

УТВЕРЖДАЮ

Зам. директора по УВР  
УрСЭИ (филиал) ОУП ВО «АТиСО»

\_\_\_\_\_ О.В. Зубкова

« 10 » июня 2020 г.

**РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ  
(МОДУЛЯ)**

**Анализ данных и машинное обучение**  
(название дисциплины в соответствии с учебным планом)

**СПЕЦИАЛЬНОСТЬ СРЕДНЕГО ПРОФЕССИОНАЛЬНОГО  
ОБРАЗОВАНИЯ**

**09.02.05 Прикладная информатика (по отраслям)**  
(код профессии, специальности СПО)

**Техник-программист**  
(наименование квалификации)

Кафедра: Гуманитарных, естественнонаучных и математических дисциплин

Разработчики программы:

Челябинск -2020

## Оглавление

<b>1. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОЧЕЙ ПРОГРАММЫ УЧЕБНОЙ ДИСЦИПЛИНЫ</b> .....	3
1.1. Область применения рабочей программы учебного предмета, курса, дисциплины (модуля) .....	3
1.2. Цели и задачи учебной дисциплины .....	3
1.3. Место учебной дисциплины в структуре программы подготовки специалистов среднего звена .....	3
1.4. Требования к результатам освоения учебной дисциплины .....	3
<b>2. СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ</b> .....	3
2.1. Объем учебной дисциплины и виды учебной работы .....	3
2.2. Тематический план и содержание учебной дисциплины .....	4
<b>3. УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ УЧЕБНОЙ ДИСЦИПЛИНЫ</b> .....	6
3.1 Требования к минимальному материально-техническому обеспечению .....	6
3.2 Информационное обеспечение реализации программы .....	6
<b>4. КОНТРОЛЬ И ОЦЕНКА РЕЗУЛЬТАТОВ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ</b> .....	7
<b>5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ</b> .....	9
<b>6. ОЦЕНОЧНЫЕ СРЕДСТВА И КОНТРОЛЬНО-ИЗМЕРИТЕЛЬНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ ПО УЧЕБНОЙ ДИСЦИПЛИНЕ</b> .....	10
6.1 Оценочные средства для проведения текущего контроля по учебной дисциплине .....	10
6.2 Контрольно-измерительные материалы для проведения текущего контроля по учебной дисциплине .....	13
<b>7. ОЦЕНОЧНЫЕ СРЕДСТВА И КОНТРОЛЬНО-ИЗМЕРИТЕЛЬНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО УЧЕБНОЙ ДИСЦИПЛИНЕ</b> .....	16
7.1. Оценочные средства для проведения промежуточной аттестации .....	16
по учебной дисциплине .....	16
7.2. Контрольно-измерительные материалы для проведения промежуточной аттестации по учебной дисциплине .....	18
<b>8. ПРОВЕРКА СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ</b> .....	21

# 1. ОБЩАЯ ХАРАКТЕРИСТИКА РАБОЧЕЙ ПРОГРАММЫ УЧЕБНОЙ ДИСЦИПЛИНЫ

## 1.1. Область применения рабочей программы учебного предмета, курса, дисциплины (модуля)

Рабочая программа учебной дисциплины «Анализ данных и машинное обучение» является частью программы подготовки специалистов среднего звена в соответствии с ФГОС по специальности СПО 09.02.05 Прикладная информатика (по отраслям), квалификация Техник-программист.

## 1.2. Цели и задачи учебной дисциплины

Цель изучения учебной дисциплины: обучить студентов современным методам интеллектуального анализа данных и познакомить с основными алгоритмами машинного обучения, особенностями их применения.

Задачи изучения учебной дисциплины:

- изучить методы интеллектуального анализа данных;
- сформировать базовые понятия машинного обучения;
- сформировать практические навыки применения алгоритмов машинного обучения.

## 1.3. Место учебной дисциплины в структуре программы подготовки специалистов среднего звена

Учебная дисциплина «Анализ данных и машинное обучение» в программе подготовки специалистов среднего звена относится к дисциплинам общепрофессионального цикла (ОП.18) специальности 09.02.05 Прикладная информатика (по отраслям).

Учебная дисциплина «Анализ данных и машинное обучение» базируется на знаниях, умениях и навыках, полученных студентами при изучении дисциплины ОУДП.01 «Информатика».

Знания, умения и навыки, полученные студентами при изучении данной дисциплины, будут использованы при прохождении производственной и преддипломной практик.

## 1.4. Требования к результатам освоения учебной дисциплины

В результате освоения дисциплины обучающийся должен освоить следующие компетенции:

- ПК 1.1. Обработать статический информационный контент.
- ПК 1.2. Обработать динамический информационный контент.

В результате изучения дисциплины студент должен:

**знать:**

- 31 методы интеллектуального анализа данных (включая их преобразование и очистку, работу с пропущенными значениями, основные способы визуализации данных, корреляционный анализ, поиск нелинейных ассоциаций);
- 32 возможности алгоритмов машинного обучения;
- 33 классы задач, решаемых с помощью алгоритмов машинного обучения.

**уметь:**

- У1 проводить анализ данных
- У2 применять на практике алгоритмы машинного обучения;
- У3 обосновать применение того или иного алгоритма машинного обучения для решения конкретной задачи.

## 2. СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

### 2.1. Объем учебной дисциплины и виды учебной работы

Вид учебной работы	Объем часов 2г10м	Объем часов 3г10м
--------------------	----------------------	----------------------

<b>Максимальная учебная нагрузка (всего)</b>	<b>66</b>	<b>66</b>
<b>Обязательная аудиторная учебная нагрузка (всего)</b>	<b>44</b>	<b>44</b>
В том числе:		
теоретическое обучение	10	10
практические занятия	34	34
<b>Самостоятельная работа обучающегося (всего)</b>		
В том числе:		
работа по темам	10	10
подготовка докладов по темам	12	12
<b>Итоговые аттестации</b>	<b>Зачет 6 семестр</b>	<b>Зачет 8 семестр</b>

## 2.2. Тематический план и содержание учебной дисциплины

Наименование разделов и тем	Содержание учебного материала	Объем часов 2г10м/3г10м	Теоретические занятия	Практические занятия	Самостоятельная работа студента	Уровень освоения	Коды формируемых компетенций
<b>Тема 1. Большие данные и машинное обучение</b>	<b>Содержание учебного материала</b>	22	4	12	6	2	ОК 02, ОК 09
	<b>Лекционные занятия</b> Общая постановка задачи машинного обучения. Обучение с учителем и без учителя, обучение с подкреплением. Задачи классификации, восстановления регрессии, предсказания. Модели алгоритмов. Признаки. Типы признаков. Понятие функционала качества. Вероятностная постановка задачи. Оценка обобщающей способности. Проблема переобучения. Критерии оценки качества работы алгоритмов машинного обучения. ROC-кривые. Примеры практических задач машинного обучения. <b>Практические занятия (лабораторные работы)</b> Первичный анализ данных с библиотекой Pandas. Визуальный анализ данных с библиотеками Seaborn и Matplotlib. <b>Самостоятельная работа</b> Изучение теоретического материала, подготовка докладов						
<b>Тема 2. Метрические методы классификации</b>	<b>Содержание учебного материала</b>	17	2	10	5	2	ОК 02, ОК 09
	<b>Лекционные занятия</b> Обобщенный метрический классификатор. Виды метрик. Метод ближайшего соседа. Алгоритм k-ближайших соседей. Взвешенная версия алгоритма k-ближайших соседей. Метод окна Парзена. Метод потенциальных функций. Понятие эталона. Отступы и классификация объектов. <b>Практические занятия (лабораторные работы)</b> Классификация, деревья решений и метод						

	ближайших соседей <b>Самостоятельная работа</b> Изучение теоретического материала , подготовка докладов						
<b>Тема 3.</b> <b>Логические</b> <b>методы</b> <b>классификации</b>	<b>Содержание учебного материала</b>	13	2	6	5	2	ОК 02, ОК 09
	<b>Лекционные занятия</b> Основные принципы построения логических алгоритмов классификации. Критерии информативности: простые критерии, статистический критерий, энтропийный критерий. Алгоритм построения дерева ID3 <b>Практические занятия (лабораторные работы)</b> Логистическая регрессия и случайный лес. Линейная регрессия, Lasso и RF-регрессия. Построение и отбор признаков. <b>Самостоятельная работа</b> Изучение теоретического материала , подготовка докладов						
<b>Тема 4.</b> <b>Линейные</b> <b>методы</b> <b>классификации</b>	<b>Содержание учебного материала</b>	14	2	6	6	2	ОК 02, ОК 09
	<b>Лекционные занятия</b> Постановка задач линейной регрессии и линейной классификации. Метод наименьших квадратов в матричной форме. Аналитическое решение. Регуляризация в задач регрессии. Мультиколлинеарность и плохая обусловленность ковариационной матрицы. Гребневая регрессия. Метод лассо. Линейные классификаторы. Метод стохастического градиента. Улучшение сходимости метода SGD. Логистическая регрессия. Метод опорных векторов. Линейно разделимые выборки. Двойственная задача. Нелинейные обобщения. Возможные виды ядер. <b>Практические занятия (лабораторные работы)</b> Линейные модели классификации и регрессии. Логистическая регрессия и случайный лес. Линейная регрессия, Lasso и RF-регрессия. Построение и отбор признаков. <b>Самостоятельная работа</b> Изучение теоретического материала , подготовка докладов						
<b>Всего</b>		66	10	34	22		

Для характеристики уровня освоения учебного материала используются следующие обозначения:

1. – **ознакомительный или минимальный уровень** (узнавание ранее изученных объектов, свойств);
2. – **репродуктивный или базовый уровень** (выполнение деятельности по образцу, инструкции или под руководством)
3. – **продуктивный или высокий уровень (планирование и самостоятельное выполнение деятельности, решение проблемных задач)**

### **3. УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ УЧЕБНОЙ ДИСЦИПЛИНЫ**

#### **3.1 Требования к минимальному материально-техническому обеспечению**

Для реализации программы учебной дисциплины предусмотрены оборудованные помещения.

Основное оборудование учебной аудитории для лекционных занятий:

- рабочие места обучающихся;
- рабочее место преподавателя;
- маркерная (или меловая) доска.
- мультимедийное оборудование.

Программное обеспечение:

- ОС Microsoft Windows;
- Пакет приложений Microsoft Office (Open Office, Libre Office).

Основное оборудование учебной аудитории для практических (лабораторных) занятий:

- рабочие места обучающихся;
- автоматизированные рабочие места обучающихся;
- рабочее место преподавателя;
- маркерная (или меловая) доска.
- мультимедийное оборудование.

Программное обеспечение:

- ОС Microsoft Windows;
- Пакет приложений Microsoft Office (Open Office, Libre Office).
- Microsoft Visual Studio;
- СУБД: SQL Server, MySQL, PostgreSQL;
- Notepad++;
- Git;
- Microsoft Visio (DIA).

#### **3.2. Информационное обеспечение реализации программы**

##### **Основная литература**

1. Кобелев, О.А. Электронная коммерция : учебное пособие / О.А. Кобелев ; под ред. С.В. Пирогова. – Москва : Дашков и К°, 2018. – 684 с. : ил. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=496127>. – Библиогр. в кн. – ISBN 978-5-394-01738-4. – Текст : электронный.
2. Вальке, А.А. Электронные средства сбора и обработки информации : учебное пособие / А.А. Вальке, В.А. Захаренко ; Минобрнауки России, Омский государственный технический университет. – Омск : Омский государственный технический университет (ОмГТУ), 2017. – 112 с. : табл., схем., ил. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=493448>. – Библиогр. в кн. – ISBN 978-5-8149-2519-0. – Текст : электронный.
3. Рябошапка, Б.В. Модели принятия решений при проектировании систем сбора данных : учебное пособие : [16+] / Б.В. Рябошапка ; Южный федеральный университет. – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2019. – 98 с. : ил. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=577904>. – Библиогр. в кн. – ISBN 978-5-9275-3179-0. – Текст : электронный.

##### **Дополнительная литература**

1. Родыгин, А.В. Информатика. MS Office : учебное пособие : [16+] / А.В. Родыгин ; Новосибирский государственный технический университет. – Новосибирск : Новосибирский государственный технический университет, 2018. – 95 с. : табл., ил. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=573861>. – Библиогр. в кн. – ISBN 978-5-7782-3638-7. – Текст : электронный

2. Киселев, Г.М. Информационные технологии в педагогическом образовании : учебник / Г.М. Киселев, Р.В. Бочкова. – 3-е изд., стер. – Москва : Дашков и К°, 2020. – 304 с. : ил. – (Учебные издания для бакалавров). – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=573270>. – Библиогр.: с. 297 - 299. – ISBN 978-5-394-03468-8. – Текст : электронный.

3. Келлехер, Д. Наука о данных: базовый курс : [16+] / Д. Келлехер, Б. Тирни ; науч. ред. З. Мамедьяров ; пер. с англ. М. Белоголовского. – Москва : Альпина Паблишер, 2020. – 224 с. : схем., табл. – Режим доступа: по подписке. – URL: <http://biblioclub.ru/index.php?page=book&id=598235>. – ISBN 978-5-9614-3170-4. – Текст : электронный.

Ресурсы информационно-телекоммуникационной сети "Интернет", необходимые для освоения дисциплины (модуля)

№ п/п	Интернет ресурс (адрес)	Описание ресурса
1.	<a href="http://citforum.ru/">http://citforum.ru/</a>	IT-портал «Сервер Информационных Технологий»;
2.	<a href="https://habrahabr.ru/">https://habrahabr.ru/</a>	ресурс для IT-специалистов
3.	<a href="http://stackoverflow.com/">http://stackoverflow.com/</a>	сайт вопросов и ответов для IT-специалистов;
4.	<a href="http://www.firststeps.ru">http://www.firststeps.ru</a>	Первые шаги – Сайт, посвященный начинающим программистам. Учебники и инструкции для по языкам программирования, алгоритмам и используемым протоколам. Вопросы безопасности.
5.	<a href="http://www.intuit.ru">http://www.intuit.ru</a>	Интернет-университет информационных технологий (ИНТУИТ)

#### 4. КОНТРОЛЬ И ОЦЕНКА РЕЗУЛЬТАТОВ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

<i>Результаты обучения, подлежащие проверке</i>	<i>Критерии оценки</i>	<i>Формы и методы оценки</i>
<p><i>Перечень знаний, осваиваемых в рамках дисциплины:</i></p> <p>31 методы интеллектуального анализа данных (включая их преобразование и очистку, работу с пропущенными значениями, основные способы визуализации данных, корреляционный анализ, поиск нелинейных ассоциаций);</p> <p>32 возможности алгоритмов машинного обучения;</p> <p>33 классы задач, решаемых с помощью алгоритмов машинного обучения.</p> <p><i>Перечень умений, осваиваемых в рамках дисциплины:</i></p> <p>У1 проводить анализ данных</p> <p>У2 применять на практике</p>	<p>«Отлично» - теоретическое содержание курса освоено полностью, без пробелов, умения сформированы, все предусмотренные программой учебные задания выполнены, качество их выполнения оценено высоко.</p> <p>«Хорошо» - теоретическое содержание курса освоено полностью, без пробелов, некоторые умения сформированы недостаточно, все предусмотренные программой учебные задания выполнены, некоторые виды заданий выполнены с ошибками.</p> <p>«Удовлетворительно» - теоретическое содержание курса освоено частично, но пробелы не</p>	<p>Письменный и устный опросы</p> <p>Тестирование</p> <p>Доклады</p> <p>Решение задач</p>

<p>алгоритмы машинного обучения; УЗ обосновать применение того или иного алгоритма машинного обучения для решения конкретной задачи.</p>	<p>носят существенного характера, необходимые умения работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые из выполненных заданий содержат ошибки. «Неудовлетворительно» - теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки.</p>	
<p>В результате освоения дисциплины обучающийся должен освоить следующие <b>компетенции:</b> ПК 1.1. Обработать статический информационный контент. ПК 1.2. Обработать динамический информационный контент.</p>		

#### Описание показателей и критериев оценивания компетенций

Показатели оценивания	Критерии оценивания компетенций	Шкала оценивания
Понимание смысла компетенции	<p>Имеет базовые общие знания в рамках диапазона выделенных задач (1 балл) Понимает факты, принципы, процессы, общие понятия в пределах области исследования. В большинстве случаев способен выявить достоверные источники информации, обработать, анализировать информацию. (2 балла) Имеет фактические и теоретические знания в пределах области исследования с пониманием границ применимости (3 балла)</p>	<p>Минимальный уровень  Базовый уровень  Высокий уровень</p>
Освоение компетенции в рамках изучения учебной дисциплины	<p>Наличие основных умений, требуемых для выполнения простых задач. Способен применять только типичные, наиболее часто встречающиеся приемы по конкретной сформулированной (выделенной) задаче (1 балл) Имеет диапазон практических умений, требуемых для решения определенных проблем в области исследования. В большинстве случаев способен выявить достоверные источники информации, обработать, анализировать информацию. (2 балла) Имеет широкий диапазон практических умений, требуемых для развития творческих решений, абстрагирования проблем. Способен выявлять проблемы и умеет находить способы решения, применяя современные методы и технологии. (3 балла)</p>	<p>Минимальный уровень  Базовый уровень  Высокий уровень</p>
Способность применять на практике знания,	<p>Способен работать при прямом наблюдении. Способен применять теоретические знания к решению конкретных задач. (1 балл)</p>	<p>Минимальный уровень</p>



полученные в ходе изучения дисциплины	Может взять на себя ответственность за завершение задач в исследовании, приспособливает свое поведение к обстоятельствам в решении проблем. Затрудняется в решении сложных, неординарных проблем, не выделяет типичных ошибок и возможных сложностей при решении той или иной проблемы (2 балла)	Базовый уровень
	Способен контролировать работу, проводить оценку, совершенствовать действия работы. Умеет выбрать эффективный прием решения задач по возникающим проблемам. (3 балла)	Высокий уровень

## **5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

### Методические указания по выполнению лекционных занятий

Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; пометить важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначить вопросы, термины, материал, который вызывает трудности, пометить и попытаться найти ответ в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на консультации, на практическом занятии.

### Методические указания по выполнению практических занятий

Проработка рабочей программы, уделяя особое внимание целям и задачам структуре и содержанию дисциплины. Конспектирование источников. Работа с конспектом лекций, подготовка ответов к контрольным вопросам, просмотр рекомендуемой литературы, работа с текстом (указать текст из источника и др.). Прослушивание аудио- и видеозаписей по заданной теме, решение расчетно-графических заданий, решение задач по алгоритму и др.

Методические указания по выполнению практических работ/индивидуальных заданий

Знакомство с основной и дополнительной литературой, включая справочные издания, зарубежные источники, конспект основных положений, терминов, сведений, требующих для запоминания и являющихся основополагающими в этой теме. Составление аннотаций к прочитанным литературным источникам и др.

### Методические указания по выполнению докладов

Доклад: Поиск литературы и составление библиографии, использование от 3 до 5 научных работ, изложение мнения авторов и своего суждения по выбранному вопросу; изложение основных аспектов проблемы.

Доклад по теме - это изучение научной, учебной, нормативной и другой литературы. Отбор необходимого материала; формирование выводов и разработка конкретных рекомендаций по решению поставленной цели и задачи; проведение практических исследований по данной проблеме или вопросу.

### Методические указания по подготовке к зачету.

Наиболее ответственным этапом в обучении студентов является экзаменационная сессия. На ней студенты отчитываются о выполнении учебной программы, об уровне и объеме полученных знаний. Это государственная отчетность студентов за период обучения, за изучение учебной дисциплины.

**Залогом успешного прохождения контроля являются систематические, добросовестные занятия студента. Однако это не исключает необходимости специальной работы перед сессией и в период сдачи зачета. Специфической задачей студента в период экзаменационной сессии являются повторение, обобщение и систематизация всего материала.**

**В процессе повторения анализируются и систематизируются все знания, накопленные при изучении программного материала: данные учебника, записи лекций, конспекты прочитанных книг, заметки, сделанные во время консультаций или семинаров, и др.**

**Консультации, которые проводятся для студентов в период экзаменационной сессии, необходимо использовать для углубления знаний, для восполнения пробелов и для разрешения всех возникших трудностей.**

При подготовке к контролю необходимо еще раз проверить себя на предмет усвоения основных категорий и ключевых понятий курса.

## **6. ОЦЕНОЧНЫЕ СРЕДСТВА И КОНТРОЛЬНО-ИЗМЕРИТЕЛЬНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ ПО УЧЕБНОЙ ДИСЦИПЛИНЕ**

### **6.1. Оценочные средства для проведения текущего контроля по учебной дисциплине**

#### **Тема 1. Большие данные и машинное обучение (ПК-1.1, ПК-1.2)**

*Вопросы к обсуждению:*

1. Общая постановка задачи машинного обучения.
2. Обучение с учителем и без учителя, обучение с подкреплением.
3. Задачи классификации, восстановления регрессии, предсказания.
4. Модели алгоритмов.
5. Признаки. Типы признаков.
6. Понятие функционала качества.
7. Вероятностная постановка задачи.
8. Оценка обобщающей способности.
9. Проблема переобучения.
10. Критерии оценки качества работы алгоритмов машинного обучения. ROC-кривые.
11. Примеры практических задач машинного обучения.

*Практические занятия (лабораторные работы):*

1. Первичный анализ данных с библиотекой Pandas.
2. Визуальный анализ данных с библиотеками Seaborn и Matplotlib.

Лабораторная работа. Первичный анализ данных с библиотекой Pandas (предобработка данных в Pandas).

План:

- 1.Использование Python для анализа данных.
- 2.Основные библиотеки Python: Scikit-learn, NumPy, SciPy, matplotlib, pandas.
- 3.Дистрибутив Anaconda.
- 4.Pandas: базовые методы.
- 5.Pandas: индексация и извлечение данных.
- 6.Pandas: применение функций к ячейкам, столбцам и строкам.
- 7.Pandas: группировка данных.
- 8.Pandas: таблицы сопряженности.
- 9.Pandas: сводные таблицы.
- 10.DataFrame в Pandas.

Вопросы:

- 1.Почему язык Python?
- 2.Для чего используется библиотека NumPy?
- 3.Для чего используются библиотеки SciPy, matplotlib, pandas?
- 4.Как выполняется индексация и извлечение данных в Pandas?

5. Как выполняется группировка данных в Pandas?
6. Для чего применяются таблицы сопряженности в Pandas?
7. Как выполняется загрузка данных в DataFrame библиотеке Pandas?
8. Как выполняется доступ к столбцам DataFrame библиотеке Pandas?

*Самостоятельная работа:* изучение теоретического материала, подготовка докладов.

## **Тема 2. Расчетные объекты конфигурации. Регистры расчета (ПК-1.1, ПК-1.2)**

*Вопросы к обсуждению:*

1. Обобщенный метрический классификатор.
2. Виды метрик.
3. Метод ближайшего соседа.
4. Алгоритм k-ближайших соседей.
5. Взвешенная версия алгоритма k-ближайших соседей.
6. Метод окна Парзена. Метод потенциальных функций.
7. Понятие эталона.
8. Отступы и классификация объектов.

*Практические занятия (лабораторные работы):* Классификация, деревья решений и метод ближайших соседей

Лабораторная работа. Метрические методы классификации в Scikit-learn.

План:

1. Метрические методы классификации.
2. Признаковые описания объекта.
3. Гипотеза компактности.
4. Метрики, виды метрик.
5. Весовая Евклидова метрика, метрика Минковского.
6. Масштабирование признаков.
7. Метод k ближайших соседей.
8. Реализация kNN в `sklearn.neighbors.KNeighborsClassifier`.
9. Кросс-валидация.
10. Алгоритм выполнения кросс-валидации по блокам.
11. Вычисление ошибки на разбиениях.

Вопросы:

1. В чем идея гипотезы компактности?
2. В чем состоит смысл обучения в метрических методах?
3. Для чего используется библиотека Scikit-learn?
4. Для чего выполняют масштабирование признаков?
5. Как обычно выполняется масштабирование количественных признаков?
6. В каком классе Scikit-learn реализован метод kNN?
7. Какой параметр метода k ближайших соседей, задает число соседей для построения прогноза?
8. В чем смысл кросс-валидации?
9. Как вычисляется весовая Евклидова метрика?
10. Приведите формулу Метрики Минковского. Что является ее параметром?

*Самостоятельная работа:* изучение теоретического материала, подготовка докладов.

## **Тема 3. Логические методы классификации (ПК-1.1, ПК-1.2)**

*Вопросы к обсуждению:*

1. Основные принципы построения логических алгоритмов классификации.
2. Критерии информативности: простые критерии, статистический критерий, энтропийный критерий.
3. Алгоритм построения дерева ID3

*Практические занятия (лабораторные работы):*

1. Логистическая регрессия и случайный лес.
2. Линейная регрессия, Lasso и RF-регрессия.
3. Построение и отбор признаков

Лабораторная работа. Деревья решений. Важность признаков

План:

1. Логическая закономерность.
2. Основные вопросы построения логических алгоритмов классификации.
3. Определение бинарного решающего дерева.
4. Реализация решающих деревьев в библиотеке scikit-learn.
5. Важность признаков.
6. Пропуски в данных.

Вопросы:

1. В каких классах scikit-learn реализуются решающие деревья для задач классификации и регрессии?
2. С помощью какой функции scikit-learn реализуется обучение модели решающих деревьев?
3. Какая переменная содержит массив "важностей" признаков?
4. С помощью какой функции можно проверить, является ли число `nan`?
5. Основные вопросы построения логических алгоритмов классификации.
6. Определение бинарного решающего дерева.
7. Реализация решающих деревьев в библиотеке scikit-learn.
8. Важность признаков

*Самостоятельная работа:* изучение теоретического материала, подготовка докладов.

#### **Тема 4. Перерасчеты (ПК-1.1, ПК-1.2)**

*Вопросы к обсуждению:*

1. Зависимость по перерасчету.
2. Стандартная табличная часть «Ведущие Виды Расчета».
3. Объект конфигурации «Перерасчет».
4. Объектная модель работы с перерасчетом.
5. Запрос для получения необходимых данных записей, подлежащих перерасчету

*Практические занятия (лабораторные работы):*

1. Линейные модели классификации и регрессии.
2. Логистическая регрессия и случайный лес.
3. Линейная регрессия, Lasso и RF-регрессия.
4. Построение и отбор признаков.

Лабораторная работа. Линейная классификация. Нормализация признаков.

План:

1. Линейные алгоритмы классификации.
2. Перцептрон.
3. Нормализация признаков. Стандартизация признаков.
4. Реализация линейных классификаторов в библиотеке scikit-learn.

5. Метрика качества.
6. Метод опорных векторов.
7. Опорные объекты.

Вопросы:

1. Сформулируйте постановку задачи линейной классификации.
2. Как выполняется стандартизация признаков?
3. В каком классе scikit-learn реализуется перцептрон?
4. Для чего используется функция sklearn.metrics.accuracy\_score?
5. Каким классом удобно воспользоваться для стандартизации признаков?
8. На что направлен функционал, который он оптимизирует метод опорных векторов? Какие объекты называют опорными

*Самостоятельная работа:* изучение теоретического материала, подготовка докладов.

### Примерные темы докладов

1. Онтологии и онтологические системы. Модели онтологии и онтологической системы.
2. Задача классификации. Методы построения деревьев решения. Методика «разделяй и властвуй».
3. Задача классификации. Методы построения деревьев решений. Алгоритм покрытия.
4. Задача классификации. Методы построения деревьев решений. Алгоритм ID 3.
5. Задача классификации. Методы построения деревьев решений. Алгоритм C4.5
6. Задача классификации. Методы построения правил классификации. Алгоритм построения 1 –правил
7. Задача классификации. Методы построения правил классификации. Метод NaiveBayes.
8. Задача кластеризации. Алгоритм k-means.
9. Задача кластеризации. Алгоритм Fuzzy C-Means.
10. Задача кластеризации. Алгоритм EM.
11. Информационный поиск в текстах. InformationRetrieval.
12. Поиск ассоциативных правил. Алгоритм Apriori.
13. Секвенциальный анализ. Алгоритм AprioriALL.
14. Секвенциальный анализ. Алгоритм GSP.
15. Методы кластеризации текстовых документов.
16. Задача аннотирования текстов. Методы извлечения фрагментов для аннотации.
17. Преобразование MTF.
18. Алгоритм сжатия BWT.
19. Алгоритм построения 1-правил.
20. Метод Naive-Base.
21. Прогнозирование временных рядов.

### 6.2. Контрольно-измерительные материалы для проведения текущего контроля по учебной дисциплине

#### Шкала оценки для проведения текущего контроля по учебной дисциплине в устной форме

№ п/п	Оценка за ответ	Характеристика ответа
1	Отлично	<ul style="list-style-type: none"> <li>- полно раскрыто содержание материала;</li> <li>- материал изложен грамотно, в определенной логической последовательности;</li> <li>- точно используется терминология;</li> </ul>

		<ul style="list-style-type: none"> <li>- показано умение иллюстрировать теоретические положения конкретными примерами, применять их в новой ситуации;</li> <li>- продемонстрированные знания и умения позволяют самостоятельно решать поставленные задачи;</li> <li>- ответ прозвучал самостоятельно, без наводящих вопросов;</li> <li>- продемонстрирована способность творчески применять знание теории к решению профессиональных задач;</li> <li>- допущены одна - две неточности при освещении второстепенных вопросов, которые исправляются по замечанию.</li> <li>- количество баллов за освоение компетенций от 8 до 9</li> </ul>
2	Хорошо	<ul style="list-style-type: none"> <li>- вопросы излагаются систематизировано и последовательно;</li> <li>- продемонстрированные знания и умения позволяют самостоятельно решать поставленные задачи, однако требуют определенного контроля;</li> <li>- продемонстрировано умение анализировать материал, однако не все выводы носят аргументированный и доказательный характер;</li> <li>- ответ удовлетворяет в основном требованиям на оценку «отлично», но при этом имеет один из недостатков: в изложении допущены небольшие пробелы, не исказившие содержание ответа; приобретенный практический опыт, знания и умения требуют незначительной корректировки в процессе выполнения задания; допущены ошибка или более двух недочетов при освещении второстепенных вопросов, которые легко исправляются по замечанию преподавателя.</li> <li>- количество баллов за освоение компетенций от 5 до 7</li> </ul>
3	Удовлетворительно	<ul style="list-style-type: none"> <li>- неполно или непоследовательно раскрыто содержание материала, но показано общее понимание вопроса и продемонстрированы умения, достаточные для дальнейшего усвоения материала;</li> <li>- при неполном знании теоретического материала выявлен недостаточный уровень знаний и умений; студент не может применить теоретические знания на практике;</li> <li>- количество баллов за освоение компетенций от 3 до 4</li> </ul>
4	Неудовлетворительно	<ul style="list-style-type: none"> <li>- не раскрыто основное содержание учебного материала;</li> <li>- обнаружено незнание или непонимание большей или наиболее важной части учебного материала;</li> <li>- допущены ошибки в определении понятий, при использовании терминологии, которые не исправлены после нескольких наводящих вопросов</li> <li>- отсутствуют практический опыт, знания и умения по предлагаемым ситуационным вопросам или задачам, количество баллов за освоение компетенций менее 3.</li> <li>- отказ от ответа или отсутствие ответа</li> </ul>

**Шкала оценки для проведения текущего контроля по учебной дисциплине в письменной форме**

№ п/п	Оценка за ответ	Характеристика ответа
1	Отлично	Материал раскрыт полностью, изложен логично, без существенных ошибок, выводы доказательны и опираются на

		теоретические знания Количество баллов за освоение материала от 8 до 9
2	Хорошо	Основные положения раскрыты, но в изложении имеются незначительные ошибки выводы доказательны, но содержат отдельные неточности Количество баллов за освоение материала от 5 до 7
3	Удовлетворительно	Изложение материала не систематизированное, выводы недостаточно доказательны, аргументация слабая. Количество баллов за освоение материала от 3 до 4
4	Неудовлетворительно	Не раскрыто основное содержание материала, обнаружено не знание основных положений темы. Не сформированы компетенции, умения и навыки. Количество баллов за освоение компетенций менее 3 Ответ на вопрос отсутствует

### Критерий оценки докладов

Критерий	Требования к докладу
Знание и понимание теоретического материала	- рассматриваемые понятия определяются четко и полно, приводятся соответствующие примеры, - используемые понятия строго соответствуют теме, - самостоятельность выполнения работы
Анализ и оценка информации	- грамотно применяется категория анализа, - методологически верно проведены расчеты показателей; - умело используются приемы сравнения и обобщения для анализа взаимосвязи понятий и явлений, - обоснованно интерпретируется текстовая информация, - дается личная оценка проблеме
Построение суждений	- изложение ясное и четкое, - приводимые доказательства логичны - выдвинутые тезисы сопровождаются грамотной аргументацией, - приводятся различные точки зрения и их личная оценка (при необходимости), - общая форма изложения полученных результатов и их интерпретации соответствует жанру проблемной научной статьи
Оценка	Критерии оценки доклада
«отлично»	1) студент легко ориентируется в содержании теоретического и аналитического материала, свободно пользуется понятийным аппаратом, обладает умением связывать теорию с практикой, высказывать и обосновывать свои суждения; 2) знает и правильно применяет формулы; 3) знает и правильно применяет нормативные документы; 4) решение аналитического (практического) задания записано понятно, аккуратно, последовательно; 5) подготовлен презентационный материал.
«хорошо»	1) студент демонстрирует полное освоение теоретического и аналитического материала, владеет понятийным аппаратом, ориентируется в изученном материале, осознанно применяет знания для решения практических задач, грамотно излагает свою позицию; 2) знает и применяет формулы и нормативные документы, но допускает небольшие неточности; 3) решение аналитического (практического) задания записано, но

	недостаточно аргументировано; 4) подготовлен презентационный материал, но недостаточно полный.
«удовлетворительно»	1) студент демонстрирует неполное освоение теоретического и аналитического материала, плохо владеет понятийным аппаратом, плохо ориентируется в изученном материале, неуверенно излагает свою позицию; 2) знает отдельные формулы и нормативные документы, но допускает значительные неточности в их применении; 3) решение аналитического (практического) задания записано неверно, аргументация отсутствует; 4) не подготовлен презентационный материал.
«неудовлетворительно»	1) студент имеет разрозненные, бессистемные знания, не умеет выделять главное и второстепенное, допускает ошибки в определении понятий, искажающие их смысл; 2) беспорядочно и неуверенно излагает материал, не может применять знания для решения практических задач; 3) решение аналитического (практического) задания записано неверно либо отсутствует; 4) не подготовлен презентационный материал.

## **7. ОЦЕНОЧНЫЕ СРЕДСТВА И КОНТРОЛЬНО-ИЗМЕРИТЕЛЬНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО УЧЕБНОЙ ДИСЦИПЛИНЕ**

### **7.1. Оценочные средства для проведения промежуточной аттестации по учебной дисциплине**

Контрольные вопросы для подготовки к контролю при проведении промежуточной аттестации по учебной дисциплине:

1. Основные понятия – информация, данные, знания. Виды информации. Обработка данных и ее виды. Data Mining. Классификация задач Data Mining.
2. Модели процессов обработки данных. Модель: конечные автоматы.
3. Модели процессов обработки данных. Модель: сети Петри.
4. Задачи обработки данных различных типов. Прикладные области обработки данных. Оцифровка сигналов. Теорема Котельникова.
5. Базы данных. OLTP–системы. Неэффективность OLTP для анализа данных. Определение свойства хранилищ данных.
6. Физические и виртуальные хранилища данных (ХД). Основные проблемы создания ХД.
7. Витрины данных.
8. Данные в хранилищах данных. ETL процесс.
9. Представление данных в виде гиперкуба. Операции над гиперкубом. Пример. Технология OLAP. Тест FASMI.
10. Многомерное представление данных и многомерный куб. Представление данных в виде гиперкуба. Пример.
11. Основные понятия гиперкубов (OLAP кубов). Структура OLAP куба. Операции над гиперкубом.
12. Архитектура OLAP. Компоненты OLAP. MOLAP, ROLAP, HOLAP.
13. Задача анализа текстов. Этапы анализа. Предобработка текста.
14. Извлечение ключевых понятий из текста.
15. Классификация текстовых документов. Методы классификации текстовых документов.
16. Большие данные. Свойства больших данных.
17. Машинное обучение, формализация задачи машинного обучения.



18. Признаковое описание объекта. Ответы и типы задач машинного обучения. Модель алгоритмов. Метод обучения. Этап обучения и этап применения.
19. Функционалы качества. Сведение задачи обучения к задаче оптимизации.
20. Переобучение и обобщение. Пример переобучения (Рунге). Эмпирические оценки обобщающей способности.
21. Примеры задач машинного обучения: задачи классификации.
22. Примеры задач машинного обучения: задачи регрессии.
23. Примеры задач машинного обучения: задача ранжирования.
24. Эксперименты в машинном обучении: эксперименты на реальных и синтетических данных.
25. Формализация метрической классификации. Обобщенный метрический классификатор.
26. Метод ближайшего соседа.
27. Метод квзвешенных ближайших соседей.
28. Метод парзеновского окна.
29. Метод потенциальных функций.
30. Отбор эталонных объектов. Понятие отступа объекта. Типы объектов в зависимости от отступа.
31. Отбор эталонов, алгоритм STOLP.
32. Логическая закономерность. Основы вопросы построения логических алгоритмов классификации. Виды закономерностей.
33. Критерии информативности: простые критерии, статистический критерий, энтропийный критерий. Схема локального поиска информативных закономерностей.
34. Определение бинарного решающего дерева. Жадный алгоритм построения дерева ID3.
35. Варианты критериев ветвления в ID3.
36. Алгоритм ID3: достоинства и недостатки.
37. Обработка пропусков в ID3, алгоритм обработки пропусков на этапе обучения и этапе классификации.
38. Стратегии редукции решающих деревьев.
39. Небрежные решающие деревья.
40. Бинаризация вещественного признака

### Итоговый тест

Вопросы	Контролируемые компетенции
1. При каком подходе к измерению информации используется тезаурусная мера? 1: Семантический; 2: Прагматический; 3: Статистический;	ПК-1.1, ПК-1.2
2. Какая операция над гиперкубом определяет переход от детального представления данных к агрегированному? 1. консолидация; 2. срез; 3. вращение;	ПК-1.1, ПК-1.2
3. Какое понятие определяет следующее высказывание «Множественная перспектива, состоящая из нескольких независимых измерений, вдоль которых могут быть проанализированы определенные совокупности данных»? 1. реляционная модель данных; 2. многомерное представление данных; 3. хранилище данных	ПК-1.1, ПК-1.2

<p>4.К какому типу задач машинного обучения относится задача предсказания цены жилья по его характеристикам?</p> <ol style="list-style-type: none"> <li>1. Классификация на два класса;</li> <li>2. Классификация на Мнепересекающихся классов;</li> <li>3. Классификация на Мпересекающихся классов;</li> <li>4. Восстановление регрессии</li> </ol>	ПК-1.1, ПК-1.2
<p>5.Ккакому типу относится признаки «Цвет глаз»?</p> <ol style="list-style-type: none"> <li>1: Бинарный;</li> <li>2: Количественный;</li> <li>3: Номинальный (категориальный);</li> <li>4: Порядковый;</li> </ol>	ПК-1.1, ПК-1.2
<p>6.К какому типу задач машинного обучения, относится задача, в которой необходимо определить независимые группы и их характеристики во всем множестве анализируемых данных?</p> <ol style="list-style-type: none"> <li>1. задача классификации;</li> <li>2. задача регрессии</li> <li>3. задача кластеризации;</li> </ol>	ПК-1.1, ПК-1.2
<p>7.Большие данные – это:</p> <ol style="list-style-type: none"> <li>1.Данные объемом более 1Тб</li> <li>2.Данные объемом более 10Тб</li> <li>3.Данные объемом более 100Тб</li> <li>4.Нет ограничений на минимальный объем</li> </ol>	ПК-1.1, ПК-1.2
<p>8.Наиболее редко на практике применяются методы машинного обучения, основанные на:</p> <ol style="list-style-type: none"> <li>1.Алгоритмах обучения без учителя</li> <li>2.Алгоритмах обучения с учителем</li> <li>3.Алгоритмах обучения с подкреплением</li> <li>4.Свёрточных нейронных сетях</li> </ol>	ПК-1.1, ПК-1.2
<p>9.Алгоритм k-средних предназначен для решения задачи:</p> <ol style="list-style-type: none"> <li>1.Классификации</li> <li>2.Кластеризации</li> <li>3.Прогнозирования</li> <li>4.Снижения размерности</li> </ol>	ПК-1.1, ПК-1.2
<p>10.Neo4j – это:</p> <ol style="list-style-type: none"> <li>1.База данных</li> <li>2.Архитектура нейронной сети</li> <li>3.Платформа распределенных вычислений</li> <li>4.Компилятор языка</li> </ol>	ПК-1.1, ПК-1.2
<p>11.Реализация метода обучения с учителем не нуждается в:</p> <ol style="list-style-type: none"> <li>1.Обучающей выборке</li> <li>2.Тестовой выборке</li> <li>3.Оценочной выборке</li> <li>4.Проверочной выборке</li> </ol>	ПК-1.1, ПК-1.2
<p>12.К какому типу задач машинного обучения, относится задача, в которой необходимо определить зависимости между объектами или событиями?</p> <ol style="list-style-type: none"> <li>1. задача распознавания образов;</li> <li>2. задача поиска ассоциативных правил;</li> <li>3. задача нормализации</li> </ol>	ПК-1.1, ПК-1.2

**7.2. Контрольно-измерительные материалы для проведения промежуточной аттестации по учебной дисциплине  
Шкала оценки для проведения промежуточной аттестации по учебной**

**дисциплине в устной форме**

№ п/п	Оценка за ответ	Характеристика ответа
1	Отлично	<ul style="list-style-type: none"> <li>- полно раскрыто содержание материала;</li> <li>- материал изложен грамотно, в определенной логической последовательности;</li> <li>- точно используется терминология;</li> <li>- показано умение иллюстрировать теоретические положения конкретными примерами, применять их в новой ситуации;</li> <li>- продемонстрированные знания и умения позволяют самостоятельно решать поставленные задачи;</li> <li>- ответ прозвучал самостоятельно, без наводящих вопросов;</li> <li>- продемонстрирована способность творчески применять знание теории к решению профессиональных задач;</li> <li>- допущены одна - две неточности при освещении второстепенных вопросов, которые исправляются по замечанию.</li> <li>- количество баллов за освоение компетенций от 8 до 9</li> </ul>
2	Хорошо	<ul style="list-style-type: none"> <li>- вопросы излагаются систематизировано и последовательно;</li> <li>- продемонстрированные знания и умения позволяют самостоятельно решать поставленные задачи, однако требуют определенного контроля;</li> <li>- продемонстрировано умение анализировать материал, однако не все выводы носят аргументированный и доказательный характер;</li> <li>- ответ удовлетворяет в основном требованиям на оценку «отлично», но при этом имеет один из недостатков: в изложении допущены небольшие пробелы, не искажившие содержание ответа; приобретенный практический опыт, знания и умения требуют незначительной корректировки в процессе выполнения задания; допущены ошибка или более двух недочетов при освещении второстепенных вопросов, которые легко исправляются по замечанию преподавателя.</li> <li>- количество баллов за освоение компетенций от 5 до 7</li> </ul>
3	Удовлетворительно	<ul style="list-style-type: none"> <li>- неполно или непоследовательно раскрыто содержание материала, но показано общее понимание вопроса и продемонстрированы умения, достаточные для дальнейшего усвоения материала;</li> <li>- при неполном знании теоретического материала выявлен недостаточный уровень знаний и умений; студент не может применить теоретические знания на практике;</li> <li>- количество баллов за освоение компетенций от 3 до 4</li> </ul>

4	Неудовлетворительно	<ul style="list-style-type: none"> <li>- не раскрыто основное содержание учебного материала;</li> <li>- обнаружено незнание или непонимание большей или наиболее важной части учебного материала;</li> <li>- допущены ошибки в определении понятий, при использовании терминологии, которые не исправлены после нескольких наводящих вопросов</li> <li>- отсутствуют практический опыт, знания и умения по предлагаемым ситуационным вопросам или задачам, количество баллов за освоение компетенций менее 3.</li> <li>- отказ от ответа или отсутствие ответа</li> </ul>
---	---------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Шкала оценки для проведения промежуточной аттестации по учебной дисциплине в письменной форме**

№ п/п	Оценка за ответ	Характеристика ответа
1	Отлично	Материал раскрыт полностью, изложен логично, без существенных ошибок, выводы доказательны и опираются на теоретические знания Количество баллов за освоение материала от 8 до 9
2	Хорошо	Основные положения раскрыты, но в изложении имеются незначительные ошибки выводы доказательны, но содержат отдельные неточности Количество баллов за освоение материала от 5 до 7
3	Удовлетворительно	Изложение материала не систематизированное, выводы недостаточно доказательны, аргументация слабая. Количество баллов за освоение материала от 3 до 4
4	Неудовлетворительно	Не раскрыто основное содержание материала, обнаружено незнание основных положений темы. Не сформированы компетенции, умения и навыки. Количество баллов за освоение компетенций менее 3 Ответ на вопрос отсутствует

**Шкала оценки в системе «зачтено – не зачтено»**

№ п/п	Оценка за ответ	Характеристика ответа
1	Зачтено	<p>Достаточный объем знаний в рамках изучения дисциплины. В ответе используется научная терминология. Стилистическое и логическое изложение ответа на вопрос правильное. Умеет делать выводы без существенных ошибок. Владеет инструментарием изучаемой дисциплины, умеет его использовать в решении стандартных (типовых) задач. Ориентируется в основных теориях, концепциях и направлениях по изучаемой дисциплине. Активен на практических (лабораторных) занятиях, допустимый уровень культуры исполнения заданий. Количество баллов за освоение компетенций от 3 до 9.</p>
2	Не зачтено	<p>Не достаточно полный объем знаний в рамках изучения дисциплины. В ответе не используется научная терминология. Изложение ответа на вопрос с существенными стилистическими и логическими ошибками. Не умеет делать выводы по результатам изучения дисциплины. Слабое владение инструментарием изучаемой дисциплины, не</p>

	компетентность в решении стандартных (типовых) задач. Не умеет ориентироваться в основных теориях, концепциях и направлениях по изучаемой дисциплине. Пассивность на практических (лабораторных) занятиях, низкий уровень культуры исполнения заданий. Не сформированы компетенции, умения и навыки. Количество баллов за освоение компетенций менее 3. Отказ от ответа или отсутствие ответа.
--	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

### Критерии формирования оценок по тестам

Оценка	Требования к знаниям
отлично	80%-100%
хорошо	65-80%
удовлетворительно	50-65%
неудовлетворительно	менее 50%
зачтено	50% и более
не зачтено	менее 50%

## 8. ПРОВЕРКА СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ

**ПК 1.1. Обработать статический информационный контент.**

**ПК 1.2. Обработать динамический информационный контент.**

*Вариант 1*

Задание 1. Анализ данных по доходу населения UCI Adult.

С помощью Pandas ответить на несколько вопросов по данным репозитория UCI Adult:

1. Каков средний возраст (признак age) женщин?
2. Какова доля граждан Германии (признак native-country)? Постройте гистограмму распределения (barplot) образования людей (признак education).
3. Каковы средние значения и среднеквадратичные отклонения возраста тех, кто получает более 50К в год (признак salary) и тех, кто получает менее 50К в год?
4. Правда ли, что люди, которые получают больше 50к, имеют как минимум высшее образование? (признак education -Bachelors, Prof-school, Assoc-acdm, Assoc-voc, Masters или Doctorate)
5. Выведите статистику возраста для каждой расы (признак race) и каждого пола. Используйте groupby и describe. Найдите таким образом максимальный возраст мужчин расы Amer-Indian-Eskimo. Среди кого больше доля зарабатывающих много (>50К): среди женатых или холостых мужчин (признак marital-status)? Женатыми считаем тех, у кого marital-status начинается с Married (Married-civ-spouse, Married-spouse-absent или Married-AF-spouse), остальных считаем холостыми.
6. Какое максимальное число часов человек работает в неделю (признак hours-per-week)? Сколько людей работают такое количество часов и каков среди них процент зарабатывающих много?
7. Посчитайте среднее время работы (hours-per-week) зарабатывающих мало и много (salary) для каждой страны (native-country).

Задание 2. Будем использовать набор данных Wine, где требуется предсказать сорт винограда, из которого изготовлено вино, используя результаты химических анализов. Нужно подобрать оптимальное значение k для алгоритма kNN.

Выполните следующие шаги:

Загрузите выборку Wine по адресу <https://archive.ics.uci.edu/ml/machinelearning-databases/wine/wine.data> Извлеките из данных признаки и классы. Класс записан в первом столбце (три варианта), признаки — в столбцах со второго по последний. Более подробно о сути признаков можно прочитать по адресу <https://archive.ics.uci.edu/ml/datasets/Wine> Оценку качества необходимо провести методом кроссвалидации по 5 блокам (5-fold).

Создайте генератор разбиений, который перемешивает выборку перед формированием блоков (`shuffle=True`). Для воспроизводимости результата, создавайте генератор `KFold` с фиксированным параметром `random_state=42`. В качестве меры качества используйте долю верных ответов (`accuracy`). Найдите точность классификации на кросс-валидации для метода `k` ближайших соседей (`sklearn.neighbors.KNeighborsClassifier`), при `k` от 1 до 50. При каком `k` получилось оптимальное качество? Чему оно равно (число в интервале от 0 до 1)?

Произведите масштабирование признаков с помощью функции `sklearn.preprocessing.scale`. Снова найдите оптимальное `k` на кросс-валидации.

Какое значение `k` получилось оптимальным после приведения признаков к одному масштабу? Как изменилось значение качества? Приведите ответы на вопросы

### Задание 3.

1. Загрузите выборку из файла `titanic.csv` с помощью пакета `Pandas`.
2. Оставьте в выборке четыре признака: класс пассажира (`Pclass`), цену билета (`Fare`), возраст пассажира (`Age`) и его пол (`Sex`).
3. Обратите внимание, что признак `Sex` имеет строковые значения.
4. Выделите целевую переменную — она записана в столбце `Survived`.
5. В данных есть пропущенные значения — например, для некоторых пассажиров неизвестен их возраст. Такие записи при чтении их в `pandas` принимают значение `nan`. Найдите все объекты, у которых есть пропущенные признаки, и удалите их из выборки.
6. Обучите решающее дерево с параметром `random_state=241` и остальными параметрами по умолчанию.

### Задание 4.

1. Загрузите обучающую и тестовую выборки из файлов `perceptrontrain.csv` и `perceptron-test.csv`. Целевая переменная записана в первом столбце, признаки — во втором и третьем.
2. Обучите перцептрон со стандартными параметрами и `random_state=241`.
3. Подсчитайте качество (долю правильно классифицированных объектов, `accuracy`) полученного классификатора на тестовой выборке.
4. Нормализуйте обучающую и тестовую выборку с помощью класса `StandardScaler`.
5. Обучите перцептрон на новых выборках. Найдите долю правильных ответов на тестовой выборке.
6. Найдите разность между качеством на тестовой выборке после нормализации и качеством до нее.

### Вариант 2

Задание 1. Анализ данных по пассажирам Титаника.

С помощью `Pandas` ответить на несколько вопросов по данным репозитория `UCI Titanic`;

1. Какое количество мужчин и женщин ехало на корабле?
2. Какой части пассажиров удалось выжить? Посчитайте долю выживших пассажиров.
3. Какую долю пассажиры первого класса составляли среди всех пассажиров?
4. Какого возраста были пассажиры? Посчитайте среднее и медиану возраста пассажиров.
5. Коррелируют ли число братьев/сестер с числом родителей/детей?

6. Посчитайте корреляцию Пирсона между признаками SibSp и Parch. Какое самое популярное женское имя на корабле?
7. Извлеките из полного имени пассажира (колонок Name) его личное имя (FirstName).

Задание 2. Исходные данные: необходимо решить задачу регрессии с помощью метода k ближайших соседей. Для это используются класс `sklearn.neighbors.KNeighborsRegressor`, метрика задается с помощью параметра `metric` (значение `'minkowski'`), параметр метрики Минковского задается с помощью параметра `p` данного класса. В данном задании набор данных Boston, где нужно предсказать стоимость жилья на основе различных характеристик расположения (загрязненность воздуха, близость к дорогам и т.д.). Результатом вызова данной функции является объект, у которого признаки записаны в поле `data`, а целевой вектор — в поле `target`.

Необходимо:

1. Приведите признаки в выборке к одному масштабу при помощи функции `sklearn.preprocessing.scale`.
2. Переберите разные варианты параметра метрики `p` по сетке от 1 до 10 с таким шагом, чтобы всего было протестировано 200 вариантов (используйте функцию `numpy.linspace`). Используйте `KNeighborsRegressor` с `n_neighbors=5` и `weights='distance'` - данный параметр добавляет в алгоритм веса, зависящие от расстояния до ближайших соседей. В качестве метрики качества используйте среднеквадратичную ошибку (параметр `scoring='mean_squared_error'` у `cross_val_score`; при использовании библиотеки `scikit-learn` версии 18.0.1 и выше необходимо указывать `scoring='neg_mean_squared_error'`).
3. Качество оценивайте с помощью кросс-валидации по 5 блокам с `random_state = 42`, не забудьте включить перемешивание выборки (`shuffle=True`).
4. Определите, при каком `p` качество на кросс-валидации оказалось оптимальным. Обратите внимание, что `cross_val_score` возвращает массив показателей качества по блокам; необходимо сделать массив показателей качества по блокам; необходимо максимизировать среднее этих показателей.

Задание 3.

1. Загрузите выборку из файла `svm-data.csv`. В нем записана двумерная выборка (целевая переменная указана в первом столбце, признаки — во втором и третьем).

2. Обучите классификатор линейным ядром, параметром `C=100000` и `random_state=241`. Такое значение параметра нужно использовать, чтобы убедиться, что SVM работает выборкой как линейно разделимой. При более низких значениях параметра алгоритм будет настраиваться с учетом слагаемого в функционале, штрафующего за маленькие отступы, из-за чего результат может не совпасть с решением классической задачи SVM для линейно разделимой выборки.

3. Найдите номера объектов, которые являются опорными (нумерация с единицы).

Задание 4.

1. Загрузите выборку из файла `titanic.csv` с помощью пакета `Pandas`.
2. Оставьте в выборке три признака: класс пассажира (`Pclass`), цену билета (`Fare`), возраст пассажира (`Age`) и его пол (`Sex`).
3. Обратите внимание, что признак `Sex` имеет строковые значения.
4. Выделите целевую переменную — она записана в столбце `Survived`.
5. В данных есть пропущенные значения — например, для некоторых пассажиров неизвестен их возраст. Такие записи при чтении их в `pandas` принимают значение `nan`. Найдите все объекты, у которых есть пропущенные признаки, и удалите их из выборки.
6. Вычислите важности признаков и найдите два признака с наибольшей важностью. Их названия будут ответами для данной задачи (в качестве ответа укажите

названия признаков через запятую без пробелов)

### ЛИСТ СОГЛАСОВАНИЯ

№п/п	Подразделение	Фамилия	Подпись	Дата
1	Кафедра ГЕиМД	И.О. Тимофеева		10.06.2020
2	Учеб.-метод. отдел	М.О. Дерябичева		10.06.2020
3	Библиотека	Г.В. Шпакова		10.06.2020