

**Чубукова И. А.**

Data Mining: учебное пособие /И.А. Чубукова.- 2-е изд., испр. — М.: Интернет-Университет Информационных Технологий; БИНОМ. Лаборатория знаний, 2008. — 382 с: ил., табл.- (Серия «Основы информационных технологий»)

# Содержание

## Раздел 1. Введение в Data Mining

### Лекция 1. Что такое Data Mining

Понятие Статистики

Понятие Машинного обучения

Понятие Искусственного интеллекта

Сравнение статистики, машинного обучения и Data mining

Развитие технологии баз данных

Понятие Data Mining

Data Mining как часть рынка информационных технологий

### Лекция 2. Данные

Что такое данные

Набор данных и их атрибутов

Измерения

Шкалы

Типы наборов данных

Форматы хранения данных

Базы данных. Основные положения

Системы управления базами данных, СУБД

Классификация видов данных

Метаданные

### Лекция 3. Методы и стадии Data Mining

Классификация стадий Data mining

Классификация методов Data Mining

Свойства методов Data Mining

### Лекция 4. Задачи Data Mining. Информация и знания

Задачи Data Mining

Связь понятий

Информация

Свойства информации

Требования, предъявляемые к информации

Знания

Сопоставление и сравнение понятий «информация», «данные», «знание»

## Лекция 5. Задачи Data Mining. Классификация и кластеризация

Задача классификации

Процесс классификации

Методы, применяемые для решения задач классификации

Точность классификации: оценка уровня ошибок

Оценивание классификационных методов

Задача кластеризации

Оценка качества кластеризации

Процесс кластеризации

Применение кластерного анализа

Выводы

## Лекция 6. Задачи Data Mining. Прогнозирование и визуализация

Задача прогнозирования

Задача визуализации

## Лекция 7. Сферы применения Data Mining

Применение Data Mining для решения бизнес-задач

Исследования для правительства

Data Mining для научных исследований

Web Mining

Text Mining

Call Mining

## **Раздел 2. Методы и алгоритмы Data Mining**

### Лекция 8. Основы анализа данных

Анализ данных в Microsoft Excel

Описательная статистика

Корреляционный анализ

Регрессионный анализ

Выводы

### Лекция 9. Методы классификации и прогнозирования.

Деревья решений

Алгоритмы

Выводы

### Лекция 10. Методы классификации и прогнозирования.

Метод опорных векторов.

Метод «ближайшего соседа».

Байесовская классификация Метод опорных векторов

Метод «ближайшего соседа» или системы рассуждений  
на основе аналогичных случаев

Байесовская классификация

Лекция 11. Методы классификации и прогнозирования.

Нейронные сети

Элементы нейронных сетей

Архитектура нейронных сетей

Обучение нейронных сетей

Модели нейронных сетей

Программное обеспечение для работы с нейронными сетями

Пример решения задачи

Пакет Matlab

Лекция 12. Нейронные сети. Самоорганизующиеся карты

Кохонена .

Классификация нейронных сетей

Подготовка данных для обучения

Выбор структуры нейронной сети

Карты Кохонена

Выводы

Лекция 13. Методы кластерного анализа. Иерархические методы

Методы кластерного анализа

Лекция 14. Методы кластерного анализа. Итеративные методы

Алгоритм k-средних (k-means)

Алгоритм PAM ( partitioning around Medoids)

Предварительное сокращение размерности

Факторный анализ

Итеративная кластеризация в SPSS

Процесс кластерного анализа. Рекомендуемые этапы

Сложности и проблемы, которые могут возникнуть при  
применении кластерного анализа

Сравнительный анализ иерархических и неиерархических  
методов кластеризации

Новые алгоритмы и некоторые модификации алгоритмов кластерного анализа

#### Лекция 15. Методы поиска ассоциативных правил

Введение в ассоциативные правила

Часто встречающиеся шаблоны или образцы

Поддержка

Характеристики ассоциативных правил

Границы поддержки и достоверности ассоциативного правила

Методы поиска ассоциативных правил

Пример работы Apriori алгоритма на базе данных D

Разновидности алгоритма Apriori

Пример решения задачи поиска ассоциативных правил

#### Лекция 16. Способы визуального представления данных.

Методы визуализации

Визуализация инструментов Data Mining

Визуализация Data Mining моделей

Методы визуализации

Качество визуализации

Представление пространственных характеристик

Основные тенденции в области визуализации

Выводы

### **Раздел 3. Процесс Data Mining**

#### Лекция 17. Комплексный подход к внедрению Data Mining, OLAP и хранилищ данных в СППР

Классификация СППР

OLAP-системы

OLAP-продукты

Интеграция OLAP и Data Mining

Хранилища данных

Преимущества использования хранилищ данных

#### Лекция 18. Процесс Data Mining. Начальные этапы

Этап 1. Анализ предметной области

Этап 2. Постановка задачи

Этап 3. Подготовка данных

Выводы

Лекция 19. Процесс Data Mining. Очистка данных

Инструменты очистки данных Проблема

Выводы по подготовке данных

Лекция 20. Процесс Data Mining. Построение и использование

модели

Моделирование

Виды моделей

Математическая модель

Этап 4. Построение модели

Этап 5. Проверка и оценка моделей

Этап 6. Выбор модели

Этап 7. Применение модели

Этап 8. Коррекция и обновление модели

Погрешности в процессе Data Mining

Выводы

Лекция 21. Организационные и человеческие факторы в Data Mining.

Стандарты Data Mining

Организационные Факторы

Человеческие факторы. Роли в Data Mining

CRISP-DM методология

SEMMA методология

Другие стандарты Data Mining

Стандарт PMML

Стандарты, относящиеся к унификации интерфейсов

#### **Раздел 4. Инструменты Data Mining**

Лекция 22. Рынок инструментов Data Mining

Поставщики Data Mining

Классификация инструментов Data Mining

Программное обеспечение Data Mining для поиска ассоциативных правил

Программное обеспечение для решения задач кластеризации и сегментации

Программное обеспечение для решения задач классификации.

Программное обеспечение Data Mining для решения задач оценивания и прогнозирования

Вывод Лекция 23. Инструменты Data Mining. SAS Enterprise Miner

Обзор программного продукта

Графический интерфейс (GUI) для анализа данных

Инструментарий для углубленного интеллектуального анализа данных

Набор инструментов для подготовки, агрегации и исследования данных

Интегрированный комплекс разнообразных методов моделирования

Интегрированные средства сравнения моделей и пакеты результатов

Скоринг по модели и простота развертывания модели

Гибкость благодаря открытости и расширяемости

Встроенная стратегия обнаружения данных

Распределенная система интеллектуального анализа данных,

ориентированная на крупные предприятия

Основные характеристики пакета SAS Enterprise Miner 5.1

Специализированное хранилище данных

Подход SAS к созданию информационно-аналитических систем

Технические требования пакета SAS\* Enterprise Miner

Лекция 24. Инструменты Data Mining. Система PolyAnalyst

Архитектура системы

PolyAnalyst Workplace - лаборатория аналитика

Аналитический инструментарий PolyAnalyst

Модули для построения числовых моделей и прогноза числовых переменных

Алгоритмы кластеризации

Алгоритмы классификации

Алгоритмы ассоциации

Модули текстового анализа

Визуализация

Эволюционное программирование

Общесистемные характеристики PolyAnalyst

PolyAnalyst Scheduler — режим пакетной обработки

WebAnalyst

Лекция 25. Инструменты Data Mining. Программные продукты Cognos и система STATISTICA Data Miner

Программные продукты Cognos

Особенности методологии моделирования с применением Cognos 4Thought

Система STATISTICA Data Miner

Средства анализа STATISTICA Data Miner

Лекция 26. Инструменты Oracle Data Mining и Deductor

Oracle Data Mining

Аналитическая платформа Deductor

Лекция 27. Инструмент KXEN

Реинжиниринг аналитического процесса

Технические характеристики продукта

Предпосылки создания KXEN

Структура KXEN Analytic Framework Version 3.0

Технология IO LAP

Лекция 28. Data Mining консалтинг

Data Mining услуги

Работа с клиентом

Примеры решения

Выводы

Список использованной и рекомендуемой литературы



