

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

Б1.В.03 Продвинутые методы анализа данных

наименование дисциплины (модуля) в соответствии с учебным планом

Направление подготовки / специальность

01.04.02 Прикладная математика и информатика

Направленность (профиль)

01.04.02.08 Анализ данных и математическое моделирование

Форма обучения

очная

Год набора

2021

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Программу составили _____

к.т.н., Доцент, Кузенков Н.П.

должность, инициалы, фамилия

1 Цели и задачи изучения дисциплины

1.1 Цель преподавания дисциплины

Целью преподавания дисциплины является ознакомление студентов с широким спектром методов машинного обучения с учителем и без учителя, анализа временных рядов и представления (визуализации) результатов их применения.

1.2 Задачи изучения дисциплины

Задачами изучения дисциплины являются:

- Ознакомление студентов с разнообразными методами решения задач анализа данных, включая анализ временных рядов, достоинствами и недостатками данных методов, способам выбора оптимальных методов для решения поставленной задачи.

- Выработка умений постановки задач анализа данных различных типов и выбора наиболее подходящих методов решения поставленной задачи.

- Получение навыков работы с экспериментальными данными и навыков использования и создания алгоритмов анализа данных.

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
ПК-4: Способен разрабатывать и внедрять новые методы и технологии исследования данных.	
ПК-4.1: Знает: состояние и перспективы развития информационных технологий, технологий данных в России и в мире; современные и перспективные методы сбора, хранения и передачи данных; источники данных, интенсивность генерации данных источниками; технические средства и среды сбора, хранения и обработки данных; современные и перспективные средства визуализации и интерпретации данных; исследование операций; машинное обучение; математическое моделирование; методы сравнительного анализа.	Основные принципы работы в средах, используемых для анализа данных и машинного обучения. Использовать среды анализа данных и машинного обучения. Навыками программирования в средах, используемых для анализа данных и машинного обучения.

ПК-4.2: Способен проводить аналитические и поисковые	Методы генерации знаний из наборов данных. Производить новые знания, используя наборы
исследования по тематике информационных технологий, технологий данных.	данных и определять область их применения. Навыками генерации знаний из наборов данных.

<p>ПК-4.3: Знает: состояние и перспективы развития информационных технологий, технологий данных в России и в мире; существующие и перспективные методы и программный инструментарий технологий данных; существующие и перспективные методы и программный инструментарий технологий данных; предметная область использования технологий больших данных; современный опыт использования технологий данных; математическое моделирование; методы сравнительного анализа методов, моделей, алгоритмов, технологий и инструментальных средств; существующие и перспективные математические методы и инструментальные средства анализа данных; существующие и перспективные методы обработки данных; методы машинного обучения; современные и перспективные методы сбора данных; источники данных, интенсивность генерации данных источниками; технические средства и среды сбора, хранения и обработки данных; современные и перспективные средства визуализации и интерпретации данных; применение технологий данных в практических</p>	<p>Основные среды, используемые для анализа данных и машинного обучения, принципы работы с ними и особенности их использования. Определять современные тенденции развития инструментов анализа данных. Навыками анализа используемых инструментов.</p>
<p>задачах предметной области; показатели эффективности технологий данных.</p>	

ПК-4.4: Способен проводить аналитические исследования	Методологию научного познания. Производить новые знания, используя наборы
по тематике информационных технологий, технологий данных; выполнять научные исследования.	данных и определять область их применения. Навыками научных и аналитических исследований.
УК-1: Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий.	
УК-1.1: Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними.	Основные типы задач анализа данных и этапы их решения. Осуществлять декомпозицию задач анализа данных. Навыками алгоритмизации этапов решения задачи анализа данных.
УК-1.2: Определяет пробелы в информации, необходимой для решения проблемной ситуации, и проектирует процессы по их устранению.	Основные характеристики выборочных данных. Нормализовывать данные, определять наличие шума, пропущенных данных и выбросов. Навыками выполнения описательного анализа данных.
УК-1.3: Критически оценивает надежность источников информации, работает с противоречивой информацией из разных источников.	Основные типы задач машинного обучения с учителем и без учителя и методы их решения, а также основные допущения к применению каждого из них. Производить выбор наиболее подходящих методов анализа. Навыками программных вычислений и визуализации основных результатов работы примененного метода.
УК-1.4: Разрабатывает и содержательно аргументирует стратегию решения проблемной ситуации на основе системного и междисциплинарных подходов.	Характерные методологические ошибки при решении задач анализа данных. Анализирует проблемную ситуацию и выбирает подходящие к ней методы решения Навыками планирования способов решения задач.
УК-1.5: Использует логико-методологический инструментарий для критической оценки современных концепций философского и социального характера в своей предметной области.	Принципы и ограничения переноса результатов анализа исследованных данных на новые данные. Интерпретировать результаты проведенного анализа на языке предметной области. Навыками определения ценности информации, полученной в результате анализа данных.
УК-2: Способен управлять проектом на всех этапах его жизненного цикла.	
УК-2.1: Формулирует на основе поставленной проблемы проектную задачу и способ ее решения через реализацию проектного управления.	Способы формализации поставленной задачи. Делить процесс решения на отдельные этапы. Навыками синтеза результатов и коррекции процесса решения задачи.

УК-2.2: Разрабатывает концепцию проекта в рамках обо-значенной проблемы: формулирует цель, задачи, обосновывает актуальность, значимость, ожидаемые результаты и возможные сферы их применения.	Методологию решения задач анализа данных. Методологически верно решать задачи анализа данных. Навыками критического оценивания результатов решения поставленных задач.
УК-2.3: Планирует необходимые ресурсы, в том числе с учетом их заменимости.	Необходимые ресурсы для решения поставленной задачи анализа данных. Выбирать необходимые ресурсы для решения задач из имеющихся в наличии. Навыками решения задач анализа данных с использованием имеющихся в наличии ресурсов.
УК-2.4: Разрабатывает план реализации проекта.	Этапы решения задач анализа данных. Делить процесс решения задач анализа данных на этапы с учётом имеющихся ресурсов. Навыками решения задач анализа данных с использованием имеющихся в наличии ресурсов.
УК-2.5: Осуществляет мониторинг хода реализации проекта, корректирует отклонения, вносит дополнительные изменения в план реализации проекта, уточняет зоны ответственности участников проекта.	Принципы и этапы решения задач анализа данных. Управлять развитием проекта по анализу данных и участниками этого проекта. Навыками декомпозиции задач анализа данных и постановки подзадач участникам проекта.

1.4 Особенности реализации дисциплины

Язык реализации дисциплины: Русский.

Дисциплина (модуль) реализуется без применения ЭО и ДОТ.

2. Объем дисциплины (модуля)

Вид учебной работы	Всего, зачетных единиц (акад.час)	е
		1
Контактная работа с преподавателем:	2 (72)	
занятия лекционного типа	1 (36)	
практические занятия	1 (36)	
Самостоятельная работа обучающихся:	2 (72)	
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. Введение									
	1. Введение	2							
	2. Визуализация многомерных зависимых данных			2					
	3. Визуализация многомерных зависимых данных							4	
2. Кластеризация									
	1. Семейство алгоритмов FOREL	2							
	2. Иерархическая кластеризация			2					
	3. Иерархическая кластеризация							4	
	4. Алгоритмы K-means и C-means	2							
	5. K-means и C-means кластеризация			2					
	6. K-means и C-means кластеризация							4	
	7. EM-алгоритм	2							
	8. GMM-модель			2					
	9. GMM-модель							4	

10. Алгоритм DBSCAN	2							
11. Кластеризация замкнутых таксонов			2					
12. Кластеризация замкнутых таксонов							4	
13. Спектральная кластеризация	2							
14. Спектральная кластеризация			2					
15. Спектральная кластеризация							4	
3. Классификация								
1. Метод k-ближайших соседей	2							
2. Метод k-ближайших соседей			2					
3. Метод k-ближайших соседей							4	
4. Наивный байесовский классификатор	2							
5. Наивный байесовский классификатор			2					
6. Наивный байесовский классификатор							4	
7. Классификационные и регрессионные деревья	4							
8. Классификационные и регрессионные деревья			4					
9. Классификационные и регрессионные деревья							8	
10. Ансамблевые методы	2							
11. Ансамблевые методы			2					
12. Ансамблевые методы							4	
4. Временные ряды								
1. Классические методы анализа и представления временных рядов	2							
2. Спектральное представление временного ряда			2					
3. Спектральное представление временного ряда							4	
4. Модель ARIMA	2							
5. Модель ARIMA			2					

6. Модель ARIMA							4	
7. Модель GARCH	2							
8. Модель GARCH			2					
9. Модель GARCH							4	
10. Корреляционная размерность	2							
11. Корреляционная размерность			2					
12. Корреляционная размерность							4	
13. Метод ложных ближайших соседей	2							
14. Метод ложных ближайших соседей			2					
15. Метод ложных ближайших соседей							4	
5. Нейроинформатика								
1. Искусственные нейронные сети	2							
2. Классификация с помощью ИНС			2					
3. Классификация с помощью ИНС							4	
4. Метод упругих карт	2							
5. Метод упругих карт			2					
6. Метод упругих карт							4	
Всего	36		36				72	

4 Учебно-методическое обеспечение дисциплины

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. MATLAB, Python

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

1. Система электронного обучения Сибирского федерального университета (<https://e.sfu-kras.ru>)
2. электронные информационно-справочные ресурсы научной библиотеки СФУ (<http://bik.sfu-kras.ru>)
3. Интернет-репозиторий статистических данных Data-Planet (<https://www.data-planet.com/>)
4. Интернет-репозиторий статистических данных на платформе kaggle.com. Режим доступа: <https://www.kaggle.com/datasets>

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 лаборатории и классы, оснащенные современными компьютерами, объединенными в локальные вычислительные сети с выходом в Интернет, а также периферийным и проекционным оборудованием.