-средних .....</b>	<b>39</b>
2.1. Поиск кластеров клиентов.....	40
2.2. Пример: профили кинозрителей .....	41
2.3. Определение кластеров.....	42
Сколько кластеров существует? .....	44
Что включают кластеры? .....	46
2.4. Ограничения .....	48
2.5. Краткие итоги.....	49
<b>Глава 3. Метод главных компонент .....</b>	<b>51</b>
3.1. Изучение пищевой ценности .....	52
3.2. Главные компоненты.....	53
3.3. Пример: анализ пищевых групп.....	56
3.4. Ограничения .....	61
3.5. Краткие итоги.....	64
<b>Глава 4. Ассоциативные правила .....</b>	<b>65</b>
4.1. Поиск покупательских шаблонов.....	66
4.2. Поддержка, достоверность и лифт .....	67

4.3. Пример: ведение продуктовых продаж .....	69
4.4. Принцип Apriori .....	72
Поиск товарных наборов с высокой поддержкой.....	73
Поиск товарных правил с высокой достоверностью или лифтом.....	74
4.5. Ограничения .....	75
4.6. Краткие итоги.....	76

**Глава 5. Анализ социальных сетей ..... 77**

5.1. Составление схемы отношений.....	78
5.2. Пример: геополитика в торговле оружием.....	80
5.3. Лувенский метод .....	84
5.4. Алгоритм PageRank .....	86
5.5. Ограничения .....	90
5.6. Краткие итоги .....	91

**Глава 6. Регрессионный анализ ..... 93**

6.1. Выведение линии тренда.....	94
6.2. Пример: предсказание цен на дома .....	95
6.3. Градиентный спуск .....	98
6.4. Коэффициенты регрессии.....	101
6.5. Коэффициенты корреляции.....	102
6.6. Ограничения .....	104
6.7. Краткие итоги.....	106

**Глава 7. Метод k-ближайших соседей и обнаружение  
аномалий ..... 107**

7.1. Пищевая экспертиза.....	108
------------------------------	-----

---

7.2. Яблоко от яблони недалеко падает .....	109
7.3. Пример: истинные различия в вине .....	111
7.4. Обнаружение аномалий.....	113
7.5. Ограничения .....	114
7.6. Краткие итоги.....	115
<b>Глава 8. Метод опорных векторов.....</b>	<b>117</b>
8.1 «Нет» или «о, нет!»?.....	118
8.2. Пример: обнаружение сердечно-сосудистых заболеваний .....	118
8.3. Построение оптимальной границы.....	120
8.4. Ограничения .....	124
8.5. Краткие итоги.....	125
<b>Глава 9. Дерево решений.....</b>	<b>127</b>
9.1. Прогноз выживания в катастрофе .....	128
9.2. Пример: спасение с тонущего «Титаника» .....	128
9.3. Создание дерева решений .....	131
9.4. Ограничения .....	133
9.5. Краткие итоги.....	135
<b>Глава 10. Случайные леса.....</b>	<b>137</b>
10.1. Мудрость толпы .....	138
10.2. Пример: предсказание криминальной активности.....	139
10.3. Ансамбли .....	144
10.4. Бэггинг.....	145



Приложение D. Другие метрики оценки.....	183
Метрики классификации .....	183
Метрики регрессии.....	186
<b>Глоссарий .....</b>	<b>188</b>
<b>Литература и ссылки на источники .....</b>	<b>199</b>
Источники на английском языке .....	199
Литература на русском языке.....	202
<b>Об авторах.....</b>	<b>204</b>