

МАСТЕР-КЛАСС

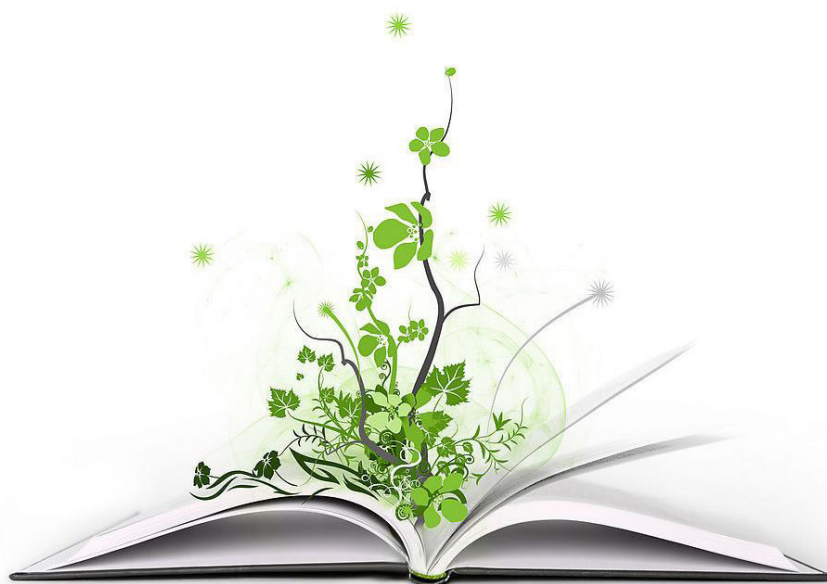
“ИСПОЛЬЗОВАНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ МЕТОДОВ АНАЛИЗА ДАННЫХ В ГИДРОМЕТЕОРОЛОГИИ”

Современные информационные технологии: средства Data Mining, которые часто называют интеллектуальными методами анализа данных, и в частности искусственные нейронные сети (ИНС), успешно применяются в самых различных областях - технике, геологии, физике, бизнесе, медицине.

ИНС вошли в практику научных и прикладных исследований, требующих решения задач аппроксимации сложных нелинейных экспериментальных функций многих переменных, прогнозирования, кластеризации, управления.

В гидрометеорологии ИНС могут быть использованы для анализа гидрометеоданных и построения вычислительных моделей сезонных изменений климата региона, краткосрочного прогнозирования погоды, ливневых дождей, града.

Особенность ИНС в том, что при построении вычислительных прогностических моделей могут использоваться как количественные, так и качественные бинарные (в общем случае, многозначные, дискретные) переменные. Основным условием построения надежных моделей погоды региона и выработки надежных ее прогнозов является полнота и качество объективных данных метеонаблюдений и измерений. Это будет являться основной задачей групп научно-образовательной экспедиционной программы и стационарных станций гидрометеонаблюдений.



РАБОЧАЯ ПРОГРАММА ПО ДИСЦИПЛИНЕ

Интеллектуальные методы анализа данных

Лекции 10 час.

Лабораторные занятия 10 час.

Всего часов по дисциплине 20 час.

Автор: д.ф.-м.н., профессор Абриков В.С.

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

1.1. Цель преподавания дисциплины

Эффективное использование накопленных знаний и информации.

Курс кратко охватывает современные подходы к анализу информации, основанные на концепции извлечения знаний из баз данных.

1.2. Задачи изучения дисциплины

- Познакомить слушателей с базовыми принципами построения моделей данных, ознакомить с концепцией Knowledge Discovery in Databases (обнаружения знаний в базах данных) и Data Mining («добыча» знаний), основными типами задач, решаемых методами Data Mining; научить эффективно пользоваться методами извлечения знаний из больших массивов данных и осуществлять тиражирование знаний.
- Научить использовать современные методы и технологии, ориентированные на управление знаниями.

1.3. В результате изучения дисциплины слушатели должны

знать и уметь:

- Эффективно управлять информацией и знанием, как на индивидуальном, так и на организационном уровнях, используя процессы создания информации и знания, их оценки, получения к ним доступа, их фильтрации, накопления, сортировки, измерения, хранения, обработки с помощью современных аналитических технологий;

- Представлять основные этапы работ по внедрению проектов Data Mining, определять критерии выбора аналитических платформ.

приобрести навыки работы на современной аналитической платформе содержащей средства Data Mining.

Предмет базируется на знании основ информатики, теории обработки данных.

2. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

1. Введение

Основы анализа данных.

Методология построения моделей сложных систем.

Модель «черного ящика».

Основные этапы построения моделей.

Методика анализа данных.

2. Методы интеллектуального анализа данных

Предпосылки развития автоматических методов добычи данных.

Многомерное представление данных.

Типы задач, решаемые Data Mining (регрессия, кластеризация, классификация, ассоциация, последовательные шаблоны).

Наиболее известные алгоритмы для решения каждого типа задач в Data Mining (нейронные сети, деревья решений, и др.).

3. Обзор примеров применения Data Mining при решении научных и прикладных задач.

4. Подготовка данных и интерпретация результатов

Этапы подготовки данных.

Формирование гипотез.

Методы сбора и систематизации фактов.

Анализ качества полученных моделей.

5. Практические аспекты

Архитектура аналитических систем.

Категории пользователей аналитических систем.

Архитектура и составные части аналитической платформы.

Итого 10 часов

3. ПЕРЕЧЕНЬ ТЕМ ЛАБОРАТОРНЫХ РАБОТ

• **Лабораторная работа № 1. Знакомство с аналитической платформой.**

Архитектура. Понятие сценария, обработчика, визуализатора. Импорт данных. Понятие процесса, измерения, факта. Создание законченных аналитических решений. Тиражирование знаний. (2 часа)

• **Лабораторная работа № 2. Многомерные отчеты и простая аналитика.**

С использованием аналитической платформы на основе прилагаемых данных (несколько таблиц с "сырыми" данными) необходимо построить сценарии для формирования многомерных отчетов и графиков. Индивидуальные задания. (2 часа)

• **Лабораторная работа № 3. Задачи кластеризации и классификации.**

С использованием аналитической платформы провести кластеризацию (сегментацию) - карты Кохонена и классификацию (деревья решений) выборки данных и проинтерпретировать результаты. Индивидуальные задания. (4 часа)

• **Лабораторная работа № 4. Временные ряды. Прогнозирование. Нейронные сети.**

С использованием аналитической платформы на основе динамического ряда построить сценарий построения прогноза - очистить данные, трансформировать к скользящему окну, обучить нейросеть, сформировать отчет. Индивидуальные задания. (2 часа)

НАГЛЯДНЫЕ ПОСОБИЯ

Курс лекций проводится с использованием презентаций подготовленных с помощью аналитической платформы.