

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

Б1.О.15 Анализ и системы управления большими  
данными (Big Data)

наименование дисциплины (модуля) в соответствии с учебным планом

Направление подготовки / специальность

09.03.03 Прикладная информатика

Направленность (профиль)

09.03.03 Прикладная информатика

Форма обучения

очная

Год набора

2019

## **РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

Программу составили \_\_\_\_\_

канд.техн.наук, Доцент, Капустина С.В.

\_\_\_\_\_  
должность, инициалы, фамилия

## 1 Цели и задачи изучения дисциплины

### 1.1 Цель преподавания дисциплины

Формирование у студентов теоретических основ технологии обработки больших данных, умений работы с большим потоком разнородной распределенной информации в структурированной и неструктурированной форме

### 1.2 Задачи изучения дисциплины

- познакомить студентов с основными понятиями больших данных;  
- познакомить студентов с основными принципами машинного обучения, видами задач, классами моделей (линейными, логическими, нейросетевыми, метриками качества и подходами к предварительной обработке 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
<b>ОПК-2: Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решения задач профессиональной деятельности;</b>	
ОПК-2.1: Знает современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности	методы поиска информационных ресурсов (CRM, ERP систем) находить, анализировать и использовать необходимую экономическую информацию для принятия обоснованных решений в профессиональной сфере навыками поиска информации посредством электронных ресурсов
ОПК-2.2: Умеет выбирать современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности	методы решения задач обучения с учителем: линейная и логистическая регрессия, классификация уметь выбирать современные технологии программными средствами
ОПК-2.3: Владеет навыками применения современных информационных технологий и программных средств, в том числе отечественного производства, при решении задач профессиональной деятельности	методы решения задач без учителя: кластеризация (сегментация клиентов и компаний) анализировать финансовую и экономическую информацию, необходимую для принятия обоснованных решений в профессиональной сфере инструментами цифровых технологий для автоматизации информационно-аналитической деятельности

### 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=29865>.

## 2. Объем дисциплины (модуля)

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

### 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
<b>1. Основные понятия больших данных</b>											
		1. Этапы анализа данных. Постановки задач машинного обучения.		4							
		2. Виды форматов данных. Разведочный анализ данных				12					
		3. Поиск информативных признаков							18		
<b>2. Методы обучения с учителем</b>											
		1. Линейная регрессия, логистическая регрессия, методы классификации		8							
		2. Классификация и оценки принадлежности классов. AUC-ROC				12					
		3. Композиция алгоритмов							18		
<b>3. Кластерный анализ</b>											
		1. Классификация алгоритмов классификации. Иерархические и неиерархические алгоритмы.		6							

2. EM-алгоритм, метод k-средних, DBSCAN			12					
3. Качество кластеризации							18	
<b>4. Нечеткая логика</b>								
1. Нечеткие множества и нечеткая логика	8							
2. Нечеткий вывод:Сугено, Мамдани.			12					
3. Решение прикладных задач							18	
<b>5. Нейронные сети</b>								
1. Структура нейронной сети. Рекуррентные, сверточные сети								
2. Разработка глубоких моделей идентификации в Python			12					
3. Техники обработки изображений							18	
<b>6. Рекомендательные системы</b>								
1. Поиск ассоциативных правил	10							
2. Алгоритмы Apriori, FPG-tree			12					
3. Машинный перевод							18	
Всего	36		72				108	

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

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

1. Коэльо Л. П. Построение систем машинного обучения на языке Python (Москва: ДМК Пресс).
2. Жуков Р.А. Язык программирования Python: практикум: Учебное пособие(Москва: ООО "Научно-издательский центр ИНФРА-М").
3. Маккинли У. Python и анализ данных(Москва: ДМК Пресс).
4. Вандер Плас Д. Python для сложных задач. Наука о данных и машинное обучение (pdf+epub)(Б. м.: б. и.).
5. Рашка С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения(Москва: ДМК Пресс).

### **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. Все практические работы проводятся на ЭВМ в среде
2. Jupyter Notebook,
3. Python,
4. пакеты numpy, pandas, scikitlearn, tensorflow, Keras

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

1. UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/>)
2. Сайт Национального открытого университета ИНТУИТ [Электронный ресурс] : – Режим доступа, открытый, [www.intuit.ru](http://www.intuit.ru).
3. Портал <http://www.machinelearning.ru>
4. <https://github.com/>
5. Dataset Kaggle [Электронный ресурс]. – URL:<https://www.kaggle.com/c/news-classification-by-categories/data>
6. Сайт Национального открытого университета ИНТУИТ [Электронный ресурс] : - Режим доступа, открытый, [www.intuit.ru](http://www.intuit.ru)
7. Центр машинного обучения и интеллектуальных систем <http://archive.ics.uci.edu/ml/>
8. Курс ШАД Яндекса от К.В.Воронцова <https://youtu.be/qLBkV4sMztk>

## **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)**

Для осуществления образовательного процесса необходимо:

- интерактивная доска INTERWRITE;
- проектор;
- компьютеры Intel Core i5-4460 3.20GHz, память 4Gb, HDD 500 Gb, монитор LOC E2460S 24"