

Содержание тем

Тема 1. Эволюция информационно-коммуникационных технологий и потенциал интеллектуального анализа данных

Этапы развития средств хранения, обработки и передачи данных. Переход от отчетности к аналитике и поддержке решений. Понятие интеллектуального анализа данных, границы применения и место в контуре управления. Аналитическая задача, объект, признаки, целевая переменная и ожидаемый результат. Качество данных, воспроизводимость, конфиденциальность и этические ограничения.

Тема 2. Концепция хранилищ данных и OLAPсистем

Назначение корпоративного хранилища данных. Предметная ориентированность, интегрированность, временная устойчивость и неизменяемость. Витрины данных, ETL/ELTпроцессы, факты и измерения, схемы «звезда» и «снежинка». OLAPкуб и операции slice, dice, drilldown, rollup, pivot. Согласование показателей, метаданные и качество аналитической отчетности.

Тема 3. Big Data и Data Mining

Характеристики больших данных: объем, скорость, разнообразие, достоверность и ценность. Распределенное хранение и обработка как организационнотехнологическая идея. Жизненный цикл исследования: понимание задачи, понимание данных, подготовка, моделирование, оценка и внедрение. Обучающая, валидационная и тестовая выборки. Переобучение, утечка данных, смещение и ограничения применимости моделей.

Тема 4. Задачи анализа данных и классификация методов

Описательный, диагностический, прогнозный и предписывающий анализ. Обучение с учителем и без учителя. Классификация, регрессия, кластеризация, поиск ассоциативных правил, сокращение размерности и выявление аномалий. Выбор метода с учетом типа данных, цели, интерпретируемости и стоимости ошибки. Метрики качества для различных типов задач.

Тема 5. Инструменты Data Mining

Классы инструментов: электронные таблицы, BIплатформы, визуальные конструкторы аналитических процессов, языки Python и R, библиотеки машинного обучения, Jupyterсреды. Импорт, профилирование, преобразование, моделирование и визуализация. Воспроизводимый аналитический процесс, документирование допущений и версий данных. Критерии выбора инструмента.

Тема 6. Работа со сложными системами

Сложная система как совокупность взаимосвязанных элементов, обратных связей и задержек. Многомерность, нелинейность, эмерджентность и неопределенность. Системные карты, сетевое и сценарное представление. Выявление прямых и косвенных связей, аномалий и узких мест. Интерпретация аналитического результата без смещения корреляции и причинности. Оценка рисков управленческого решения.

Практические занятия

№ темы	Наименование темы, по которой предусмотрено практическое занятие	Форма и содержание практического занятия
3	Тема 3. Big Data и Data Mining	Практикум «Подготовка данных и проект аналитического исследования». Слушатели выявляют пропуски, дубликаты, выбросы и несогласованности в учебном наборе данных, предлагают способы очистки и преобразования, определяют порядок разделения данных на выборки и фиксируют риски утечки данных. Результат: паспорт качества данных и план исследования.
4	Тема 4. Задачи анализа данных и классификация методов	Решение ситуационных задач по выбору метода анализа. Для предложенных прикладных ситуаций слушатели определяют тип аналитической задачи, целевую переменную, подходящий класс методов и метрику качества, оценивают последствия ошибок. Результат: таблица соответствия «задача – метод – метрика – ограничение».
5	Тема 5. Инструменты Data Mining	Практическая работа «Мини-исследование в аналитическом инструменте». Слушатели загружают учебные данные, выполняют профилирование и преобразование, получают визуализацию или результат простой модели, интерпретируют полученный результат. Результат: файл исследования и краткое аналитическое заключение.
6	Тема 6. Работа со сложными системами	Анализ конкретной ситуации «Диагностика сложной системы». Слушатели строят карту факторов и связей, выявляют обратные связи, задержки и возможные узкие места, формируют показатели наблюдения и сценарии развития. Результат: карта системы и обоснованные выводы с указанием границ достоверности.