

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  
(The Basics of Machine Learning)

наименование дисциплины (модуля) в соответствии с учебным планом

Направление подготовки / специальность

01.04.02 Прикладная математика и информатика

Направленность (профиль)

01.04.02.09 Data Science and Mathematical Modeling

Форма обучения

очная

Год набора

2022

## **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
<b>ПК-4: Способен разрабатывать и внедрять новые методы и технологии исследования данных.</b>	
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.	

ПК-4.2: Способен проводить аналитические и поисковые исследования по тематике информационных технологий, технологий данных.	
<b>УК-1: Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий.</b>	
УК-1.1: Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними.	
УК-1.2: Определяет пробелы в информации, необходимой для решения проблемной ситуации, и проектирует процессы по их устранению.	
УК-1.3: Критически оценивает надежность источников информации, работает с противоречивой информацией из разных источников.	
УК-1.4: Разрабатывает и содержательно аргументирует стратегию решения проблемной ситуации на основе системного и междисциплинарных подходов.	
УК-1.5: Использует логико-методологический инструментарий для критической оценки современных концепций философского и социального характера в своей предметной области.	
<b>УК-2: Способен управлять проектом на всех этапах его жизненного цикла.</b>	
УК-2.1: Формулирует на основе поставленной проблемы проектную задачу и способ ее решения через реализацию проектного управления.	

УК-2.2: Разрабатывает концепцию проекта в рамках обозначенной проблемы: формулирует цель, задачи, обосновывает актуальность, значимость, ожидаемые	
результаты и возможные сферы их применения.	
УК-2.3: Планирует необходимые ресурсы, в том числе с учетом их заменимости.	
УК-2.4: Разрабатывает 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.	

#### **1.4 Особенности реализации дисциплины**

Язык реализации дисциплины: Английский.

Дисциплина (модуль) реализуется без применения ЭО и ДОТ.

## 2. Объем дисциплины (модуля)

Вид учебной работы	Всего, зачетных единиц (акад.час)	е
		1
<b>Контактная работа с преподавателем:</b>	<b>2 (72)</b>	
занятия лекционного типа	1 (36)	
практические занятия	1 (36)	
<b>Самостоятельная работа обучающихся:</b>	<b>3 (108)</b>	
курсовое проектирование (КП)	Нет	
курсовая работа (КР)	Нет	
<b>Промежуточная аттестация (Экзамен)</b>	<b>1 (36)</b>	

### 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
<b>1. Basic components of machine learning</b>									
	1. The ingredients of machine learning	2							
	2. Basics of programming using R			2					
	3. Data Preparation	2							
	4. Descriptive statistics and visual data analysis			2					
	5. Basic components of machine learning							14	
<b>2. Linear methods of regression</b>									
	1. Probabilistic justification of the regression model	2							
	2. Multiple linear regression with Least square method			4					
	3. Multiple linear regression model	4							
	4. Metrics in regression models	2							
	5. Quantile regression. Metrics of regression			4					
	6. Linear methods of regression							22	
<b>3. Linear methods of classification</b>									

1. Classification problem. Metrics for classification	2							
2. Bayesian classifier. Linear discriminant analysis. Logistic Regression.	4							
3. Support vector classifier. Support vector machine	2							
4. Linear perceptron	2							
5. Linear discriminant analysis. Logistic Regression.			4					
6. Support vector classifier. Support vector machine			2					
7. Elemental perceptron			2					
8. Linear methods of classification							24	
<b>4. Variance and Covariance analysis</b>								
1. Analysis of variance	2							
2. Analysis of covariance	4							
3. Analysis of covariance			6					
4. Variance and Covariance analysis							22	
<b>5. Model quality issues</b>								
1. Multicollinearity problem	2							
2. Feature selection. Principal component analysis	2							
3. Overfitting problem. Regularization.	2							
4. Outliers and missed values	2							
5. Feature selection for classification model with Caret package			2					
6. Stepwise feature selection in regression models			4					
7. Principle component analysis			2					
8. Outlier and high leverage points detection			2					
9. Model quality issues							26	
Bcero	36		36				108	



## **4 Учебно-методическое обеспечение дисциплины**

### **4.1 Печатные и электронные издания:**

1. Neter J., Wasserman W., Kutner M. H. Applied Linear Statistical Models: Regression, Analysis of Variance and Experimental Designs(Boston: Irwin).
2. Dunn P. F. Measurement and Data Analysis for Engineering and Science: научное издание(Boca Raton: CRC Press).
3. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников(Москва: Физматлит).
4. Кабаков Р., Волкова П. А. R в действии. Анализ и визуализация данных на языке R(Москва: ДМК Пресс).
5. Джеймс Г. Введение в статистическое обучение с примерами на языке R (Москва: ДМК Пресс).
6. Волкова П. А., Шипунов А. Б. Статистическая обработка данных в учебно-исследовательских работах: Учебное пособие(Москва: Издательство "ФОРУМ").
7. Рубан А. И. Методы анализа данных: учебное пособие(Красноярск: СФУ).
8. 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").

### **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.