МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ
УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ»
(ТУСУР)

УТВЕРЖДАЮ
Проректор по учебной работе
_____П.В. Сенченко
«23» _____12_____ 2020 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

АНАЛИЗ БОЛЬШИХ ДАННЫХ

Уровень образования: **высшее образование - магистратура**
Направление подготовки / специальность: **09.04.04 Программная инженерия**
Направленность (профиль) / специализация: **Методы и технологии индустриального проектирования программного обеспечения**
Форма обучения: **очная**
Факультет: **Факультет систем управления (ФСУ)**
Кафедра: **Кафедра автоматизации обработки информации (АОИ)**
Курс: **1**
Семестр: **2**
Учебный план набора 2021 года

Объем дисциплины и виды учебной деятельности

Виды учебной деятельности	2 семестр	Всего	Единицы
Лекционные занятия	18	18	часов
Лабораторные занятия	36	36	часов
Самостоятельная работа	90	90	часов
Общая трудоемкость (включая промежуточную аттестацию)	144	144	часов
	4	4	з.е.

Формы промежуточной аттестация	Семестр
Зачет с оценкой	2

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Сенченко П.В.
Должность: Проректор по УР
Дата подписания: 23.12.2020
Уникальный программный ключ:
a1119608-cdff-4455-b54e-5235117c185c

Томск

Согласована на портале № 63028

1. Общие положения

1.1. Цели дисциплины

1. Освоение теоретических основ и методов анализа данных, применяемых при решении прикладных задач с помощью языка программирования Python.

1.2. Задачи дисциплины

1. Формирование навыков обработки и анализа информации для получения общей оценки массива данных, и выявления тенденций, закономерностей в изучаемых наборах данных. Освоение языка программирования Python, Interactive Python в виде Jupyter Notebook, библиотеки Pandas, применяемых при анализе данных.

2. Место дисциплины в структуре ОПОП

Блок дисциплин: Б1. Дисциплины (модули).
Часть блока дисциплин: Обязательная часть.
Модуль дисциплин: Специализированный модуль (hard skills - HS).
Индекс дисциплины: Б1.О.02.03.
Реализуется с применением электронного обучения, дистанционных образовательных технологий.

3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Процесс изучения дисциплины направлен на формирование следующих компетенций в соответствии с ФГОС ВО и основной образовательной программой (таблица 3.1):
Таблица 3.1 – Компетенции и индикаторы их достижения

Компетенция	Индикаторы достижения компетенции	Планируемые результаты обучения по дисциплине
Универсальные компетенции		
-	-	-
Общепрофессиональные компетенции		

Согласована на портале № 63028

ОПК-7. Способен применять при решении профессиональных задач методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе, в глобальных компьютерных сетях	ОПК-7.1. Знает методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе, в глобальных компьютерных сетях	Умение использовать средства Python для обработки и анализа данных. Понимание структуры современного конвейера обработки данных и средств работы со сверхбольшими объемами данных.
	ОПК-7.2. Умеет применять методы и средства получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе, в глобальных компьютерных сетях	Умение выполнить очистку и объединение данных, представленных в разных форматах и из разных источников средствами Python.
	ОПК-7.3. Владеет навыками, методами и средствами получения, хранения, переработки и трансляции информации посредством современных компьютерных технологий, в том числе, в глобальных компьютерных сетях	Умение развернуть, настроить и применить кластер Hadoop для обработки сверхбольших массивов данных.
Профессиональные компетенции		
-	-	-

4. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 4 зачетных единиц, 144 академических часов.

Распределение трудоемкости дисциплины по видам учебной деятельности представлено в таблице 4.1.

Таблица 4.1 – Трудоемкость дисциплины по видам учебной деятельности

Виды учебной деятельности	Всего часов	Семестры
		2 семестр
Контактная аудиторная работа обучающихся с преподавателем, всего	54	54
Лекционные занятия	18	18
Лабораторные занятия	36	36
Самостоятельная работа обучающихся, в т.ч. контактная внеаудиторная работа обучающихся с преподавателем, всего	90	90
Подготовка к зачету с оценкой	36	36
Подготовка к лабораторной работе, написание отчета	18	18
Подготовка к тестированию	22	22
Подготовка к устному опросу / собеседованию	14	14
Общая трудоемкость (в часах)	144	144
Общая трудоемкость (в з.е.)	4	4

5. Структура и содержание дисциплины

Согласована на портале № 63028

5.1. Разделы (темы) дисциплины и виды учебной деятельности

Структура дисциплины по разделам (темам) и видам учебной деятельности приведена в таблице 5.1.

Таблица 5.1 – Разделы (темы) дисциплины и виды учебной деятельности

Названия разделов (тем) дисциплины	Лек. зан., ч	Лаб. раб.	Сам. раб., ч	Всего часов (без экзамена)	Формируемые компетенции
2 семестр					
1 Введение.	2	4	10	16	ОПК-7
2 Основные принципы и этапы анализа данных.	2	4	10	16	ОПК-7
3 Анализ данных может быть разным.	2	4	10	16	ОПК-7
4 Подготовка и очистка данных.	2	4	10	16	ОПК-7
5 Выявление зависимостей и связей в данных.	2	4	10	16	ОПК-7
6 Методы машинного обучения для обработки данных.	2	12	12	26	ОПК-7
7 Графическое представление данных.	2	4	12	18	ОПК-7
8 Большие данные. Экосистема Hadoop.	2	-	8	10	ОПК-7
9 Распределенные системы анализа данных.	2	-	8	10	ОПК-7
Итого за семестр	18	36	90	144	
Итого	18	36	90	144	

5.2. Содержание разделов (тем) дисциплины

Содержание разделов (тем) дисциплины (в т.ч. по лекциям) приведено в таблице 5.2.

Таблица 5.2 – Содержание разделов (тем) дисциплины (в т.ч. по лекциям)

Названия разделов (тем) дисциплины	Содержание разделов (тем) дисциплины (в т.ч. по лекциям)	Трудоемкость (лекционные занятия), ч	Формируемые компетенции
2 семестр			
1 Введение.	Причины появления термина "Большие данные". Роль данных в информационном мире.	2	ОПК-7
	Итого	2	
2 Основные принципы и этапы анализа данных.	Методика CRISP-DM. Данные в покое и движении. Структурированные и неструктурированные данные.	2	ОПК-7
	Итого	2	
3 Анализ данных может быть разным.	Описательный, прогнозный и рекомендательный анализ.	2	ОПК-7
	Итого	2	
4 Подготовка и очистка данных.	Методы преобразования форматов в Python. Методы статистической обработки данных.	2	ОПК-7
	Итого	2	

Согласована на портале № 63028

5 Выявление зависимостей и связей в данных.	Выявление тренда, линейная и нелинейная регрессия. Методы кластеризации.	2	ОПК-7
	Итого	2	
6 Методы машинного обучения для обработки данных.	Общие принципы работы нейронных сетей, глубокое обучение, понятие "искусственный интеллект" и "машинное обучение", методы машинного обучения.	2	ОПК-7
	Итого	2	
7 Графическое представление данных.	Библиотека Python Matplotlib, возможности, использование в Jupyter Notebook, особенности работы и настройки. Библиотеки PyPlot и Plotly.	2	ОПК-7
	Итого	2	
8 Большие данные. Экосистема Hadoop.	Принцип Map-Reduce, основные характеристики кластера Hadoop. Файловая система HDFS, инструменты настройки YARN, управление доступом к данным Pig, Hive. Другие инструменты экосистемы.	2	ОПК-7
	Итого	2	
9 Распределенные системы анализа данных.	Процесс ETL. Традиционная архитектура обработки данных и архитектура Lambda, потоковая обработка данных в реальном времени. Архитектура Kappa, проблемы распределенной обработки.	2	ОПК-7
	Итого	2	
Итого за семестр		18	
Итого		18	

5.3. Практические занятия (семинары)

Не предусмотрено учебным планом

5.4. Лабораторные занятия

Наименование лабораторных работ приведено в таблице 5.4.

Таблица 5.4 – Наименование лабораторных работ

Названия разделов (тем) дисциплины	Наименование лабораторных работ	Трудоемкость, ч	Формируемые компетенции
2 семестр			
1 Введение.	Анализ криминальной обстановки в Сан-Франциско	4	ОПК-7
	Итого	4	
2 Основные принципы и этапы анализа данных.	Измерение и анализ данных Интернет-соединения	4	ОПК-7
	Итого	4	

Согласована на портале № 63028

3 Анализ данных может быть разным.	Измерение и анализ Интернет-соединения с хранилищем данных в виде SQL базы данных.	4	ОПК-7
	Итого	4	
4 Подготовка и очистка данных.	Описательная статистика в Python	4	ОПК-7
	Итого	4	
5 Выявление зависимостей и связей в данных.	Корреляционный анализ в Python	4	ОПК-7
	Итого	4	
6 Методы машинного обучения для обработки данных.	Простая линейная регрессия в Python	4	ОПК-7
	Классификация с помощью дерева решений	4	ОПК-7
	Обработка ошибок при использовании метода линейной регрессии	4	ОПК-7
	Итого	12	
7 Графическое представление данных.	Измерение и анализ Интернет-соединения визуализация результатов.	4	ОПК-7
	Итого	4	
Итого за семестр		36	
Итого		36	

5.5. Курсовой проект / курсовая работа

Не предусмотрено учебным планом

5.6. Самостоятельная работа

Виды самостоятельной работы, трудоемкость и формируемые компетенции представлены в таблице 5.6.

Таблица 5.6 – Виды самостоятельной работы, трудоемкость и формируемые компетенции

Названия разделов (тем) дисциплины	Виды самостоятельной работы	Трудоемкость, ч	Формируемые компетенции	Формы контроля
2 семестр				
1 Введение.	Подготовка к зачету с оценкой	4	ОПК-7	Зачёт с оценкой
	Подготовка к лабораторной работе, написание отчета	2	ОПК-7	Лабораторная работа
	Подготовка к тестированию	2	ОПК-7	Тестирование
	Подготовка к устному опросу / собеседованию	2	ОПК-7	Устный опрос / собеседование
	Итого	10		

Согласована на портале № 63028

2 Основные принципы и этапы анализа данных.	Подготовка к зачету с оценкой	4	ОПК-7	Зачёт с оценкой
	Подготовка к лабораторной работе, написание отчета	2	ОПК-7	Лабораторная работа
	Подготовка к тестированию	2	ОПК-7	Тестирование
	Подготовка к устному опросу / собеседованию	2	ОПК-7	Устный опрос / собеседование
	Итого	10		
3 Анализ данных может быть разным.	Подготовка к зачету с оценкой	4	ОПК-7	Зачёт с оценкой
	Подготовка к лабораторной работе, написание отчета	2	ОПК-7	Лабораторная работа
	Подготовка к тестированию	2	ОПК-7	Тестирование
	Подготовка к устному опросу / собеседованию	2	ОПК-7	Устный опрос / собеседование
	Итого	10		
4 Подготовка и очистка данных.	Подготовка к зачету с оценкой	4	ОПК-7	Зачёт с оценкой
	Подготовка к лабораторной работе, написание отчета	2	ОПК-7	Лабораторная работа
	Подготовка к тестированию	2	ОПК-7	Тестирование
	Подготовка к устному опросу / собеседованию	2	ОПК-7	Устный опрос / собеседование
	Итого	10		
5 Выявление зависимостей и связей в данных.	Подготовка к зачету с оценкой	4	ОПК-7	Зачёт с оценкой
	Подготовка к лабораторной работе, написание отчета	2	ОПК-7	Лабораторная работа
	Подготовка к тестированию	2	ОПК-7	Тестирование
	Подготовка к устному опросу / собеседованию	2	ОПК-7	Устный опрос / собеседование
	Итого	10		

Согласована на портале № 63028

6 Методы машинного обучения для обработки данных.	Подготовка к зачету с оценкой	4	ОПК-7	Зачёт с оценкой
	Подготовка к лабораторной работе, написание отчета	4	ОПК-7	Лабораторная работа
	Подготовка к тестированию	2	ОПК-7	Тестирование
	Подготовка к устному опросу / собеседованию	2	ОПК-7	Устный опрос / собеседование
	Итого	12		
7 Графическое представление данных.	Подготовка к зачету с оценкой	4	ОПК-7	Зачёт с оценкой
	Подготовка к лабораторной работе, написание отчета	4	ОПК-7	Лабораторная работа
	Подготовка к тестированию	2	ОПК-7	Тестирование
	Подготовка к устному опросу / собеседованию	2	ОПК-7	Устный опрос / собеседование
	Итого	12		
8 Большие данные. Экосистема Hadoop.	Подготовка к зачету с оценкой	4	ОПК-7	Зачёт с оценкой
	Подготовка к тестированию	4	ОПК-7	Тестирование
	Итого	8		
9 Распределенные системы анализа данных.	Подготовка к зачету с оценкой	4	ОПК-7	Зачёт с оценкой
	Подготовка к тестированию	4	ОПК-7	Тестирование
	Итого	8		
Итого за семестр		90		
Итого		90		

5.7. Соответствие компетенций, формируемых при изучении дисциплины, и видов учебной деятельности

Соответствие компетенций, формируемых при изучении дисциплины, и видов учебной деятельности представлено в таблице 5.7.

Таблица 5.7 – Соответствие компетенций, формируемых при изучении дисциплины, и видов занятий

Формируемые компетенции	Виды учебной деятельности			Формы контроля
	Лек. зан.	Лаб. раб.	Сам. раб.	
ОПК-7	+	+	+	Зачёт с оценкой, Устный опрос / собеседование, Лабораторная работа, Тестирование

6. Рейтинговая система для оценки успеваемости обучающихся

6.1. Балльные оценки для форм контроля

Балльные оценки для форм контроля представлены в таблице 6.1.

Таблица 6.1 – Балльные оценки

Согласована на портале № 63028

Формы контроля	Максимальный балл на 1-ую КТ с начала семестра	Максимальный балл за период между 1КТ и 2КТ	Максимальный балл за период между 2КТ и на конец семестра	Всего за семестр
2 семестр				
Зачёт с оценкой	0	0	0	0
Устный опрос / собеседование	5	5	10	20
Лабораторная работа	20	20	20	60
Тестирование	5	5	10	20
Итого максимум за период	30	30	40	100
Нарастающим итогом	30	60	100	100

6.2. Пересчет баллов в оценки за текущий контроль

Пересчет баллов в оценки за текущий контроль представлен в таблице 6.2.

Таблица 6.2 – Пересчет баллов в оценки за текущий контроль

Баллы на дату текущего контроля	Оценка
≥ 90% от максимальной суммы баллов на дату ТК	5
От 70% до 89% от максимальной суммы баллов на дату ТК	4
От 60% до 69% от максимальной суммы баллов на дату ТК	3
< 60% от максимальной суммы баллов на дату ТК	2

6.3. Пересчет суммы баллов в традиционную и международную оценку

Пересчет суммы баллов в традиционную и международную оценку представлен в таблице 6.3.

Таблица 6.3 – Пересчет суммы баллов в традиционную и международную оценку

Оценка	Итоговая сумма баллов, учитывает успешно сданный экзамен	Оценка (ECTS)
5 (отлично) (зачтено)	90 – 100	A (отлично)
4 (хорошо) (зачтено)	85 – 89	B (очень хорошо)
	75 – 84	C (хорошо)
	70 – 74	D (удовлетворительно)
3 (удовлетворительно) (зачтено)	65 – 69	
	60 – 64	E (посредственно)
2 (неудовлетворительно) (не зачтено)	Ниже 60 баллов	F (неудовлетворительно)

7. Учебно-методическое и информационное обеспечение дисциплины

7.1. Основная литература

1. Маккинни, У. Python и анализ данных / У. Маккинни ; перевод с английского А. А. Слинкина. — 2-ое изд., испр. и доп. — Москва : ДМК Пресс, 2020. — 540 с. — ISBN 978-5-97060-590-5. — Текст : электронный // Лань : электронно-библиотечная система. — Режим доступа: для авториз. пользователей. [Электронный ресурс]: — Режим доступа: https://e.lanbook.com/book/131721.

7.2. Дополнительная литература

Согласована на портале № 63028

1. Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с. — Текст : электронный // Лань : электронно-библиотечная система. — Режим доступа: для авториз. пользователей. [Электронный ресурс]: — Режим доступа: https://e.lanbook.com/book/182452.

2. Программные системы статистического анализа. Обнаружение закономерностей в данных с использованием системы R и языка Python : учебное пособие / В. М. Волкова, М. А. Семёнова, Е. С. Четвертакова, С. С. Вожов. — Новосибирск : НГТУ, 2017. — 74 с. — ISBN 978-5-7782-3183-2. — Текст : электронный // Лань : электронно-библиотечная система. — Режим доступа: для авториз. пользователей. [Электронный ресурс]: — Режим доступа: https://e.lanbook.com/book/118287.

7.3. Учебно-методические пособия

7.3.1. Обязательные учебно-методические пособия

1. Широбокова, С. Н. Программирование на языке Python для лабораторных занятий : учебное пособие / С. Н. Широбокова, А. А. Кацупеев, А. В. Сулыз. — Новочеркасск : ЮРГПУ, 2020. — 104 с. — ISBN 978-5-9997-0725-3. — Текст : электронный // Лань : электронно-библиотечная система. — Режим доступа: для авториз. пользователей. [Электронный ресурс]: — Режим доступа: https://e.lanbook.com/book/180938.

2. Учебно-методическое пособие по лабораторному практикуму для студентов технических направлений подготовки и специальностей / Е.Ю. Агеев. – Томск: Томский государственный университет систем управления и радиоэлектроники, 2021. – 72 с. [Электронный ресурс]: — Режим доступа: https://sdo.tusur.ru/mod/resource/view.php?id=281680.

7.3.2. Учебно-методические пособия для лиц с ограниченными возможностями здоровья и инвалидов

Учебно-методические материалы для самостоятельной и аудиторной работы обучающихся из числа лиц с ограниченными возможностями здоровья и инвалидов предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации.

Для лиц с нарушениями зрения:

– в форме электронного документа;

– в печатной форме увеличенным шрифтом.

Для лиц с нарушениями слуха:

– в форме электронного документа;

– в печатной форме.

Для лиц с нарушениями опорно-двигательного аппарата:

– в форме электронного документа;

– в печатной форме.

7.4. Современные профессиональные базы данных и информационные справочные системы

1. При изучении дисциплины рекомендуется обращаться к современным базам данных, информационно-справочным и поисковым системам, к которым у ТУСУРа открыт доступ: https://lib.tusur.ru/ru/resursy/bazy-dannyh.

8. Материально-техническое и программное обеспечение дисциплины

8.1. Материально-техническое и программное обеспечение для лекционных занятий

Для проведения занятий лекционного типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации используется учебная аудитория с достаточным количеством посадочных мест для учебной группы, оборудованная доской и стандартной учебной мебелью. Имеются мультимедийное оборудование и учебно-наглядные пособия, обеспечивающие тематические иллюстрации по лекционным разделам дисциплины.

8.2. Материально-техническое и программное обеспечение для лабораторных работ

Компьютерный класс: учебная аудитория для проведения занятий практического типа, учебная аудитория для проведения занятий лабораторного типа, помещение для курсового

Согласована на портале № 63028

проектирования (выполнения курсовых работ), помещение для проведения групповых и индивидуальных консультаций; 634034, Томская область, г. Томск, Вершинина улица, д. 74, 407 ауд.

Описание имеющегося оборудования:

- Интерактивный плоскопанельный дисплей SMART VIZION DC75-E4;
- Магнитно-маркерная доска;
- Комплект специализированной учебной мебели;
- Рабочее место преподавателя.

Программное обеспечение:

- Google Chrome, Open Source;
- Microsoft Office 2013 Standard;
- Python, свободно распространяемое ПО совместимое с GNU GPL;
- Windows 10 Professional, AZURE DEV TOOLS FOR TEACHING;

8.3. Материально-техническое и программное обеспечение для самостоятельной работы

Для самостоятельной работы используются учебные аудитории (компьютерные классы), расположенные по адресам:

- 634050, Томская область, г. Томск, Ленина проспект, д. 40, 233 ауд.;
- 634045, Томская область, г. Томск, ул. Красноармейская, д. 146, 201 ауд.;
- 634034, Томская область, г. Томск, Вершинина улица, д. 47, 126 ауд.;
- 634034, Томская область, г. Томск, Вершинина улица, д. 74, 207 ауд.

Описание имеющегося оборудования:

- учебная мебель;
- компьютеры;
- компьютеры подключены к сети «Интернет» и обеспечивают доступ в электронную информационно-образовательную среду ТУСУРа.

Перечень программного обеспечения:

- Microsoft Windows;
- OpenOffice;
- Kaspersky Endpoint Security 10 для Windows;
- 7-Zip;
- Google Chrome.

8.4. Материально-техническое обеспечение дисциплины для лиц с ограниченными возможностями здоровья и инвалидов

Освоение дисциплины лицами с ограниченными возможностями здоровья и инвалидами осуществляется с использованием средств обучения общего и специального назначения.

При занятиях с обучающимися с **нарушениями слуха** предусмотрено использование звукоусиливающей аппаратуры, мультимедийных средств и других технических средств приема/передачи учебной информации в доступных формах, мобильной системы преподавания для обучающихся с инвалидностью, портативной индукционной системы. Учебная аудитория, в которой занимаются обучающиеся с нарушением слуха, оборудована компьютерной техникой, аудиотехникой, видеотехникой, электронной доской, мультимедийной системой.

При занятиях с обучающимися с **нарушениями зрения** предусмотрено использование в лекционных и учебных аудиториях возможности просмотра удаленных объектов (например, текста на доске или слайда на экране) при помощи видеоувеличителей для комфортного просмотра.

При занятиях с обучающимися с **нарушениями опорно-двигательного аппарата** используются альтернативные устройства ввода информации и другие технические средства приема/передачи учебной информации в доступных формах, мобильной системы обучения для людей с инвалидностью.

9. Оценочные материалы и методические рекомендации по организации изучения дисциплины

9.1. Содержание оценочных материалов для текущего контроля

Согласована на портале № 63028

и промежуточной аттестации

Для оценки степени сформированности и уровня освоения закрепленных за дисциплиной компетенций используются оценочные материалы, представленные в таблице 9.1.

Таблица 9.1 – Формы контроля и оценочные материалы

Названия разделов (тем) дисциплины	Формируемые компетенции	Формы контроля	Оценочные материалы (ОМ)
1 Введение.	ОПК-7	Зачёт с оценкой	Перечень вопросов для зачета с оценкой
		Устный опрос / собеседование	Примерный перечень вопросов для устного опроса / собеседования
		Лабораторная работа	Темы лабораторных работ
		Тестирование	Примерный перечень тестовых заданий
2 Основные принципы и этапы анализа данных.	ОПК-7	Зачёт с оценкой	Перечень вопросов для зачета с оценкой
		Устный опрос / собеседование	Примерный перечень вопросов для устного опроса / собеседования
		Лабораторная работа	Темы лабораторных работ
		Тестирование	Примерный перечень тестовых заданий
3 Анализ данных может быть разным.	ОПК-7	Зачёт с оценкой	Перечень вопросов для зачета с оценкой
		Устный опрос / собеседование	Примерный перечень вопросов для устного опроса / собеседования
		Лабораторная работа	Темы лабораторных работ
		Тестирование	Примерный перечень тестовых заданий
4 Подготовка и очистка данных.	ОПК-7	Зачёт с оценкой	Перечень вопросов для зачета с оценкой
		Устный опрос / собеседование	Примерный перечень вопросов для устного опроса / собеседования
		Лабораторная работа	Темы лабораторных работ
		Тестирование	Примерный перечень тестовых заданий

Согласована на портале № 63028

5 Выявление зависимостей и связей в данных.	ОПК-7	Зачёт с оценкой	Перечень вопросов для зачета с оценкой
		Устный опрос / собеседование	Примерный перечень вопросов для устного опроса / собеседования
		Лабораторная работа	Темы лабораторных работ
		Тестирование	Примерный перечень тестовых заданий
6 Методы машинного обучения для обработки данных.	ОПК-7	Зачёт с оценкой	Перечень вопросов для зачета с оценкой
		Устный опрос / собеседование	Примерный перечень вопросов для устного опроса / собеседования
		Лабораторная работа	Темы лабораторных работ
		Тестирование	Примерный перечень тестовых заданий
7 Графическое представление данных.	ОПК-7	Зачёт с оценкой	Перечень вопросов для зачета с оценкой
		Устный опрос / собеседование	Примерный перечень вопросов для устного опроса / собеседования
		Лабораторная работа	Темы лабораторных работ
		Тестирование	Примерный перечень тестовых заданий
8 Большие данные. Экосистема Hadoop.	ОПК-7	Зачёт с оценкой	Перечень вопросов для зачета с оценкой
		Тестирование	Примерный перечень тестовых заданий
9 Распределенные системы анализа данных.	ОПК-7	Зачёт с оценкой	Перечень вопросов для зачета с оценкой
		Тестирование	Примерный перечень тестовых заданий

Шкала оценки сформированности отдельных планируемых результатов обучения по дисциплине приведена в таблице 9.2.

Таблица 9.2 – Шкала оценки сформированности планируемых результатов обучения по дисциплине

Оценка	Баллы за ОМ	Формулировка требований к степени сформированности планируемых результатов обучения		
		знать	уметь	владеть
2 (неудовлетворительно)	< 60% от максимальной суммы баллов	отсутствие знаний или фрагментарные знания	отсутствие умений или частично освоенное умение	отсутствие навыков или фрагментарные применение навыков

Согласована на портале № 63028

3 (удовлетворительно)	от 60% до 69% от максимальной суммы баллов	общие, но не структурированные знания	в целом успешно, но не систематически осуществляемое умение	в целом успешное, но не систематическое применение навыков
4 (хорошо)	от 70% до 89% от максимальной суммы баллов	сформированные, но содержащие отдельные проблемы знания	в целом успешное, но содержащие отдельные пробелы умение	в целом успешное, но содержащие отдельные пробелы применение навыков
5 (отлично)	≥ 90% от максимальной суммы баллов	сформированные систематические знания	сформированное умение	успешное и систематическое применение навыков

Шкала комплексной оценки сформированности компетенций приведена в таблице 9.3.

Таблица 9.3 – Шкала комплексной оценки сформированности компетенций

Оценка	Формулировка требований к степени компетенции
2 (неудовлетворительно)	Не имеет необходимых представлений о проверяемом материале или Знать на уровне **ориентирования**, представлений. Обучающийся знает основные признаки или термины изучаемого элемента содержания, их отнесенность к определенной науке, отрасли или объектам, узнает в текстах, изображениях или схемах и знает, к каким источникам нужно обращаться для более детального его усвоения.
3 (удовлетворительно)	Знать и уметь на **репродуктивном** уровне. Обучающихся знает изученный элемент содержания репродуктивно: произвольно воспроизводит свои знания устно, письменно или в демонстрируемых действиях.
4 (хорошо)	Знать, уметь, владеть на **аналитическом** уровне. Зная на репродуктивном уровне, указывать на особенности и взаимосвязи изученных объектов, на их достоинства, ограничения, историю и перспективы развития и особенности для разных объектов усвоения.
5 (отлично)	Знать, уметь, владеть на **системном** уровне. Обучающийся знает изученный элемент содержания системно, произвольно и доказательно воспроизводит свои знания устно, письменно или в демонстрируемых действиях, учитывая и указывая связи и зависимости между этим элементом и другими элементами содержания дисциплины, его значимость в содержании дисциплины.

9.1.1. Примерный перечень тестовых заданий

1. Как с помощью NumPy найти ковариацию двух переменных?
2. Какова была точность прогноза на обучающей и тестовой выборке данных при выполнении части 3 лабораторной работы "Классификация с помощью дерева решений"?
3. Линейную регрессию в Python можно выполнить разными способами, назовите несколько?
4. Является ли линейная регрессия методом машинного обучения?
5. Какие библиотеки применяются для визуализации данных Python, назовите и кратко опишите пять таких библиотек.
6. Каков точный формат задания подписей к осям графика, наименования графика,

Согласована на портале № 63028

добавления легенды для выводимых на график величин?
7. Если переменные сравниваемой пары данных представляют корреляцию, близкую к нулю, то что это означает?
8. Какие статистические значения для датафрейма позволяет получить метод describe() ?
9. Как при выводе значения переменной в текстовой строке оставить только 2 цифры после десятичной точки, если исходное значение содержит больше цифр?
10. Как добавить в датафрейм новый столбец с данными?

9.1.2. Перечень вопросов для зачета с оценкой

1. У вас есть координаты расположения 10 АЗС в городской черте. Напишите программу отображения АЗС в сервисе OpenStreetMaps.
2. Вам дан датафрейм, содержащий записи в 10 колонках и 100 строках. Нужно удалить 7 и 9 колонку, и добавить две новые: одну со средним значением для строки, вторую с дисперсией в строке. Некоторые записи в исходном датафрейме содержат NaN. Как выполнить задачу с помощью Python?
3. Записи в датафрейме содержат URL веб-ресурсов. Как с помощью Python выделить для каждой записи только доменное имя первого уровня?
4. Мини-кластер Hadoop создан на одном PC. Какой выигрыш производительности при обработке задачи с хорошим параллелизмом можно получить, если запускать ее на таком кластере и применить алгоритм Map-Reduce?
5. Какие признаки должны иметь данные для отнесения их к категории "Большие данные"?

9.1.3. Примерный перечень вопросов для устного опроса / собеседования

1. Опишите методы загрузки данных в датафрейм в библиотеке Pandas.
2. Сравните возможности построения сложных и наглядных графиков в библиотеках Matplotlib, PyPlot и Plotly.
3. В каких случаях следует использовать Spark при обработке больших данных, а в каких Map-Reduce?
4. Чем архитектура потоковой обработки данных Kappa отличается от архитектуры Lambda?
5. Каковы минимальные аппаратные требования к кластеру Hadoop?

9.1.4. Темы лабораторных работ

1. Анализ криминальной обстановки в Сан-Франциско
2. Измерение и анализ данных Интернет-соединения
3. Измерение и анализ Интернет-соединения с хранилищем данных в виде SQL базы данных.
4. Описательная статистика в Python
5. Корреляционный анализ в Python
6. Простая линейная регрессия в Python
7. Классификация с помощью дерева решений
8. Обработка ошибок при использовании метода линейной регрессии
9. Измерение и анализ Интернет-соединения визуализация результатов.

9.2. Методические рекомендации

Учебный материал излагается в форме, предполагающей самостоятельное мышление студентов, самообразование. При этом самостоятельная работа студентов играет решающую роль в ходе всего учебного процесса.

Начать изучение дисциплины необходимо со знакомства с рабочей программой, списком учебно-методического и программного обеспечения. Самостоятельная работа студента включает работу с учебными материалами, выполнение контрольных мероприятий, предусмотренных учебным планом.

В процессе изучения дисциплины для лучшего освоения материала необходимо регулярно обращаться к рекомендуемой литературе и источникам, указанным в учебных материалах; пользоваться через кабинет студента на сайте Университета образовательными ресурсами электронно-библиотечной системы, а также общедоступными интернет-порталами, содержащими научно-популярные и специализированные материалы, посвященные различным аспектам

Согласована на портале № 63028

учебной дисциплины.

При самостоятельном изучении тем следуйте рекомендациям:

– чтение или просмотр материала осуществляйте со скоростью, достаточной для индивидуального понимания и освоения материала, выделяя основные идеи; на основании изученного составить тезисы. Освоив материал, попытаться соотнести теорию с примерами из практики;

– если в тексте встречаются незнакомые или малознакомые термины, следует выяснить их значение для понимания дальнейшего материала;

– осмысливайте прочитанное и изученное, отвечайте на предложенные вопросы.

Студенты могут получать индивидуальные консультации, в т.ч. с использованием средств телекоммуникации.

По дисциплине могут проводиться дополнительные занятия, в т.ч. в форме вебинаров. Расписание вебинаров и записи вебинаров публикуются в электронном курсе / электронном журнале по дисциплине.

9.3. Требования к оценочным материалам для лиц с ограниченными возможностями здоровья и инвалидов

Для лиц с ограниченными возможностями здоровья и инвалидов предусмотрены дополнительные оценочные материалы, перечень которых указан в таблице 9.2.

Таблица 9.2 – Дополнительные материалы оценивания для лиц с ограниченными возможностями здоровья и инвалидов

Категории обучающихся	Виды дополнительных оценочных материалов	Формы контроля и оценки результатов обучения
С нарушениями слуха	Тесты, письменные самостоятельные работы, вопросы к зачету, контрольные работы	Преимущественно письменная проверка
С нарушениями зрения	Собеседование по вопросам к зачету, опрос по терминам	Преимущественно устная проверка (индивидуально)
С нарушениями опорно-двигательного аппарата	Решение дистанционных тестов, контрольные работы, письменные самостоятельные работы, вопросы к зачету	Преимущественно дистанционными методами
С ограничениями по общемедицинским показаниям	Тесты, письменные самостоятельные работы, вопросы к зачету, контрольные работы, устные ответы	Преимущественно проверка методами, определяющимися исходя из состояния обучающегося на момент проверки

9.4. Методические рекомендации по оценочным материалам для лиц с ограниченными возможностями здоровья и инвалидов

Для лиц с ограниченными возможностями здоровья и инвалидов предусматривается доступная форма предоставления заданий оценочных средств, а именно:

– в печатной форме;

– в печатной форме с увеличенным шрифтом;

– в форме электронного документа;

– методом чтения ассистентом задания вслух;

– предоставление задания с использованием сурдоперевода.

Лицам с ограниченными возможностями здоровья и инвалидам увеличивается время на подготовку ответов на контрольные вопросы. Для таких обучающихся предусматривается доступная форма предоставления ответов на задания, а именно:

– письменно на бумаге;

– набор ответов на компьютере;

– набор ответов с использованием услуг ассистента;

– представление ответов устно.

Согласована на портале № 63028

Процедура оценивания результатов обучения лиц с ограниченными возможностями здоровья и инвалидов по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

– в форме электронного документа;

– в печатной форме увеличенным шрифтом.

Для лиц с нарушениями слуха:

– в форме электронного документа;

– в печатной форме.

Для лиц с нарушениями опорно-двигательного аппарата:

– в форме электронного документа;

– в печатной форме.

При необходимости для лиц с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения может проводиться в несколько этапов.

Согласована на портале № 63028

ЛИСТ СОГЛАСОВАНИЯ

Рассмотрена и одобрена на заседании кафедры АОИ
протокол № 6 от «10» 12 2020 г.

СОГЛАСОВАНО:

Должность	Инициалы, фамилия	Подпись
Заведующий выпускающей каф. АОИ	А.А. Сидоров	Согласовано, d4895b45-5d88-49f8-b7c7-e8bf0196776a
Заведующий обеспечивающей каф. АОИ	А.А. Сидоров	Согласовано, d4895b45-5d88-49f8-b7c7-e8bf0196776a
Начальник учебного управления	Е.В. Саврук	Согласовано, fa63922b-1fce-4a6a-845d-9ce7670b004c

ЭКСПЕРТЫ:

Доцент, каф. АОИ	Н.Ю. Салмина	Согласовано, ed28a52c-a209-461c-b4ed-4e958affbfc7
Заведующий кафедрой, каф. АОИ	А.А. Сидоров	Согласовано, d4895b45-5d88-49f8-b7c7-e8bf0196776a

РАЗРАБОТАНО:

Доцент, каф. АОИ	Е.Ю. Агеев	Разработано, 1380771b-dd3c-4ac1-8e1d-30fb96b5fa40

Согласована на портале № 63028