

Министерство науки и высшего образования РФ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б1.О.02 Методы и алгоритмы распознавания и обработки
данных

наименование дисциплины (модуля) в соответствии с учебным планом

Направление подготовки / специальность

09.04.04 Программная инженерия

Направленность (профиль)

09.04.04.02 Технологии индустриального производства программного
обеспечения интеллектуальных систем управления

Форма обучения

очная

Год набора

2021

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Программу составили _____

канд. техн. наук, доцент, Тынченко В.В.

должность, инициалы, фамилия

1 Цели и задачи изучения дисциплины

1.1 Цель преподавания дисциплины

Обучение магистрантов современным методам и алгоритмам решения интеллектуальных задач распознавания и обработки данных, а также их практическому применению в сфере индустриального производства программного обеспечения интеллектуальных систем управления.

1.2 Задачи изучения дисциплины

1. Освоить теоретические основы технологии Data Mining и возможности ее применения для распознавания и обработки данных при решении задач в сфере индустриального производства программного обеспечения интеллектуальных систем управления.

2. Приобрести умения по выбору, настройке и применению эффективных методов и алгоритмов Data Mining для решения различных классов интеллектуальных задач распознавания и обработки данных.

3. Овладеть практическими навыками применения модельно-алгоритмического и программного инструментария Data Mining для распознавания и обработки данных при решении широкого круга задач классификации, кластеризации, прогнозирования, аппроксимации, ассоциации и пр.

1.3 Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Код и наименование индикатора достижения компетенции	Запланированные результаты обучения по дисциплине
ОПК-7: Способен применять при решении профессиональных задач методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе, в глобальных компьютерных сетях;	
ОПК-7.1: Знать методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе, в глобальных компьютерных сетях	Знает методы, алгоритмы, программный инструментарий технологии Data Mining для решения задач распознавания и обработки данных в сфере индустриального производства программного обеспечения интеллектуальных систем управления Умеет применять методы, алгоритмы, программный инструментарий технологии Data Mining для решения задач распознавания и обработки данных в сфере программной инженерии и кибернетики Владеет навыками применения методов, алгоритмов, программного инструментария технологии Data Mining для решения задач распознавания и обработки данных в сфере программной инженерии и кибернетики

ОПК-7.2: Уметь применять методы и средства получения, хранения, переработки и трансляции информации	Умеет применять методы, алгоритмы, программный инструментарий технологии Data Mining для решения задач распознавания и обработки данных в сфере индустриального производства программного
посредством современных компьютерных технологий, в том числе, в глобальных компьютерных сетях	обеспечения интеллектуальных систем управления
ОПК-7.3: Иметь навыки методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе, в глобальных компьютерных сетях	Владеет практическими навыками применения методов, алгоритмов, программного инструментария технологии Data Mining для решения задач распознавания и обработки данных в сфере индустриального производства программного обеспечения интеллектуальных систем управления

1.4 Особенности реализации дисциплины

Язык реализации дисциплины: Русский.

Дисциплина (модуль) реализуется с применением ЭО и ДОТ

URL-адрес и название электронного обучающего курса: <https://e.sfu-kras.ru/course/view.php?id=8026>.

2. Объем дисциплины (модуля)

Вид учебной работы	Всего, зачетных единиц (акад.час)	е
		1
Контактная работа с преподавателем:	1,5 (54)	
занятия лекционного типа	0,5 (18)	
практические занятия	1 (36)	
Самостоятельная работа обучающихся:	2,5 (90)	
курсовое проектирование (КП)	Нет	
курсовая работа (КР)	Нет	
Промежуточная аттестация (Экзамен)	1 (36)	

3 Содержание дисциплины (модуля)

3.1 Разделы дисциплины и виды занятий (тематический план занятий)

		Контактная работа, ак. час.							
№ п/п	Модули, темы (разделы) дисциплины	Занятия лекционного типа		Занятия семинарского типа				Самостоятельная работа, ак. час.	
				Семинары и/или Практические занятия		Лабораторные работы и/или Практикумы			
		Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС	Всего	В том числе в ЭИОС
1. Технологии анализа данных									
	1. Введение в анализ данных. Принципы анализа данных. Структурированные данные. Подготовка данных к анализу. Технология Data Mining.	1							
	2. Фильтрация данных. Обработка дубликатов и противоречий. Выявление аномальных значений. Восстановление пропущенных значений.			2					
	3. Самостоятельное изучение теоретического материала по теме "Технология KDD". Подготовка отчета по практической работе, подготовка к защите практической работы.							8	8
2. Консолидация данных									

1. Задачи консолидации. Введение в хранилища данных. Основные концепции хранилищ данных. Многомерные хранилища данных. Реляционные хранилища данных. Гибридные хранилища данных. Виртуальные хранилища данных.	1								
2. Самостоятельное изучение теоретического материала по теме "Введение в ETL".							8	8	
3. Трансформация данных									
1. Введение в трансформацию данных. Трансформация упорядоченных данных. Группировка данных. Слияние данных. Квантование. Нормализация и кодирование данных.	2								
2. Квантование. Нормализация и кодирование данных.			4						
3. Самостоятельное углубленное изучение теоретического материала по темам раздела. Подготовка отчета по практической работе, подготовка к защите практической работы.							8	8	
4. Визуализация данных									
1. Введение в визуализацию. Визуализаторы общего назначения. OLAP-анализ. Визуализаторы для оценки качества моделей. Визуализаторы для интерпретации результатов анализа.	2								
2. Самостоятельное углубленное изучение теоретического материала по темам раздела.							8	8	
5. Очистка и преобработка данных									
1. Оценка качества данных. Очистка и преобработка. Фильтрация данных. Восстановление пропущенных значений. Введение в сокращение размерности. Сокращение числа признаков. Сэмплинг.	2								

2. Сокращение размерности. Сокращение числа признаков. Сэмплинг.			2						
3. Самостоятельное изучение теоретического материала по темам "Обработка дубликатов и противоречий", "Выявление аномальных значений". Подготовка отчета по практической работе, подготовка к защите практической работы.							10	10	
6. Задачи ассоциации									
1. Ассоциативные правила. Алгоритм a priori. Иерархические ассоциативные правила.	2								
2. Алгоритм a priori. Иерархические ассоциативные правила.			4						
3. Самостоятельное углубленное изучение теоретического материала по темам раздела. Подготовка отчета по практической работе, подготовка к защите практической работы.							10	10	
7. Задачи кластеризации									
1. Введение в кластеризацию. Алгоритм кластеризации k-means. Карты Кохонена. Проблема алгоритмов кластеризации.	2								
2. Алгоритм кластеризации k-means.			4						
3. Карты Кохонена.			4						
4. Самостоятельное углубленное изучение теоретического материала по темам раздела. Подготовка отчета по практической работе, подготовка к защите практической работы.							10	10	
8. Классификация и регрессия									
1. Линейная регрессия. Регрессия с категориальными входными переменными. Логистическая регрессия.	1								

2. Деревья решений. Алгоритмы ID3, C4.5, CART. Нейронные сети.	1								
3. Линейная регрессия. Регрессия с категориальными входными переменными. Логистическая регрессия.			4						
4. Деревья решений. Алгоритмы ID3, C4.5, CART. Нейронные сети.			4						
5. Самостоятельное углубленное изучение теоретического материала по темам раздела. Подготовка отчета по практической работе, подготовка к защите практической работы.							10	10	
9. Анализ и прогнозирование временных рядов									
1. Введение в прогнозирование. Временной ряд и его компоненты. Методы прогнозирования.	2								
2. Методы прогнозирования.			4						
3. Самостоятельное углубленное изучение теоретического материала по темам раздела. Подготовка отчета по практической работе, подготовка к защите практической работы.							12	12	
10. Ансамбли моделей									
1. Введение в ансамбли моделей. Бэггинг. Бустинг. Альтернативные методы построения ансамблей. Сравнение моделей.	2								
2. Ансамбли моделей. Бэггинг. Бустинг.			4						
3. Самостоятельное углубленное изучение теоретического материала по темам раздела. Подготовка отчета по практической работе, подготовка к защите практической работы.							6	6	
4.									

Bcero	18		36				90	90
-------	----	--	----	--	--	--	----	----

4 Учебно-методическое обеспечение дисциплины

4.1 Печатные и электронные издания:

1. Рубан А. И. Методы анализа данных: учебное пособие(Красноярск: СФУ).
2. Системный анализ, оптимизация и принятие решений.: Учебник. (Москва: ООО "КУРС").
3. Барский А.Б. Нейронные сети: распознавание, управление, принятие решений(М.: Финансы и статистика).
4. Маккинли У. Python и анализ данных(Москва: ДМК Пресс).
5. Семенкин Е. С., Семенкина О. Э., Антамошкин А. Н., Терсков В. А., Тынченко В. В. Методы оптимизации: электрон. учеб.-метод. комплекс дисциплины(Красноярск: ИПК СФУ).
6. Рубан А. И., Кузнецов А. В. Методы анализа данных: электрон. учеб.-метод. комплекс дисциплины(Красноярск: ИПК СФУ).

4.2 Лицензионное и свободно распространяемое программное обеспечение, в том числе отечественного производства (программное обеспечение, на которое университет имеет лицензию, а также свободно распространяемое программное обеспечение):

1. Deductor Academic, Microsoft Visual Studio, Microsoft Office Professional Plus

4.3 Интернет-ресурсы, включая профессиональные базы данных и информационные справочные системы:

1. Не требуется.

5 Фонд оценочных средств

Оценочные средства находятся в приложении к рабочим программам дисциплин.

6 Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине (модулю)

Учебные аудитории для проведения лекционных и лабораторных занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, содержащие специализированную мебель, компьютеры с подключением к сети Интернет (неограниченный доступ) и обеспечением доступа к системе виртуальных машин; демонстрационное оборудование(интерактивная доска обратной проекции, проектор, экран для проектора), маркерная доска, доступ к беспроводной сети WI-FI, а также помещение для самостоятельной работы оснащенное компьютерами с подключением к сети Интернет (неограниченный доступ) и обеспечением доступа в электронную информационно-образовательную среду СФУ.