

УДК 004.056.5(075.8)

ББК 32.97я73

Ц349

*Печатается по решению кафедры информационно-аналитических систем безопасности Института компьютерных технологий и информационной безопасности Южного федерального университета (протокол № 9 от 13 мая 2022 г.)*

#### **Рецензенты:**

доктор технических наук, профессор, заведующий кафедрой информатики Таганрогского института имени А. П. Чехова (филиал) РГЭУ (РИНХ)

*Я. Е. Ромм*

доктор технических наук, профессор, профессор кафедры информационно-аналитических систем безопасности Южного федерального университета *А. В. Боженюк*

#### **Целых, А. Н.**

Ц349 Извлечение знаний методами машинного обучения : учебное пособие по курсам «Модели и методы инженерии знаний», «Методы машинного обучения» / А. Н. Целых, Э. М. Котов ; Южный федеральный университет. – Ростов-на-Дону ; Таганрог : Издательство Южного федерального университета, 2022. – 105 с.

ISBN 978-5-9275-4215-4

Пособие посвящено рассмотрению теоретических подходов к извлечению информации из текстов, рассмотрению и систематизации прогностических методов и моделей в маркетинге и интернет-рекламе, на основе выборки больших данных. В данном пособии рассмотрено применение машинного обучения при построении рекомендательных систем.

Пособие предназначено для студентов высших учебных заведений, обучающихся по направлениям 10.03.01 "Информационная безопасность" (направленность программы: "Информационно-аналитические системы безопасности") по курсу "Модели и методы инженерии знаний" и 09.04.03 "Прикладная информатика" (направленность программы: "Машинное обучение и технологии больших данных") по курсу "Методы машинного обучения".

УДК 004.056.5(075.8)

ББК 32.97я73

ISBN 978-5-9275-4215-4

© Южный федеральный университет, 2022

© Целых А. Н., Котов Э. М., 2022

© Оформление. Макет. Издательство

Южного федерального университета, 2022



## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	5
1. ОБЗОР ПРАКТИКИ ПРИМЕНЕНИЯ БОЛЬШИХ ДАННЫХ .....	8
1.1. Анализ использования больших данных российскими компа- ниями .....	8
1.2. Анализ российских кейсов .....	13
2. БОЛЬШИЕ ДАННЫЕ В РЕКЛАМЕ И МАРКЕТИНГЕ .....	24
2.1. Большие данные и рекламные форматы .....	24
2.2. Большие данные в продажах рекламного инвентаря .....	40
2.3. Экосистема рынка больших данных в интернет-рекламе: основ- ные участники и технологии .....	43
2.4. Влияние Больших Данных на рекламный рынок .....	47
3. БОЛЬШИЕ ДАННЫЕ И РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ .....	57
3.1. Основные понятия в области рекомендательных систем .....	57
3.2. Что такое рекомендательная система .....	58
3.3. Выгоды от внедрения системы рекомендаций .....	61
3.4. Рекомендательная система основанная на контенте .....	61
3.4.1. <i>Высокоуровневая архитектура систем на основе контента</i> .....	62
3.4.2. <i>Преимущества и недостатки фильтрации на основе кон-                 тента</i> .....	67
3.5. Коллаборативная фильтрация .....	68
3.5.1. <i>Типы коллаборативной фильтрации</i> .....	69
3.5.2. <i>Фильтрация на основе памяти</i> .....	69
3.5.3. <i>Фильтрация на основе моделей</i> .....	71
3.5.4. <i>Характеристики и проблемы совместной фильтрации</i> .....	74
3.6. Гибридная рекомендательная система .....	79
3.6.1. <i>Взвешенная модель</i> .....	79

*Содержание*

---

3.6.2. Смешанный подход .....	80
3.6.3. Комбинаторный подход .....	80
3.6.4. Гибридные переключения .....	80
3.6.5. Способствующая модель .....	82
3.6.6. Каскадная модель .....	82
3.6.7. Гибрид метатурвня .....	82
4. ПОСТРОЕНИЕ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ .....	85
4.1. Подготовка данных .....	85
4.2. Выбор модели построения рекомендательной систем .....	86
4.3. Построение модели системы .....	87
4.4. Выбор средства реализации .....	88
4.5. Экспериментальные тесты .....	91
4.6. Эффективность выбранного решения .....	93
4.7. Описание работы системы .....	94
4.8. Предварительная подготовка данных .....	94
4.9. Построение модели .....	97
4.10. Оценка модели .....	97
ЗАКЛЮЧЕНИЕ .....	99
СПИСОК ЛИТЕРАТУРЫ .....	102