

**Аннотация рабочей программы дисциплины/практики**  
**Б1.О.13 Большие данные**  
**Специальность/направление подготовки: 09.03.02 Информационные системы и технологии**  
**Специализация/профиль: Информационные системы и технологии на транспорте**

**1. Цели освоения дисциплины(модуля)/практики**

Формирование у студентов необходимой теоретической базы и практических навыков, которые позволят всесторонне и системно понимать современные проблемы прикладной математики и информатики, проблемы обработки и анализа информации, а также разрабатывать и анализировать концептуальные и теоретические модели при решении научных и прикладных задач в области информационных технологий. Одна из главных проблем современной обработки и анализа данных - рост объемов данных, поэтому вопросам обработки большого объема данных посвящена данная дисциплина. Главная задача курса - сформировать целостное представление о современных проблемах анализа и обработки больших данных, помочь овладеть опытом разработки и анализа концептуальных и теоретических моделей прикладных задач анализа больших данных с применением моделей Data Mining.

**2. Компетенции, формируемые в результате освоения дисциплины (модуля) практики**

**УК-2: Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений**

Индикатор	УК-2.1. Знает виды ресурсов и ограничений для решения профессиональных задач; основные методы оценки разных способов решения задач; действующее законодательство и правовые нормы, регулирующие профессиональную деятельность.
Индикатор	УК-2.2 Умеет проводить анализ поставленной цели и формулировать задачи, которые необходимо решить для ее достижения; анализировать альтернативные варианты для достижения намеченных результатов; использовать нормативно-правовую документацию в сфере профессиональной деятельности.
Индикатор	УК-2.3. Владеет методиками разработки цели и задач проекта; методами оценки потребности в ресурсах, продолжительности; навыками работы с нормативно-правовой документацией.

**УК-6: Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни**

Индикатор	УК-6.1. Знает основные приемы эффективного управления собственным временем; основные методики самоконтроля, саморазвития и самообразования на протяжении всей жизни.
Индикатор	УК-6.2. Умеет эффективно планировать и контролировать собственное время; использовать методы саморегуляции, саморазвития и самообучения.
Индикатор	УК-6.3. Владеет методами управления собственным временем; технологиями приобретения, использования и обновления социокультурных и профессиональных знаний, умений и навыков; методиками саморазвития и самообразования

**ОПК-2: Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности;**

Индикатор	ОПК-2.1. Знает современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности
Индикатор	ОПК-2.2. Умеет выбирать современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности
Индикатор	ОПК-2.3. Имеет навыки применения современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности

**3. В результате освоения дисциплины (модуля)/практики обучающийся должен**

<b>3.1</b>	<b>Знать:</b>
3.1.1	методы системного анализа и математической статистики
<b>3.2</b>	<b>Уметь:</b>
3.2.1	использовать методы системного анализа и математической статистики для решения социально-экономические задачи
<b>3.3</b>	<b>Владеть:</b>
3.3.1	терминологией и основными методами математической статистики;
3.3.2	навыками применения статистических методов для обработки и анализа больших объемов информации

<b>4. Структура и содержание дисциплины (модуля)/практики</b>
<b>Наименование разделов</b>
<b>Раздел 1. Технологии анализа данных</b>
Большие данные (Big Data): современные подходы к обработке и хранению. Проблема множественного сравнения данных. /Лек/
Процесс анализа. Общая схема анализа. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных. /Лек/
Технологии KDD и Data Mining. Подготовка данных к анализу. Методика извлечения знаний. Data Mining. Мультидисциплинарный характер Data Mining. Причины распространения KDD и Data Mining. Актуальность технологий Data Mining как средств обработки больших объемов информации. /Лек/
Программное обеспечение в области анализа данных. Аналитические платформы: классификация и особенности применения. Языки визуального моделирования. /Лек/
Начало работы. /Пр/
Понятие сценария. /Пр/
Понятие узла обработки. /Пр/
Консолидация данных. /Пр/
Визуализация данных. /Пр/
Трансформация данных. /Пр/
Проработка необходимых вопросов для подготовки к лекциям и практическим занятиям. /Ср/
Подготовка к практическим занятиям /Ср/
Подготовка к лабораторным занятиям /Ср/
<b>Раздел 2. Интеллектуальный анализ данных</b>
Ассоциативные правила. Аффинитивный анализ, предметный набор. Поддержка и достоверность ассоциативного правила. Значимость ассоциативных правил, лифт и левередж. Поиск ассоциативных правил. Частые предметные наборы и их обнаружение. Алгоритм генерации ассоциативных правил. Иерархические ассоциативные правила. Методы поиска иерархических ассоциативных правил. /Лек/
Определение кластеризации. Постановка задачи кластеризации. Цели кластеризации в Data Mining. Примеры кластеризации в различных областях. Виды метрик. Шаги алгоритма. Меры расстояний. Пример работы алгоритма k-means. Проблемы алгоритмов кластеризации. /Лек/
Применение классификации и регрессии. Обзор методов классификации и регрессии. Статистические методы. Методы, основанные на обучении, разнообразие подходов. /Лек/
Основные понятия теории нейронных сетей. Основные парадигмы нейронных сетей. Многослойный перцептрон: класс решаемых задач, архитектура. /Лек/
Определение дерева решений. Причины популярности и условия применимости. Структура дерева решений. Выбор атрибута разбиения в узле. Алгоритм ID3, критерий выбора атрибута разбиения ID3, пример работы алгоритма. Проблема переобучения, Неизвестные значения атрибутов, алгоритм C4.5. /Лек/
Ассоциативные правила. Поиск ассоциативных правил. /Пр/
Кластеризация. Алгоритм кластеризации k-means. /Пр/
Прогнозирование с помощью линейной регрессии. /Пр/
Классификация с помощью нейросети. /Пр/
Классификация с помощью деревьев решений. /Пр/
<b>Раздел 3. Самостоятельная работа</b>
Подготовка к экз /Ср/
<b>Раздел 4. Аттестация</b>
Экзамен /КЭ/

Трудоёмкость: 5 ЗЕ.