

Объем дисциплины

Виды учебной работы	Объем в часах по семестрам		
	Всего	5 семестр	— семестр
1. Контактная работа обучающихся с преподавателем по видам учебных занятий (всего)	144		
Аудиторная работа (всего)	68	68	
· Лекции (Л)	34	34	
· семинары (С)			
· лабораторные работы (ЛР)	34	34	
2. Самостоятельная работа обучающихся (СР) (всего)	76	76	
· Проработка учебного материала лекций	10	10	
· Подготовка к семинарам			
· Подготовка к лабораторным работам	16	16	
· Подготовка к рубежному контролю, контрольной работе	6	6	
· Подготовка докладов, рефератов, презентации			
· Выполнение домашнего задания	21	21	
· Выполнение курсового проекта / курсовой работы			
· Подготовка к зачету	15	15	
· Другие, отдельные виды СРС по Положению «Об организации внеаудиторной самостоятельной работы студентов МГТУ им. Н.Э. Баумана»	8	8	
Вид промежуточной аттестации обучающегося	зачет	зачет	

- **Содержание дисциплины, структурированное по модулям.**

Модуль 1. OLAP системы. (18 часов)

Лекции по модулю 1. Содержание. Объем часов:

Лекция1.

Введение в оперативный анализ данных - OLAP систем. Цели и задачи курса. Архитектура OLAP систем: OLAP-клиент, OLAP-сервер.

(2 часа)

Лекция2.

Основные способы реализации OLAP систем. MOLAP структура и основные свойства. ROLAP структура и основные свойства. HOLAP структура и основные свойства

(2 часа)

Лекция3.

Математическая модель OLAP кубов.

Многомерное представление данных. Кубы. Измерения. Иерархии. Агрегирующие функции, меры. Виды запросов к кубам. Операции, выполняемые над гиперкубом.

(2 часа)

Лекция 4.

Общая схема и алгоритм работы с OLAP системой. Подготовка данных, загрузка данных в гиперкуб, реализация гиперкуба и построения срезов..

(2 часа)

Лекция 5.

Источники данных. Классификация типов источников данных. Предназначение, структура и свойства OLTP систем. Преимущества, недостатки, проблемы возникающие при прямом доступе к источникам данных.

(2 часа)

Лекция 6.

Извлечение и преобразование данных ETL. Подходы к реализации ETL процесса. Элементы ETL процесса. Диаграммы движения, преобразования и управления преобразованием данных. Моделирование движения данных

(2 часа)

Лекция 7.

Введение в хранилище данных (ХД). Обзор архитектур ХД. Типовая структура хранилищ данных. Основные особенности концепции ХД. Основные требования к ХД. Основные концепции ХД

(2 часа)

Лекция 8.

Задачи решаемые ХД. Многомерные хранилища данных. Гибридные ХД. Витрины данных. Централизованное ХД с витринами данных. Виртуальные ХД

(2 часа)

Лекция 9..

Трансформация данных. Основные методы трансформации данных. Группировка данных. Разгруппировка данных. Слияние данных. Операции над таблицами: объединение, внутреннее соединение, внешнее соединение, полное внешнее соединение

(2 часа)

Семинары по модулю 1. Содержание. Объем часов

РПД не предусмотрены.

Лабораторные работы по модулю 1. Содержание. Объем часов:18

1. Машинное обучение. Исследование и анализ точности распознавания символов с применением метода линейной регрессии.

(4 часа)

2. Машинное обучение. Исследование и анализ точности распознавания символов с применением метода логистической регрессии.

(4 часа)

3. Анализ временных рядов: Анализ временных рядов при помощи многослойной нейронной сети.

(4 часов)

4. Эволюционные алгоритмы: Общая схема эволюционных алгоритмов. Применение эволюционных алгоритмов к задачам оптимизации и анализа данных .

(6 часа)

Самостоятельная работа (по видам самостоятельной работы). Объем часов:

Проработка разделов лекционного курса 4 .

Подготовка к лабораторным работам 5,5.

Подготовка к рубежному контролю, контрольной работе 3 .

-

Модуль 2. Data Mining

(16 часов)

Лекции по модулю 2. Содержание. Объем часов:

Лекция 1.

1. Интеллектуальный анализ данных Data Mining. Задачи Data Mining: классификация, регрессия, ассоциативные правила, кластеризация.

(2 часа)

Лекция 2.

Модели Data Mining: описание и предсказание. Методы Data Mining: базовые методы, нечеткая логика, генетические алгоритмы, нейронные сети.

(2 часа)

Лекция 3.

Методы Data Mining – Базовые методы. Нечеткая логика

(2 часа)

Лекция 4.

Методы Data Mining – Генетические алгоритмы. Нейронные сети.

(2 часа)

Лекция 5.

Классификация. Постановка задачи. Правила классификации. Деревья решений. Методы построения правил классификации, деревьев решений

(2 часа)

Лекция 6.

Кластеризация. Постановка задачи. Меры близости основанные на расстояниях. Алгоритмы кластеризации.

(2 часа)

Лекция 7

Прогнозирование. Постановка задачи. Задачи прогнозирования. Методы и модели прогнозирования. Классификация методов прогнозирования. Аналитические методы прогнозирования. Статистические методы прогнозирования. Временные ряды

(2 часа)

Лекция 8.

Визуализация данных. Введение. Цели и задачи визуализации. Инструменты визуализации. Методы визуализации. Методы геометрических преобразований. Отображение иконок. Методы, ориентированные на пиксели. Иерархические образы

(2 часа)

Семинары по модулю 2. Содержание. Объем часов

РПД не предусмотрены.

Визуализация данных на базе построенной ассоциативной модели в среде BI QlickView

Лабораторные работы по модулю 2. Содержание. Объем часов: 16

1. Освоить инструментарий ПО Pentaho Report Designer. Научиться добавлять источники данных. Освоить навык создания отчетов и диаграмм и провести анализ полученных результатов.

(4 часа)

2. Освоить объекты бизнес - аналитики BI QlickView». Освоить инструментарий. Запустить мастер построения диаграмм. Загрузить данные из Excel. Анализ результатов.

(4 часа)

3. Подключение к разным источникам данных. Выбор таблиц с необходимыми атрибутами. Отладка скриптов и датологической модели. Построение ассоциативной модели данных на базе BI QlickView.

(4 часа)

4. Визуализация данных на базе построенной ассоциативной модели в среде BI QlickView.

Построение таблиц, диаграмм и графиков по выбранным измерениям.

(4 часа)