

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН ДИСЦИПЛИНЫ
2020/2021 учебный год

ОЦЕНКИ			<i>«Обработка больших объемов данных»</i> для студентов 1 курса (магистр) ИЯТШ по направлению 01.04.02 Прикладная математика и информатика	Лекции	8	час.
«Отлично»	A	90 - 100 баллов		Практ. занятия	16	час.
«Хорошо»	B	80 – 89 баллов		Лаб. занятия	24	час.
	C	70 – 79 баллов		Всего ауд. работа	48	час.
«Удовл.»	D	65 –69 баллов		СРС	60	час.
	E	55 –64 баллов				
Зачтено	P	55 - 100 баллов		ИТОГО	108	час.
Неудовлетвори тельно / незачтено	F	0 - 54 баллов			3	з.е.

Результаты обучения по дисциплине:

РД1	Выполнять исследования процессов создания, накопления и обработки информации, включая анализ и создание моделей данных и знаний, языков их описания и манипулирования.
РД2	Владеть методами исследования и обработки данных и их применению в самостоятельной научно-исследовательской и профессиональной деятельности.
РД3	Владение методами и инструментами визуализации и ординации многомерных объектов

Оценочные мероприятия:

Для дисциплин с формой контроля – зачет

Оценочные мероприятия		Кол-во	Баллы
Текущий контроль:			
П	Посещение занятий	48	-
ТК1	Защита практических работ	8	24
ТК2	Защита отчета по лабораторной работе	12	24
КР	Контрольная работа	1	20
ЭС	Эссе по СРС	12	12
ПА1	Зачет	1	20
ИТОГО			100

Неделя	Дата начала недели	Результат обучения по дисциплине	Учебная деятельность	Кол-во часов		Оценочное мероприятие	Кол-во баллов	Информационное обеспечение		
				Ауд.	Сам.			Учебная Литература	Интернет-ресурсы	Видео-ресурсы
1	2	3	4	5	6	7	8	9	10	11
1 24	10.02	РД1	Раздел 1. Стратегии работы с большими объемами данных							-
			Лекция 1. Введение в большие данные. Стратегии работы с большими объемами данных	2		П				
			ЛР 1. Функции чтения/записи данных из/в текстовый файл.	2		ТК2	2		ЭК1	
			СРС: Структурированный язык запросов		5	ЭС	1		ЭК1	
2 25	17.02	РД1	ЛР 2. Справочная система R/Python.	2		ТК2	2		ЭК1	
			Пр. занятие 1. Языки, программные среды, фреймворки для обработки данных.	2		ТК2	3	ОСН1	ЭК1	
			СРС: Подход Not Only SQL		5	ЭС	1		ОСН1	
3 26	24.02	РД2	Лекция 2. Базовые алгоритмы интеллектуального анализа данных (Data mining). Подготовка исходных данных к обработке.	2		П				
			ЛР 3. Стратегии работы с большими массивами данных.	2		ТК2	2	ОСН2	ЭК1	
			СРС: Локальный Spark-кластер		5	ЭС	1	ОСН1	ЭК3	
4 27	02.03	РД2	ЛР 4. Кэширование вычислений.	2		ТК2	2	ОСН1		
			Пр. занятие 2. Платформы для выполнения распределенных вычислений.	2		ТК2	3	ОСН1	ЭК1	
			СРС: Установка и использование Java Virtual Machine		5	ЭС	1		ЭК3	
5 28	09.03	РД2	Раздел 2. Обработка больших объемов данных							
			Лекция 3. Выбор признаков (Feature Selection), экземпляров (Instance Selection), дискретизация для классификации (Discretization).	2		П		ДОП1	ЭК1	
			ЛР 5. Обработка пропусков в данных.	2		ТК2	2			
6 29	16.03	РД2, РД3	СРС: Установка и использование Hadoop		5	ЭС	1			
			ЛР 6. Графический способ определения выбросов.	2		ТК2	2			
			Пр. занятие 3. Графический разведочный анализ данных.	2		ТК2	3	ОСН1		
7 30	23.03	РД2	СРС: Бутстреп-процедура		5	ЭС	1	ДОП3	ЭК1	
			Раздел 3. Визуализация и ординация многомерных объектов							
			Лекция 4. Алгоритмы сжатия информационного пространства	2		П		ДОП4	ЭК1	
8 31	30.03	РД2	ЛР 7. Преобразование и дискретизация данных.	2		ТК2	2	ОСН2		
			СРС: Функции графических пакетов matplotlib		5	ЭС	1			
			ЛР 8. Формирование обучающей и тестовой выборки.	2		ТК2	2	ДОП5		
		РД2	Пр. занятие 4. Статистические методы заполнения пропусков и определения выбросов.	2		ТК2	3	ОСН1	ЭК1	
			СРС: подготовка к контрольной работе		5				ЭК1	
9 32	06.04		Конференц-неделя 1							
			Контрольная работа				20			
			Всего по контрольной точке (аттестации) 1	32	40		48			
10 33	13.04	РД2, РД3	ЛР 9. Системы визуализации ggplot2, GGVIS.	2		ТК2	2	ОСН1	ЭК1	
			Ординационные диаграммы.							
			СРС: Статистическое оценивание качества модели		2	ЭС	1	ОСН5	ЭК1	
11 34	20.04	РД2	ЛР 10. Метод главных компонент, координат.	2		ТК2	2	ОСН3	ЭК2	
			СРС: Метрики качества кластеризации		2	ЭС	1	ДОП1		
12 35	27.04	РД2	Пр. занятие 5. Определение важности признаков.	2		ТК2	3	ДОП4		
			СРС: Использование GitHub для коллективной работы		2	ЭС				
13 36	04.05	РД2, РД3	Пр. занятие 6. Классификация и регрессия. Линейные, нелинейные модели.	2		ТК2	3	ОСН3	ЭК1	
			СРС: Оценка качества регрессионных моделей		2	ЭС	1	ДОП5	ЭК1	
14 37	11.05	РД2	ЛР 11. Кластерный анализ. Дендограммы.	2		ТК2	2	ОСН2		
			СРС: Типы измерительных шкал		2	ЭС	1	ОСН5		
15 38	18.05	РД2, РД3	Пр. занятие 7. Методы визуализации многомерных объектов	2		ТК2	3	ОСН1		
			СРС: Нормализация, стандартизация данных		2	ЭС	1	ДОП2		
16 39	25.05	РД2	ЛР 12. Неметрическое многомерное шкалирование.	2		ТК2	2	ОСН1		
			СРС: Эвристические алгоритмы перебора для кластеризации		2	ЭС	1	ОСН 5		
17 40	01.06	РД2, РД3	Пр. занятие 8. Методы ординации многомерных объектов.	2		ТК2	3	ОСН2		
			СРС: Подготовка к зачету		6			все		
18 41	08.06		Конференц-неделя 2							
			Зачет				20			
			Всего по контрольной точке (аттестации) 2	8	20		100			
			Общий объем работы по дисциплине	40	60		100			

Информационное обеспечение:

№ (код)	Основная учебная литература (ОСН)	№ (код)	Название электронного ресурса (ЭР)	Адрес ресурса
---------	-----------------------------------	---------	------------------------------------	---------------

ОСН 1	Кабаков Р.И. R в действии. Анализ и визуализация данных в программе R. Москва: ДМК Пресс, 2014. — 588 с. // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/58703	ЭК1	Курс «Big Data»	https://www.coursera.org/specializations/big-data
ОСН 2	Буховец А. Г. Алгоритмы вычислительной статистики в системе R: учебное пособие / А. Г. Буховец, П. В. Москалев. — 2-е изд., перераб. и доп. — Санкт-Петербург : Лань, 2015. — 160 с. // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/68459	ЭК2	Конференция по большим данным и искусственному интеллекту	https://bigdatadays.ru/
ОСН 3	Введение в статистическое обучение с примерами на языке R / Г. Джеймс, Д. Уиттон, Т. Хастингс, Р. Тибширани ; перевод с английского С. Э. Мостицкого. — Москва : ДМК Пресс, 2017. — 456 с. // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/93580	ЭК3	Локальный Spark-кластер	https://r-analytics.blogspot.com/2020/02/spark-r-connect.html
ОСН 4	Москвитин, А. А.. Данные, информация, знания: методология, теория, технологии : монография [Электронный ресурс] / Москвитин А. А.. — Санкт-Петербург: Лань, 2019. — 236 с.			
ОСН 5	Шитиков В.К., Мостицкий С.Э. (2017) Классификация, регрессия и другие алгоритмы Data Mining с использованием R. 351 с. – Электронная книга, адрес доступа: https://github.com/ranalytics/data-mining			
№ (код)	Дополнительная учебная литература (ДОП)			
ДОП 1	Olvera-López J., Carrasco-Ochoa J. Martínez-Trinidad J. F. and Kittler J. (2010). A review of instance selection methods. Artif. Intell. Rev. 34. 133-143.			
ДОП 2	X. Wu et al. (2008) Top 10 algorithms in data mining. Knowl. Inf. Syst. 14. 1–37.			
ДОП 3	Лесковец, Юре. Анализ больших наборов данных : пер. с англ. / Ю. Лесковец, А. Раджараман, Дж. Ульман. — Москва: ДМК Пресс, 2016. — 498 с.			
ДОП 4	Фрэнкс, Билл. Укрощение больших данных. Как извлекать знания из массивов информации с помощью глубокой аналитики : пер. с англ. / Б. Фрэнкс. — Москва: Манн, Иванов и Фербер, 2014. — 340 с.			
ДОП 5	Хименко, Виталий Иванович. Случайные данные: структура и анализ: учебник / В. И. Хименко. — Москва: Техносфера, 2019. — 424 с.			

Составил: _____ (Семенов М.Е.)
«31» 08 2020 г.

Согласовано:
Заведующий кафедрой –
руководитель отделения (на правах кафедры) _____ (Лидер А.М.)
«31» 08 2020 г.