

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Гнатюк Максим Александрович
Должность: Первый проректор
Дата подписания: 11.07.2022 09:51:21
Уникальный программный ключ:
8873f497f100e798ae8c92c0d38e105c818d5410

МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА
Федеральное государственное бюджетное образовательное учреждение высшего образования
САМАРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПУТЕЙ СООБЩЕНИЯ

Большие данные

рабочая программа дисциплины (модуля)

Направление подготовки 09.03.02 Информационные системы и технологии
Направленность (профиль) Информационные системы и технологии на транспорте

Квалификация **бакалавр**

Форма обучения **очная**

Общая трудоемкость **5 ЗЕТ**

Виды контроля в семестрах:
экзамены 5

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	5 (3.1)		Итого	
	Неделя 17,7			
Вид занятий	УП	РП	УП	РП
Лекции	18	18	18	18
Практические	36	36	36	36
Конт. ч. на аттест. в	2,35	2,35	2,35	2,35
Итого ауд.	54	54	54	54
Контактная работа	56,35	56,35	56,35	56,35
Сам. работа	90	90	90	90
Часы на контроль	33,65	33,65	33,65	33,65
Итого	180	180	180	180

Программу составил(и):

д.т.н., профессор, Тюгашев А.А.

Рабочая программа дисциплины

Большие данные

разработана в соответствии с ФГОС ВО:

Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 09.03.02 Информационные системы и технологии (приказ Минобрнауки России от 19.09.2017 г. № 926)

составлена на основании учебного плана: 09.03.02-20-12-ИСТб изм.plm.plx

Направление подготовки 09.03.02 Информационные системы и технологии Направленность (профиль) Информационные системы и технологии на транспорте

Рабочая программа одобрена на заседании кафедры

Прикладная математика, информатика и информационные системы

Зав. кафедрой *д.т.н., доцент Тюгашев А.А.* _____

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

1.1	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. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Цикл (раздел) ОП:	Б1.О.14
-------------------	---------

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

ОПК-8 Способен применять математические модели, методы и средства проектирования информационных и автоматизированных систем

ОПК-8.1 Применяет математические модели для проектирования информационных и автоматизированных систем

ОПК-8.3 Использует средства проектирования информационных и автоматизированных систем

В результате освоения дисциплины (модуля) обучающийся должен

3.1 Знать:	
3.1.1	Математические модели для проектирования информационных и автоматизированных систем для работы с большими данными;
3.1.2	Средства проектирования информационных и автоматизированных систем для работы с большими данными.
3.2 Уметь:	
3.2.1	Применять математические модели для проектирования информационных и автоматизированных систем для работы с большими данными;
3.2.2	Использовать средства проектирования информационных и автоматизированных систем для работы с большими данными.
3.3 Владеть:	
3.3.1	Работы с математическими моделями для проектирования информационных и автоматизированных систем для работы с большими данными;
3.3.2	Использования средств проектирования информационных и автоматизированных систем для работы с большими данными.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Примечание
	Раздел 1. Технологии анализа данных			
1.1	Большие данные (Big Data): современные подходы к обработке и хранению. Проблема множественного сравнения данных. /Лек/	5	2	
1.2	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/	5	2	
1.3	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/	5	2	
1.4	Программное обеспечение в области анализа данных. Аналитические платформы: классификация и особенности применения. Языки визуального моделирования. /Лек/	5	2	
1.5	Начало работы. /Пр/	5	3	
1.6	Понятие сценария. /Пр/	5	3	
1.7	Понятие узла обработки. /Пр/	5	4	
1.8	Консолидация данных. /Пр/	5	3	

1.9	Визуализация данных. /Пр/	5	4	
1.10	Трансформация данных. /Пр/	5	3	
1.11	Проработка необходимых вопросов для подготовки к лекциям и практическим занятиям. /Ср/	5	15	
1.12	Подготовка к правтическим занятиям /Ср/	5	11	
1.13	Подготовка к лабораторным занятиям /Ср/	5	14	
Раздел 2. Интеллектуальный анализ данных				
2.1	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/	5	2	
2.2	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/	5	2	
2.3	Применение классификации и регрессии. Обзор методов классификации и регрессии. Статистические методы. Методы, основанные на обучении, разнообразии подходов. /Лек/	5	2	
2.4	Основные понятия теории нейронных сетей. Основные парадигмы нейронных сетей. Многослойный перцептрон: класс решаемых задач, архитектура. /Лек/	5	2	
2.5	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/	5	2	
2.6	Ассоциативные правила. Поиск ассоциативных правил. /Пр/	5	4	
2.7	Кластеризация. Алгоритм кластеризации k-means. /Пр/	5	3	
2.8	Прогнозирование с помощью линейной регрессии. /Пр/	5	3	
2.9	Классификация с помощью нейросети. /Пр/	5	3	
2.10	Классификация с помощью деревьев решений. /Пр/	5	3	
Раздел 3. Самостоятельная работа				
3.1	Подготовка к экз /Ср/	5	50	
Раздел 4. Аттестация				
4.1	Экзамен /КЭ/	5	2,35	

5. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

Оценочные материалы для проведения промежуточной аттестации обучающихся приведены в приложении к рабочей программе дисциплины.

Формы и виды текущего контроля по дисциплине (модулю), виды заданий, критерии их оценивания, распределение баллов по видам текущего контроля разрабатываются преподавателем дисциплины с учетом ее специфики и доводятся до сведения обучающихся на первом учебном занятии.

Текущий контроль успеваемости осуществляется преподавателем дисциплины (модуля), как правило, с использованием ЭИОС или путем проверки письменных работ, предусмотренных рабочими программами дисциплин в рамках контактной работы и самостоятельной работы обучающихся. Для фиксирования результатов текущего контроля может использоваться ЭИОС.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

6.1. Рекомендуемая литература

6.1.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год	Эл. адрес
--	---------------------	----------	-------------------	-----------

Л1.1	Советов Б. Я., Цехановский В. В., Чертовской В. Д.	Базы данных: учебник для прикладного бакалавриата	Москва: Юрайт, 2015	
6.1.2. Дополнительная литература				
	Авторы, составители	Заглавие	Издательс тво, год	Эл. адрес
Л2.1	Кумскова И.А.	Базы данных.	Москва: КноРус, 2020	http://www.book.ru/book/932493
6.2 Информационные технологии, используемые при осуществлении образовательного процесса по дисциплине (модулю)				
6.2.1 Перечень лицензионного и свободно распространяемого программного обеспечения				
6.2.1.1	Операционная система Microsoft® Windows Professional 8 Russian Upgrade OLP NL Academic Edition Договор на поставку № 0342100004813000011 от года.			
6.2.1.2	Microsoft Office 2013 Professional Договор № 0342100004814000045			
6.2.2 Перечень профессиональных баз данных и информационных справочных систем				
6.2.2.1	База книг и публикаций Электронной библиотеки "Наука и Техника" - http://www.n-t.ru			
7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)				
7.1	Учебные аудитории для проведения занятий лекционного типа, укомплектованные специализированной мебелью и техническими средствами обучения: мультимедийное оборудование для предоставления учебной информации большой аудитории и/или звукоусиливающее оборудование (стационарное или переносное).			
7.2	Учебные аудитории для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные специализированной мебелью и техническими средствами обучения: мультимедийное оборудование и/или звукоусиливающее оборудование (стационарное или переносное)			
7.3	Помещения для самостоятельной работы, оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду университета.			
7.4	Помещения для хранения и профилактического обслуживания учебного оборудования			