

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

Б1.В.02 Базовые модели и методы машинного обучения

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 _____

должность, инициалы, фамилия

1 Цели и задачи изучения дисциплины

1.1 Цель преподавания дисциплины

Целью изучения дисциплины является ознакомление студентов с основными методами и моделями анализа данных.

1.2 Задачи изучения дисциплины

Задачами изучения дисциплины являются:

- Ознакомление с задачами обучения на размеченных данных, основными линейными моделями, методами оценки качества прогнозирования, возможными причинами низкого качества моделей;
- Выработка умений постановки задач анализа данных, обоснованного выбора и применения изученных методов в прикладных исследованиях
- Получение практических навыков программирования алгоритмов машинного обучения средствами языка R.

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.	Знать современные и перспективные средства визуализации данных, возможности использования таких визуализаций для выдвижения гипотез о взаимосвязях в данных; линейные методы прогнозирования и возможности их обобщения; особенности интерпретации результатов прогнозирования с помощью линейных алгоритмов.

ПК-4.2: Способен проводить аналитические и поисковые исследования по тематике информационных технологий, технологий данных.	Способен анализировать качество моделей прогнозирования, осуществлять подбор гиперпараметров моделей в целях улучшения качества прогноза, способен производить отбор предикторов моделей
---	--

<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 задачах предметной области;</p>	<p>Знает способы сравнительного анализа моделей машинного обучения</p>
<p>показатели эффективности технологий данных.</p>	

ПК-4.4: Способен проводить аналитические исследования	Способен формулировать задачи анализа данных, возникающие в научных исследованиях, как задачи
по тематике информационных технологий, технологий данных; выполнять научные исследования.	машинного обучения, выбирать релевантные методы машинного обучения для решения этих задач
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.	Использует методы статистического анализа для выявления причинно-следственных связей и зависимостей
УК-1.2: Определяет пробелы в информации, необходимой для решения проблемной ситуации, и проектирует процессы по их устранению.	Способен обрабатывать пропуски в данных математическими методами, определять возможность и эффективность использования методов заполнения пропусков
УК-1.3: Критически оценивает надежность источников информации, работает с противоречивой информацией из разных источников.	Способен стандартизировать данные, поступающие из разных источников с учетом применяемых шкал и единиц изменения. Использует статистические методы выявления противоречивой информации.
УК-1.4: Разрабатывает и содержательно аргументирует стратегию решения проблемной ситуации на основе системного и междисциплинарных подходов.	Учитывает особенности применимости методов машинного обучения к данным разной природы, учитывает известные причинно-следственные связи в предметных областях или фундаментальные законы, на основании этого разрабатывает стратегию проведения исследований и решения возникающих проблем
УК-1.5: Использует логико-методологический инструментарий для критической оценки современных концепций философского и социального характера в своей предметной области.	
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.	Формулирует задачи предметной области как задачи машинного обучения, способен интерпретировать ожидаемые результаты решения задачи машинного обучения на языке предметной области, определять границы применимости результатов исходя из математических обоснований применяемых методов машинного обучения

УК-2.2: Разрабатывает концепцию проекта в рамках	Формулирует задачи предметной области как задачи машинного обучения, способен интерпретировать
обо-значенной проблемы: формулирует цель, задачи, обосновывает актуальность, значимость, ожидаемые результаты и возможные сферы их применения.	ожидаемые результаты решения задачи машинного обучения на языке предметной области, определять границы применимости результатов исходя из математических обоснований применяемых методов машинного обучения
УК-2.3: Планирует необходимые ресурсы, в том числе с учетом их заменимости.	Планирует время на предварительную подготовку данных, проведение аналитической работы, а также технические ресурсы с учетом возможных проблем с достаточностью и зашумленностью сырых данных, вычислительными затратами на обучение и отладку моделей машинного обучения
УК-2.4: Разрабатывает план реализации проекта.	Встраивает все необходимые этапы решения входящей в проект задачи машинного обучения в общий план реализации проекта
УК-2.5: Осуществляет мониторинг хода реализации проекта, корректирует отклонения, вносит дополнительные изменения в план реализации проекта, уточняет зоны ответственности участников проекта.	Корректирует процесс обучения моделей прогнозирования (меняет или модифицирует используемые методы, меняет состав предикторов моделей), учитывая результаты предварительного анализа данных

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=21687>.

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:	3 (108)	
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)

№ п/п		Модули, темы (разделы) дисциплины		Контактная работа, ак. час.							
				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. Основные компоненты машинного обучения											
		1. Основные компоненты машинного обучения		2							
		2. Основы программирования на языке R				2					
		3. Предобработка данных		2							
		4. Предобработка данных. Описательная статистика и визуальный анализ данных				2					
		5. Основные компоненты машинного обучения							14		
2. Линейные методы регрессии											
		1. Вероятностное обоснование модели регрессии		2							
		2. Множественная линейная регрессия методом наименьших квадратов				4					
		3. Классическая модель множественной линейной регрессии		4							
		4. Метрики качества в задаче регрессии		2							
		5. Квантильная регрессия. Метрики качества				4					

6. Линейные методы регрессии								22	
3. Линейные методы классификации									
1. Постановка задачи классификации. Метрики качества в задачах классификации	2								
2. Байесовский классификатор. Линейный дискриминантный анализ и логистическая регрессия	4								
3. Метод опорных векторов и машина опорных векторов	2								
4. Перцептрон	2								
5. Логистическая регрессия и линейный дискриминантный анализ			4						
6. Линейная классификация методом опорных векторов. Машина опорных векторов			2						
7. Элементарный перцептрон.			2						
8. Линейные методы классификации								24	
4. Дисперсионный и ковариационный анализ									
1. Дисперсионный анализ	2								
2. Ковариационный анализ	4								
3. Ковариационный анализ			6						
4. Дисперсионный и ковариационный анализ								22	
5. Проблемы качества моделей									
1. Проблема мультиколлинеарности	2								
2. Отбор значимых признаков. Метод главных компонент	2								
3. Проблема переобучения. Регуляризация	2								
4. Работа с выбросами и пропущенными значениями	2								
5. Отбор признаков в задаче классификации с помощью средств пакета Caret			2						

6. Пошаговые методы отбора признаков в задаче регрессии			4					
7. Метод главных компонент			2					
8. Работа с выбросами			2					
9. Проблемы качества моделей							26	
Всего	36		36				108	

4 Учебно-методическое обеспечение дисциплины

4.1 Печатные и электронные издания:

1. Кабаков Р., Волкова П. А. R в действии. Анализ и визуализация данных на языке R(Москва: ДМК Пресс).
2. Джеймс Г. Введение в статистическое обучение с примерами на языке R (Москва: ДМК Пресс).
3. Волкова П. А., Шипунов А. Б. Статистическая обработка данных в учебно-исследовательских работах: Учебное пособие(Москва: Издательство "ФОРУМ").
4. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников(Москва: Физматлит).
5. Зарова Е. В. Applied Multivariate Statistical Analysis: Presentations for Lecturing and Working Examples with R=Прикладной многомерный статистический анализ: Презентации для лекций и примеры решений с использованием пакета R: Учебное пособие на английском языке (Москва: ООО "Научно-издательский центр ИНФРА-М").
6. Рубан А. И. Методы анализа данных: учебное пособие(Красноярск: 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. Для выполнения индивидуальных заданий по программированию алгоритмов анализа данных необходима программная среда вычислений R версии не младше, чем 3.4.3, а также редактор кода RStudio версии 1.1 или старше.

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.