

Министерство науки и высшего образования РФ
Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

СОГЛАСОВАНО

Заведующий кафедрой

**Кафедра бизнес-информатики и
моделирования бизнес-процессов**

наименование кафедры

подпись, инициалы, фамилия

«___» _____ 20__ г.

институт, реализующий ОП ВО

УТВЕРЖДАЮ

Заведующий кафедрой

**Кафедра бизнес-информатики и
моделирования бизнес-процессов**

наименование кафедры

Кашина Е.В.

подпись, инициалы, фамилия

«___» _____ 20__ г.

институт, реализующий дисциплину

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
УПРАВЛЕНИЕ БОЛЬШИМИ
ДАНЫМИ (BIG DATA)**

Дисциплина Б1.В.ДВ.04.02 Управление большими данными (Big Data)

Направление подготовки /
специальность

Направленность
(профиль)

Форма обучения

Год набора

очная

2019

Красноярск 2021

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

составлена в соответствии с Федеральным государственным образовательным стандартом высшего образования с учетом профессиональных стандартов по укрупненной группе

090000 «ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА»

Направление подготовки /специальность (профиль/специализация)

09.03.03 Прикладная информатика

Программу
составили

ассистент, Карасева Т.С.

1 Цели и задачи изучения дисциплины

1.1 Цель преподавания дисциплины

Целью преподавания дисциплины «Управление большими данными (Big Data)» является фундаментальная подготовка в области обработки больших данных, знакомство со средствами обработки больших данных.

1.2 Задачи изучения дисциплины

- изучение современных тенденций в области управления большими данными;
- освоение методов прогнозирования;
- освоение методов регрессионного анализа;
- освоение методов кластерного анализа;
- обзор технологий обработки больших данных.

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

ПК-7:Способен анализировать информационные потребности пользователей информационных ресурсов	
ПК-7.1:Знает: терминологию и ключевые параметры веб-статистики; основные принципы и методы сбора статистики посещаемости веб-сайтов; популярные сервисы для сбора веб-статистики	
Уровень 1	терминологическую базу в области анализа больших данных
Уровень 1	анализировать научно-техническую литературу в области анализа больших данных
Уровень 1	международным терминологическим базисом в области анализа больших данных
ПК-7.2:Умеет: анализировать структурированную и неструктурированную информацию; - владеть популярными сервисами для оценки посещаемости характеристик аудитории сайтов	
Уровень 1	основные инструменты для анализа данных
Уровень 1	структурировать необработанные данные
Уровень 1	методами предобработки больших данных
ПК-7.3:Владеет навыками: проведения общего анализа посещаемости сайта; выявления наиболее популярных страниц; выработки решений по наполнению сайта контентом	
Уровень 1	основные этапы анализа данных
Уровень 1	строить выводы на основании полученного анализа
Уровень 1	методами анализа больших данных

1.4 Место дисциплины (модуля) в структуре образовательной

программы

Для дисциплины "«Управление большими данными (Big Data)»" требуются знания и компетенции, полученные в рамках дисциплин:

- Базы данных;
- Язык Python для анализа данных;
- Системы искусственного интеллекта.

Знания, полученные в рамках изучения дисциплины, необходимы для освоения дисциплин:

- Интеллектуальный анализ данных;
- Машинное обучение.

1.5 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)	Семестр	
		7	8
Общая трудоемкость дисциплины	7 (252)	3 (108)	4 (144)
Контактная работа с преподавателем:	2,5 (90)	1,5 (54)	1 (36)
занятия лекционного типа	0,83 (30)	0,5 (18)	0,33 (12)
занятия семинарского типа			
в том числе: семинары			
практические занятия	1,67 (60)	1 (36)	0,67 (24)
практикумы			
лабораторные работы			
другие виды контактной работы			
в том числе: групповые консультации			
индивидуальные консультации			
иная внеаудиторная контактная работа:			
групповые занятия			
индивидуальные занятия			
Самостоятельная работа обучающихся:	4,5 (162)	1,5 (54)	3 (108)
изучение теоретического курса (ТО)			
расчетно-графические задания, задачи (РГЗ)			
реферат, эссе (Р)			
курсовое проектирование (КП)	Нет	Нет	Нет
курсовая работа (КР)	Нет	Нет	Нет
Промежуточная аттестация (Зачёт)			

3 Содержание дисциплины (модуля)

3.1 Разделы дисциплины и виды занятий (тематический план занятий)

№ п/п	Модули, темы (разделы) дисциплины	Занятия лекционного типа (акад. час)	Занятия семинарского типа		Самостоятельная работа, (акад. час)	Формируемые компетенции
			Семинары и/или Практические занятия (акад. час)	Лабораторные работы и/или Практикумы (акад. час)		
1	2	3	4	5	6	7
1	Обзор Big-Data. Методы и средства. Используемые программы.	2	16	0	10	ПК-7.1 ПК-7.2 ПК-7.3
2	Прогнозное моделирование	2	6	0	10	ПК-7.1 ПК-7.2 ПК-7.3
3	Прогнозное моделирование: работа с регрессионными моделями	6	0	0	10	ПК-7.1 ПК-7.2 ПК-7.3
4	Прогнозное моделирование: нейронные сети.	4	6	0	12	ПК-7.1 ПК-7.2 ПК-7.3
5	Кластерный анализ	4	8	0	12	ПК-7.1 ПК-7.2 ПК-7.3
6	Оценка модели анализа данных	3	12	0	10	ПК-7.1 ПК-7.2 ПК-7.3
7	Анализ данных с использованием Python	9	12	0	98	ПК-7.1 ПК-7.2 ПК-7.3
Всего		30	60	0	162	

3.2 Занятия лекционного типа

№ п/п	№ раздела дисциплины	Наименование занятий	Объем в акад. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме

1	1	Big-Data. Инструменты. Технологии. Методы анализа.	2	0	0
2	2	Область прикладных задач с использованием прогнозного моделирования. Проклятие размерности. Избавление от бесполезных и избыточных входных переменных. Создание обучающих и проверочных данных	2	0	0
3	3	Регрессия. Логистическая регрессия. Полиномиальные регрессии. Оценка параметров. Обработка пропущенных значений. Выбор входных переменных. Оптимизация сложности. Интерпретация регрессии.	4	0	0
4	3	Регрессии с преобразованными входными переменными. Категориальные переменные в регрессионной модели. Область прикладных задач с использованием прогнозного моделирования. Проклятие размерности. Создание обучающих и проверочных данных	2	0	0
5	4	Особенности нейронных сетей. Структура однослойных и многослойных нейронных сетей. Применение нейронных сетей.	2	0	0

6	4	Обучение нейронной сети, классификация алгоритмов обучения. Инструмент AutoNeural.	2	0	0
7	5	Основные задачи, решаемые с помощью кластерного анализа, сфера применения. ПОДХОДЫ И МЕТОДЫ КЛАСТЕРНОГО АНАЛИЗА	2	0	0
8	5	Кластерный анализ методом k-средних. Понимание кластера и центра кластера, суть алгоритма. Требования к исходным данным. Определение количества кластеров, критерии оценки качества кластеров. Интерпретация кластеров.	2	0	0
9	6	Принципы оценки качества моделей для анализа данных	3	0	0
10	7	Анализ данных. Введение в Python	3	0	0
11	7	Основные библиотеки языка Python для анализа данных	3	0	0
12	7	Математический аппарат, требуемый для анализа данных	3	0	0
Всего			20	0	0

3.3 Занятия семинарского типа

№ п/п	№ раздела дисциплины	Наименование занятий	Объем в акад. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме

1	1	SAS Interprise Miner. Введение. Возможности. Инструменты (Меню. Принцип анализа данных SEMMA. Основные инструменты и узлы. Возможности построения моделей)	8	0	0
2	1	Создание проекта, библиотеки и диаграмм SAS. Настройки источника данных. Типы переменных. Изменение размера выборки. Создание диаграмм. Исследование взаимосвязей между переменными.	8	0	0
3	2	Прогнозная модель, использующая дерево решений (Создание дерева решений: структура. Алгоритм построения. Поиск разбиений. Прогнозная модель использующая дерево решений: построение, создание правила разбиения. Оптимизация сложности деревьев решений. Оценка качества дерева решений).	6	0	0
4	4	Прогнозная модель, использующая искусственные нейронные сети(Создание ИНС: структура. Алгоритм обучения т.д.)	6	0	0
5	5	Построение модели кластеризации, разбиение объектов в классы в системах распознавания образов "без учителя"	8	0	0
6	6	Оценка модели классификации	4	0	0
7	6	Оценка регрессионной модели	4	0	0
8	6	Оценка кластерной модели	4	0	0

9	7	Проверка статистических гипотез на языке Python	4	0	0
10	7	Обзор библиотек Python для анализа данных	4	0	0
11	7	Реализация нейронных сетей на языке Python	4	0	0
Итого			60	0	0

3.4 Лабораторные занятия

№ п/п	№ раздела дисциплины	Наименование занятий	Объем в акад. часах		
			Всего	в том числе, в инновационной форме	в том числе, в электронной форме
Итого					

4 Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

	Авторы, составители	Заглавие	Издательство, год
Л1.1	Лейнартас Д. Е.	Анализ данных. Математические методы анализа данных: учеб.-метод. пособие для самостоят. работы [для студентов напр. 080500.62 «Бизнес-информатика» очной формы обуч.]	Красноярск: СФУ, 2013
Л1.2	Крутиков В. И., Мешечкин В. В.	Анализ данных	Кемерово: КемГУ, 2014

5 Фонд оценочных средств для проведения промежуточной аттестации

Оценочные средства находятся в приложении к рабочим программам дисциплин.

6 Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

6.1. Основная литература			
	Авторы, составители	Заглавие	Издательство, год
Л1.1	Редько В. Г., Малинецкий Г. Г.	Эволюция, нейронные сети, интеллект: модели и концепции эволюционной кибернетики	Москва: URSS, 2011
Л1.2	Емельянов С. В.	Искусственный интеллект и принятие решений: Вып. 1	Москва: URSS, 2011

Л1.3	Богданов Е. П.	Интеллектуальный анализ данных: практикум для подготовки магистрантов направления 09.04.03 «прикладная информатика» профиль подготовки «информационные системы и технологии корпоративного управления»	Волгоград: Волгоградский ГАУ, 2019
6.2. Дополнительная литература			
	Авторы, составители	Заглавие	Издательство, год
Л2.1	Смолин Д. В.	Введение в искусственный интеллект: конспект лекций: методические указания	Москва: Физматлит, 2007
Л2.2	Цехановский В. В., Чертовской В. Д.	Управление данными: учебник	Санкт-Петербург: Лань, 2015
6.3. Методические разработки			
	Авторы, составители	Заглавие	Издательство, год
Л3.1	Лейнартас Д. Е.	Анализ данных. Математические методы анализа данных: учеб.-метод. пособие для самостоят. работы [для студентов напр. 080500.62 «Бизнес-информатика» очной формы обуч.]	Красноярск: СФУ, 2013
Л3.2	Крутиков В. И., Мешечкин В. В.	Анализ данных	Кемерово: КемГУ, 2014

7 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)

Э1	Сайт Национального открытого университета ИНТУИТ	www.intuit.ru
Э2	Открытое образование, курс "Машинное обучение"	https://openedu.ru/course/hse/INTRML/?session=fall_2020
Э3	Открытое образование, курс "Прикладной искусственный интеллект"	https://openedu.ru/course/ITMOUniversity/APPARTINT2035/

8 Методические указания для обучающихся по освоению дисциплины (модуля)

Организация процесса работы по дисциплине:

- для успешного освоения дисциплины предусмотрены лекции, в том числе с применением презентационного материала;
- для закрепления теоретического материала курс содержит практические работы, вариант задания к работе предоставляется непосредственно преподавателем (практические занятия включают такие формы интерактивного обучения, как работа в малых группах, решение кейсов, анализ прикладных ситуаций, проблемное обучение);
- после завершения изучения курса студент имеет возможность

сдать экзамен. Для этого необходимо в течение семестра выполнить все практические работы, выступить с докладом и ответить устно преподавателю на экзаменационные вопросы.

Объем самостоятельной работы – 54 ч. (7 семестр) и 108 ч. (8 семестр) Самостоятельная работа студентов включает в себя следующие виды учебной деятельности:

- изучение теоретического материала с использованием как рекомендуемой, так и другой литературы по разделам дисциплины;
- выполнение индивидуальных и типовых заданий, оформление отчетов по практическим работам;
- подготовку результатов по изученному материалу в форме отчетов, конспектов лекций.
- выступление с докладом.

Самостоятельная работа распределяется следующим образом:

- Обзор Big-Data. Методы и средства. Используемые программы (10 ч.)
- Прогнозное моделирование (10 ч.)
- Прогнозное моделирование: работа с регрессионными моделями (10 ч.)
- Прогнозное моделирование: нейронные сети (12 ч.)
- Кластерный анализ (12 ч.)
- Оценка модели анализа данных (10 ч.)
- Анализ данных с использованием Python (98 ч.)

Сроки самостоятельной работы по дисциплине распределяются в течение семестра в соответствии с расписанием практических и лекционных занятий.

Для самостоятельной работы обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья учебно-методические материалы для самостоятельной работы разрабатываются под соответствующую адаптированную или частично адаптированную ОП (при наличии).

9 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю) (при необходимости)

9.1 Перечень необходимого программного обеспечения

9.1.1	1. регулярно обновляемый интернет-браузер (Mozilla Firefox, Google Chrome, Yandex Browser, Opera, Internet Explorer, Safari, либо иной);
-------	--

9.1.2	2. офисный пакет (MS Office, Libre Office, Open Office, либо иной).
9.1.3	3. SAS Interprise Miner
9.1.4	4. Произвольная среда для работы с языком Python.

9.2 Перечень необходимых информационных справочных систем

9.2.1	1. Сайт библиотеки СФУ. Режим доступа: http://bik.sfu-kras.ru/
9.2.2	2. Электронный каталог библиотеки СФУ. Режим доступа: http://catalog.sfu-kras.ru/
9.2.3	

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

Для осуществления образовательного процесса необходимо:

- для проведения лекционных занятий – оснащенные проекционной и компьютерной техникой учебные аудитории, позволяющие выступающему (преподавателю, студенту) демонстрировать слайды в форматах pdf, PowerPoint и других графических форматах на экране с одновременным выступлением перед аудиторией;
- для проведения практических работ – компьютерный класс с установленным ПО из п.9.1 и доступом Интернет.