

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Андрей Драгомирович Хлутков
Должность: директор
Дата подписания: 12.08.2026 12:44:46
Уникальный программный ключ:
880f7c07c583b07b775f6604a630281b13ca9fd2

Приложение 4
к образовательной программе

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.ДЭ.08.02 Аналитика больших данных

(индекс, наименование дисциплины в соответствии с учебным планом)

38.03.05 Бизнес-информатика

(код, наименование направления подготовки/специальности)

Бизнес-аналитика

(наименование образовательной программы)

Очная форма обучения

(форма обучения)

Год набора - 2026

Санкт-Петербург

Автор(ы)-составитель(и) РПД:

Зеленина Лариса Ивановна, кандидат технических наук, доцент, доцент кафедры бизнес-информатики.

Заведующий кафедрой:

Наумов Владимир Николаевич, доктор военных наук, профессор, заведующий кафедрой бизнес-информатики

Рабочая программа дисциплины Б1.В.ДЭ.08.02 Аналитика больших данных одобрена на заседании кафедры бизнес-информатики СЗИУ РАНХиГС.

протокол № 6 от «26» марта 2026 г.

СОДЕРЖАНИЕ

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы
2. Объем и место дисциплины в структуре образовательной программы
3. Содержание и структура дисциплины
4. Типы оценочных материалов, показатели и критерии их оценивания
5. Формы аттестации, типовые оценочные материалы для текущего контроля успеваемости обучающихся, критерии и шкалы оценивания по контрольным точкам
6. Формы промежуточной аттестации, критерии и шкала оценивания, типовые оценочные материалы по дисциплине
7. Методические материалы по освоению дисциплины
8. Учебная литература и ресурсы информационно-телекоммуникационной сети «Интернет»
9. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Дисциплина Б1.В.ДЭ.08.02 Аналитика больших данных обеспечивает формирование у обучающихся следующих универсальных, общепрофессиональных и профессиональных компетенций*:

ОТФ/ТФ и реквизиты ПС (при наличии)**	Код компетенции **	Наименование Компетенции **	Код индикатора достижения компетенции **	Наименование индикатора достижения компетенций **	Образовательный результат **
06.014 Менеджер по информационным технологиям, утв. приказом Министерства труда и социальной защиты РФ от 30.08.2021 г. № 588н А- Управление операционной деятельностью организации в области ИТ; А/01.6 Управление изменениями ИТ	ПКС-1.	Способен управлять ресурсами ИТ, инфраструктурой, информационной безопасностью, качеством ИТ	ПКС-1.1	Демонстрирует умение управлять ИТ-инфраструктурой, информационными процессами и системами	ПКС -1.1. 3-1. Знает международные и отечественные стандарты, лучшие практики и фреймворки по управлению изменениями в ИТ; ПКС-1.1. У-1. Умеет моделировать, анализировать и декомпозировать цели управления изменениями ИТ ПКС-1.1. У-3. Умеет осуществлять мониторинг и контроль управления изменениями ИТ

* Дисциплина может формировать компетенцию полностью или частично.

** Должно соответствовать Приложению 1 к образовательной программе

2. Объем и место дисциплины в структуре образовательной программы

Общий объем дисциплины:

Общий объем дисциплины: 3 з.е., 108 ак.час

Контактная работа обучающихся с преподавателем по видам учебных занятий: 49 ак. час на контактную работу с преподавателем, из них 16 ак.час на лекции и 24 ак.час на практические занятия, 59 ак. час на самостоятельную работу обучающихся (9 ак.часа на каттэк).

Б1.В.ДЭ.08.02 Аналитика больших данных реализуется в 8-м семестре 4-го курса. Преподавание дисциплины Б1.В.ДЭ.08.02 Аналитика больших данных опирается на дисциплины Теория вероятностей и математическая статистика, Эконометрическое моделирование, Нечеткая логика и нейронные сети.

В свою очередь она создаёт необходимые предпосылки для при выполнении научно-исследовательской и выпускной квалификационной работы.

Объем дисциплины, реализуемый с применением СДО: количество академических часов, выделенных на самостоятельную работу обучающихся: всего с применением СДО – 59 ак.ч.

Знания, умения и навыки, полученные при изучении дисциплины, используются студентами при подготовке и сдаче государственного экзамена.

Формой промежуточной аттестации в соответствии с учебным планом является зачет с оценкой.

3. Содержание и структура дисциплины

3.1. Структура дисциплины

Очная форма обучения

№ п/п	Наименование тем и (или) разделов	ВСЕГО	Объем дисциплины, ак.час											Форма текущего контроля успеваемости, промежуточной аттестации	
			Контактная работа обучающихся с преподавателем по видам учебных занятий								Самостоятельная работа				
			Период теоретического обучения				Период промежуточной аттестации (сессия)								
			Занятия лекционного типа		Занятия семинарского типа		ИК	КСР	КЭ	Катт эк	Контроль	СРкр	СРэк		СР
			Л	ВЛ	ЛР	ПЗ									
Тема 1	Жизненный цикл аналитики данных. Высокопроизводительные вычисления	45	6			10							29	Тестирование, Практическое контрольное задание	
Тема 2	Масштабирование и многоуровневое хранение данных. Визуализация данных и результатов анализа	54	10			14							30	Тестирование, Практическое контрольное задание	
Промежуточная аттестация		9								9				Зачет с оценкой	
Итого		108	16			24				9			59		

Используемые сокращения:

Л – лекции - занятия, предусматривающие преимущественную передачу учебной информации обучающимся педагогическими работниками организации и (или) лицами,

привлекаемыми организацией к реализации образовательных программ на иных условиях,).

ВЛ – видео лекции.

ЛР – лабораторные работы.

ПЗ – практические занятия (за исключением лабораторных работ).

ИК – индивидуальные консультации.

КСР – контроль самостоятельной работы

КЭ – консультации перед экзаменом

Каттэк – контактная работа на аттестацию в период экзаменационных сессий

Контроль - контактная работа на аттестацию в период экзаменационных сессий для заочной формы обучения

СРкр – самостоятельная работа на подготовку курсовой работы/ курсового проекта.

СРэк – самостоятельная работа на подготовку к экзамену.

СР – самостоятельная работа в семестре на подготовку к учебным занятиям.

Т – тестирование.

ПКЗ – практические контрольные задания.

ПИЗ – профессионально-исследовательские задания.

В процессе обучения применяются следующие интерактивные формы: лекция-диалог, работа в малых группах, спарринг-партнерство.

Темы 1-2 могут быть освоены с применением ЭО и ДОТ с контролем в системе электронного обучения Академии.

3.2. Содержание дисциплины

Тема 1. Жизненный цикл аналитики данных. Высокопроизводительные вычисления. ПКС-1.1.

Понятие «Big Data». Основная цель обработки Big Data. Характеристики Big Data. Принцип трех "V". Квази-структурированные данные.

Business Intelligence. OLAP-технологии. Средства предоставления информации и средства интеграции в Business Intelligence. Цели Data Science. Жизненный цикл аналитики данных «предиктивное моделирование. ETL. BI-аналитика.

Apache Hadoop. Решения на базе Hadoop. MapReduce: достоинства и недостатки. Принцип HDFS.

Тема 2. Масштабирование и многоуровневое хранение данных. Визуализация данных и результатов анализа. ПКС-1.1.

Особенность NoSQL. Применение NoSQL хранилища. Способ хранения данных в MongoDB. Нереляционные БД. Группы пользователей в MongoDB. Шардинг. Теорема CAP.

Языки сценариев. Традиционные виды визуализации. Основные возможности языка R. Дедупликация данных. Основные задачи визуализации. Требования к визуализации. Типы визуализации.

4. Типы оценочных материалов, показатели и критерии оценивания

4.1. Оценочные материалы по дисциплине Б1.В.ДЭ.08.02 Аналитика больших данных входят в состав оценочных материалов по образовательной программе. Совокупность оценочных материалов по всем дисциплинам образовательной программы составляет фонд оценочных средств (далее – ФОС). ФОС используется при проведении текущего контроля успеваемости и промежуточной аттестации обучающихся с целью оценивания достижения обучающимися планируемых результатов обучения.

4.2. ФОС разработан как комплекс проверочных заданий различного типа и уровня сложности, включает критерии и шкалы оценивания, а также «ключи» правильных ответов. ФОС формируется как отдельный документ и хранится в электронном виде, доступ к ФОС предоставлен ограниченному кругу лиц.

4.3. Для самостоятельной работы обучающихся при подготовке к текущему контролю успеваемости и промежуточной аттестации в рабочих программах дисциплин размещены типовые проверочные задания, которые можно условно разделить на задания закрытого, комбинированного и открытого типов.

Задания закрытого типа — это тестовые задания, в которых каждый вопрос сопровождается готовыми вариантами ответов, из которых необходимо выбрать один или несколько правильных.

Задания комбинированного типа – это тестовые задания, в которых каждый вопрос сопровождается готовыми вариантами ответов, из которых необходимо выбрать один или несколько правильных и обосновать свой выбор.

Задания открытого типа — это задания, в которых на каждый вопрос должен быть предложен развернутый обоснованный ответ.

В зависимости от типа задания рекомендованы определенная последовательность выполнения и система оценивания выполнения заданий.

4.4. Типы заданий, сценарии выполнения, критерии оценивания

ТИП ЗАДАНИЯ	ИНСТРУКЦИЯ	СЦЕНАРИИ ВЫПОЛНЕНИЯ	КРИТЕРИИ ОЦЕНИВАНИЯ
Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных	Прочитайте текст, выберите правильный ответ	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные вариант-ты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).	Ответ считается верным, если правильно указана цифра или буква
Задание закрытого типа на установление соответствия	Прочитайте текст и установите соответствие	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидаются пары элементов. 2. Внимательно прочитать оба списка: список 1 – вопросы, утверждения, факты, понятия и т.д.; список 2 – утверждения, свойства объектов и т.д. 3. Сопоставить элементы списка 1 с элементами списка 2, сформировать пары элементов. 4. Записать попарно буквы и цифры (в зависимости от задания) вариантов ответа (например, А1 или Б4).	Ответ считается верным, если правильно указаны цифры или буквы
Задание закрытого типа с выбором нескольких правильных ответов из нескольких вариантов предложенных	Прочитайте текст, выберите правильные ответы	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается несколько правильных ответов из предложенных вариантов. 2. Внимательно прочитать предложенные вариант-ты ответа. 3. Выбрать несколько правильных ответов. 4. Записать только номера (или буквы)	Ответ считается верным, если правильно установлены все соответствия (позиции из одного столбца верно сопоставлены с позициями другого)

		выбранного варианта ответа (например, 1 4 или А Г).	
Задание закрытого типа на установление последовательности	Прочитайте текст и установите последовательность	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается последовательность элементов. 2. Внимательно прочитать предложенные варианты ответа. 3. Построить верную последовательность из предложенных элементов. 4. Записать буквы/цифры (в зависимости от задания) вариантов ответа в нужной последовательности (например, БВА или 135).	Ответ считается верным, если правильно указана вся последовательность цифр
Задание комбинированного типа с выбором одного правильного ответа из предложенных и обоснованием выбора	Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа. 5. Записать аргументы, обосновывающие выбор ответа (например, 4 текст обоснования).	Ответ считается верным, если правильно указана цифра или буква и приведены корректные аргументы, используемые при выборе ответа
Задание открытого типа с развернутым ответом	Прочитайте текст и запишите развернутый обоснованный ответ	1. Внимательно прочитать текст задания и понять суть вопроса. 2. Продумать логику и полноту ответа. 3. Записать ответ, используя четкие компактные формулировки. 4. В случае расчетной задачи, записать решение и ответ	Ответ считается верным: 1. Отсутствие фактических ошибок. 2. Раскрытие объема используемых понятий (полнота ответа). 3. Обоснованность ответа (наличие аргументов). 4. Логическая последовательность излагаемого материала.

4.5. Общая шкала оценивания результатов текущего контроля успеваемости и промежуточной аттестации обучающихся с применением БРС

Итоговая балльная оценка	Традиционная система	Бинарная система	ECTS	
			Для традиционной системы	Для бинарной системы
95-100	Отлично	Зачтено	A	P/ Passed
85-94			B	P/ Passed
75-84	Хорошо		C	P/ Passed
65-74			D	P/ Passed
55-64	Удовлетворительно		E	P/ Passed
0-54	Неудовлетворительно	Не зачтено	F	F/Failed

Соотношение баллов за текущий контроль успеваемости и промежуточную аттестацию, а также повторную промежуточную аттестацию:

Максимальная сумма баллов за текущий контроль успеваемости	Максимальная сумма баллов за промежуточную аттестацию	Максимальная итоговая балльная оценка	Максимальная сумма баллов за повторную промежуточную аттестацию
60 баллов	40 баллов	100 баллов	100 баллов

5. Формы аттестации, типовые оценочные материалы для текущего контроля успеваемости обучающихся, критерии и шкалы оценивания по контрольным точкам

5.1. В ходе реализации дисциплины используются следующие формы текущего контроля успеваемости обучающихся (в том числе, задания к контрольным точкам):

Т – тестирование, ПКЗ – практические контрольные задания.

Тема 1. Жизненный цикл аналитики данных. Высокопроизводительные вычисления.

Тестовые задания:

Задание открытого типа с развернутым ответом

1. Внимательно прочитать текст задания и понять суть вопроса.
2. Продумать логику и полноту ответа.
3. Записать ответ, используя четкие компактные формулировки.

1. Что означает термин «Big Data» в информационных технологиях?
2. Что является основной целью обработки Big Data?
3. Кто и в каком году впервые ввел термин «Big Data»?
4. Какие главные характеристики Big Data?
5. Что является главным результатом процесса Business Intelligence?
3. Что означает термин «Business Intelligence» в информационных технологиях?
4. Расшифруйте аббревиатуру OLAP.
5. Что относится к средствам предоставления информации в Business Intelligence?
6. Что относится к средствам интеграции в «Business Intelligence»?
7. Что такое Apache Hadoop? В чем преимущества решений на базе Hadoop?
8. Что такое MapReduce?
9. Какими достоинствами и недостатками обладает MapReduce?
10. Какому основному принципу следует HDFS?
11. Какой размер блока по умолчанию в HDFS?
12. Какие функции выполняет NameNode в HDFS?

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.
2. Внимательно прочитать предложенные вариант-ты ответа.
3. Выбрать один верный ответ.
4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В)..

1. Что является средством анализа в BI?

- А) **Карты показателей;**
- Б) Совместная работа и управление рабочими процессами;
- В) Информационные панели;
- Г) BI инфраструктура.

2. Основное умение исследователя данных?

- А) Умение находить наиболее важные элементы в хранимой информации
- Б) Уметь прогнозировать исход работы системы
- В) **Находить скрытые логические связи в системе собранной информации**
- Г) Отличать неструктурированные данные от структурированных

3. Какой язык программирования из перечисленных является наиболее важным для аналитика?

A) C++ Б) PHP В) F# Г) **R**

4. Что означает термин «Business Intelligence» в информационных технологиях?

A) Комплексный набор методов для создания бизнес-планов.

Б) **Методы и инструменты для перевода необработанной информации в осмысленную, удобную для восприятия форму.**

В) Файлы, содержащие информацию о бизнес-плане.

Г) Технологии, направленные на развитие бизнеса.

5. Языком, на котором был разработан RabbitMQ, является:

A) Java Б) Python В) C++ Г) **Erlang**

6. Что является главным результатом процесса Business Intelligence?

A) **Возможность принятия решений для бизнеса**

Б) Результаты интеллектуального анализа данных

В) Возможность использования искусственного интеллекта

Г) Получение структуризации данных после выполнения всех шагов процесса

7. Что из перечисленного не является средством анализа?

A) Продвинутое визуализация

Б) **Reporting**

В) Predictive Modelling

Г) Data Mining

8. Что относится к средствам предоставления информации в «Business Intelligence»?

A) **Генератор нерегламентированных запросов**

Б) Совместная работа и управление рабочими процессами

В) Предиктивное моделирование и Data Mining

Г) Карты показателей

9. Процессом создания и выбора модели для предсказания вероятности наступления некоторого события является:

A) OLAP Б) Data Mining В) **Predictive Modelling** Г) Data Science

ПКЗ по теме 1:

Методы оптимизации в глубоком обучении

В общем случае задачу оптимизации можно записать в виде:

$$\text{minimize}_x f(x) \text{ при } \begin{cases} g_i(x) \leq 0, & i = 1, \dots, m \\ h_j(x) = 0, & j = 1, \dots, p \end{cases},$$

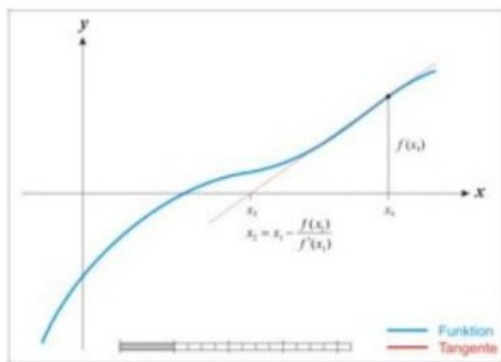
где $f: \mathbb{R}^n \rightarrow \mathbb{R}$ – целевая функция, минимизация которой осуществляется путем поиска значений n -мерного вектора x ;

$g_i(x) \leq 0$ – ограничения в виде неравенств;

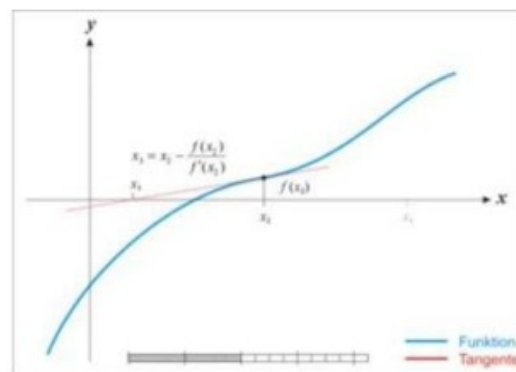
$h_j(x) = 0$ – ограничения в виде равенств;

$m \geq 0, p \geq 0$.

В рамках практической работы сосредоточимся на рассмотрении дифференциальных методов оптимизации заданной функции нескольких переменных и вариациях метода градиентного спуска. Метод Ньютона Метод Ньютона – метод поиска корней уравнения, использующий линейную аппроксимацию. Предполагается, что точка x_n – некоторое приближенное решение уравнения $f(x) = 0$. В точке производится расчет линейного приближения функции $f(x)$, после чего в качестве нового приближенного решения x_{n+1} берется пересечение графика аппроксимации с осью абсцисс (рисунок 1.1).



а



б

... Задание 1. Скачать исходные данные

Задание 2. Реализовать функцию для алгоритма обратного распространения ошибки в виде градиентного спуска для трёхслойной полносвязной сети с 32-16-10 нейронами в слоях соответственно и среднеквадратичной функции ошибок для задачи бинарной классификации.

Задание 3. Визуализировать процесс обучения сети, построив график изменения лосс-функции.

Задание 4. Реализовать алгоритм стохастического градиентного спуска и его модификации: а) метод моментов; б) RMSprop; в) Адам. Задание

5. Сделать сравнительный анализ точности решения оптимизационной задачи с помощью различных методов.

Тема 2. Масштабирование и многоуровневое хранение данных. Визуализация данных и результатов анализа.

Тестовые задания:

Задание открытого типа с развернутым ответом

1. Внимательно прочитайте текст задания и понять суть вопроса.
 2. Продумать логику и полноту ответа.
 3. Записать ответ, используя четкие компактные формулировки.
 4. В случае расчетной задачи, записать решение и ответ
-
1. Чем анализ больших данных отличается от традиционного анализа?
 2. Какие основные типы Data Mining?
 3. Какие категории Web Mining можно выделить?
 4. В чем основная задача Web Content Mining?
 5. В чем основные задачи интеллектуального анализа текстов?
 6. Что изучает статистика?
 7. К каким алгоритмам классификации относится метод ближайших соседей?
 8. Что является целью кластеризации?
 9. Чем анализ больших данных отличается от традиционного анализа?
 10. Какие основные типы Data Mining?
 11. Какие категории Web Mining можно выделить?
 12. В чем основная задача Web Content Mining?
 13. В чем основные задачи интеллектуального анализа текстов?
 14. К каким алгоритмам классификации относится метод ближайших соседей?

Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных.

1. Внимательно прочитайте текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов.
 2. Внимательно прочитайте предложенные варианты-ты ответа.
 3. Выбрать один верный ответ.
 4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В)..
-
1. Вертикальное масштабирование...
 - А) Требуется изменений в прикладных программах, работающих на таких системах
 - Б) Не требует никаких изменений в прикладных программах,

работающих на таких системах

- В) Уменьшает производительность каждого компонента БД
 - Г) Увеличивает скорость загрузки данных
2. Для достижения какого свойства в БД типа NoSQL нет JOIN операций?
- А) Intercepting 20 Б) Concurrency В) **Consistency** Г) Capacity
3. Что, согласно теореме CAP (теореме Брюера), возможно обеспечить в любой реализации распределённых вычислений?
- А) Только согласованность данных
 - Б) Только доступность данных
 - В) Согласованность данных, доступность данных, устойчивость к разделению
 - Г) **Не более двух свойств из трёх вышеприведённых**
4. Выберите верное определение понятия AP-система:
- А) Система, во всех узлах которой данные согласованы и обеспечена доступность, жертвует устойчивостью к распаду на секции
 - Б) Распределённая система, в каждый момент обеспечивающая целостный результат и способная функционировать в условиях распада
 - В) **Распределённая система, отказывающаяся от целостности результата**
 - Г) Система, автоматически изменяющая данные алгоритма своего с целью сохранения оптимального состояния
5. Что означает термин NoSQL?
- А) Не SQL Б) **Не только SQL** В) Без SQL Г) SQL – плохо
6. Разбиение системы на более мелкие структурные компоненты и разнесение их по отдельным физическим машинам (или их группам), и (или) увеличение количества серверов, параллельно выполняющих одну и ту же функцию, это:
- А) **Горизонтальное масштабирование**
 - Б) Вертикальное масштабирование
 - В) Master- slave репликация
 - Г) Peer-to-peer репликация
7. Что из перечисленного относится к графо-ориентированным хранилищам (Graph Store)?
- А) **Neo4j** Б) BaseX В) Elasticsearch Г) Ничего
8. Что поддерживает NoSQL?
- А) **Операцию Insert**
 - Б) Полностью стандарт SQL
 - В) Операцию Join

Г) Операцию Group by

9. Какие три свойства фигурируют в определении теоремы CAP?

А) **Согласованность данных** Б) Сложность В) Доступность Г) Устойчивость к разделению

10. Выделение таблицы или группы таблиц на отдельный сервер это...

А) Горизонтальное масштабирование

Б) Вертикальное масштабирование

В) Горизонтальный шардинг

Г) **Вертикальный шардинг**

ПКЗ по теме 2:

Задачи обработки естественного языка.

Задачи обработки естественного языка – одни из самых востребованных в современном машинном обучении. Они включают в себя такие области, как машинный перевод, аннотирование текстов, классификация документов, генерация текста, text-to-image и text-to-video задачи, построение чат-ботов и диалоговых систем и многие другие. Решение любой задачи в сфере NLP (natural language processing) начинается с предобработки текста, которая обычно включает в себя: а) удаление всех не относящиеся к естественному языку символов из текста: @, #, & и т.д. (если это не противоречит цели исследования); б) токенизация текста, т.е., разделение его на отдельные слова: слово = «токен»; в) удаление стоп-слов, встречающихся во всех без исключения текстах и не несущих смысловой нагрузки для решения текущей задачи; г) приведение текста к нижнему регистру, чтобы модель распознавала такие слова, как, например, «привет» и «Привет», одинаково; д) стэмминг (обрезка слова до его основания) и лемматизация текста (приведение слов к единой форме, например, («красивая», «красивый», «красивое») → «красивый»). Чтобы использовать методы машинного обучения на текстовых документах, нужно перевести текстовое содержимое (слова на естественном языке) в числовой вектор признаков. Наиболее интуитивно понятный способ сделать описанное выше преобразование — это представить текст в виде набора слов, после чего приписать каждому слову в тексте уникальный целочисленный индекс, соответствующий частоте его появления в документах обучающей выборки. Для этого в каждом документе i следует посчитать количество употреблений каждого слова w и сохранить это число в ячейке $X[i, j]$. Это будет соответствовать значению признака j , где j — это индекс слова w в словаре. Такое преобразование текста в матрицу частот употреблений слов носит название «мешка слов» (или «bag of words»). Оно подразумевает, что

вероятность появления слова в разных документах одинакова, и строит матрицу объекты-признаки исходя из предположения, что количество признаков соответствует количеству уникальных слов в корпусе документов. «Мешок слов» чаще всего является высокоразмерным разреженным набором данных (с большим количеством нулей).

D1 - "I am feeling very happy today"

D2 - "I am not well today"

D3 - "I wish I could go to play"

	I	am	feeling	very	happy	today	not	well	wish	could	go	to	play
D1	1	1	1	1	1	1	0	0	0	0	0	0	0
D2	1	1	0	0	0	1	1	1	0	0	0	0	0
D3	2	0	0	0	0	0	0	0	1	1	1	1	1

слов для корпуса документов из трёх предложений (источник изображения:

deeplearning.ai by Andrew Ng)

Рисунок 1 – Пример построения матрицы частот упоминаний уникальных

Однако, даже если документы посвящены одной теме, в относительно длинных документах среднее количество словоупотреблений будет выше, чем в коротких, что может приводить к некорректной настройке моделей машинного обучения. Чтобы избежать потенциальных несоответствий, применяется подход, называемый TF-IDF – «term frequency – inverse document frequency», который позволяет оценить «важность» слова для данного документа. Этот подход позволяет снизить влияние низкоинформативных слов и, наоборот, повысить «вес» важных с точки зрения оценки принадлежности документа к определённой тематике. Для каждого слова вычисляется следующая величина:

$$tf - idf(t, d, D) = tf(t, d) \times idf(t, D),$$

$$tf(t, d) = \frac{n_t}{\sum_k n_k},$$

$$idf(t, D) = \log \frac{|D|}{|\{d_i \in D | t \in d_i\}|},$$

где n_t - частота слова (токена) t в документе d , $|D|$ - количество документов в коллекции; $|\{d_i \in D | t \in d_i\}|$ - количество документов в коллекции, в которых встречается слово t .

...Сформированные таким образом вектора позволяют вычислять «семантическое расстояние» между словами: например, слова «король» и «королева» будут находиться в этом представлении близко друг к другу, а слова «король» и «яблоко» - далеко.

Задание 1. Скачать английскую книгу "Alice's Adventures in Wonderland"

<http://www.gutenberg.org/files/11/11-0.txt>

Задание 2. Реализовать пайплайн обработки выбранного текста на английском языке, включая всю необходимую предварительную обработку текста, включая приведение слов к нижнему регистру, удаление стоп-слов, цифр/неалфавитных символов, знаков пунктуации. Задание

3. Разделить текст на главы и в каждой главе отобразить Топ-20 слов с помощью алгоритма TF-IDF. Задание

4. Реализовать LDA алгоритм и сравнить результаты с полученными ранее с помощью TF-IDF. Сделать выводы о применимости реализованных подходов.

Дополнительные вопросы и задания

1. Какое слово имеет максимальное значение метрики TF-IDF для двенадцатой главы под названием «Alice's Evidence»?

2. Как бы вы назвали главы книги «Alice in Wonderland», основываясь на найденных для каждой главы важных словах?

**5.2. Типовые оценочные материалы для текущего контроля успеваемости обучающихся (вне контрольных точек):
приведены в п.6.2.**

5.3. Один или несколько тематических блоков дисциплины завершаются контрольной точкой (далее – КТ). Текущий контроль успеваемости по дисциплине предусматривает не менее 2 (двух) и не более 10 (десяти) КТ в течение периода освоения дисциплины.

Максимальное количество баллов за любой тип работ в рамках КТ составляет 100 (сто) баллов.

Распределение весовых коэффициентов по КТ в рамках текущего контроля успеваемости по дисциплине и формулы расчета:

Наименование контрольной точки	Максимальное количество баллов за работу в рамках КТ, которое может набрать студент	Коэффициент веса контрольной точки	Результат контрольной точки, участвующий в формировании итоговой балльной оценки по дисциплине (отражается в журнале БРС в СДО)
КТ - 1	100	0, 23	23
КТ - 2	100	0,23	23
КТ- 3	100	0,14	14
Итого:	x	0,6	60

Формула расчета результата контрольной точки:

Результат контрольной точки = Количество баллов за работу в рамках КТ

X Коэффициент веса контрольной точки.

5.4. Формы текущего контроля успеваемости обучающихся в рамках КТ и типовые оценочные материалы:

КТ-1

Тема 1.

Практическое контрольное задание (ПКЗ)

Тестирование.

КТ-2

Тема 2.

Тестирование.

КТ-3

Тема 2.

Практическое контрольное задание (ПКЗ)

Для каждой формы текущего контроля успеваемости обучающихся в рамках КТ определены критерии оценивания результатов выполнения задания.

1. Критерии оценивания тестирования:

Критерии оценки	Диапазон баллов	Описание критерия
<i>Количество правильных ответов</i>	<i>0</i>	<i>Количество правильных ответов менее 55%</i>
	<i>25</i>	<i>Количество правильных ответов от 55% до 64%</i>
	<i>50</i>	<i>Количество правильных ответов от 65% до 74%</i>
	<i>75</i>	<i>Количество правильных ответов от 75% до 84%</i>
	<i>100</i>	<i>Количество правильных ответов от 85% до 100%</i>
Итого максимально:	100	

2. Критерии оценивания ПКЗ:

Критерии оценки	Диапазон баллов	Описание критерия
<i>Содержание и раскрытие выбранных понятий</i>	<i>41-70</i>	<i>Детальное, последовательное описание всех понятий на примере выбранной системы</i>
	<i>21-40</i>	<i>Поверхностное описание без привязки к выбранной системе</i>
	<i>0-20</i>	<i>Понятия раскрыты минимально или не раскрыты вовсе</i>

<i>Количество выполненных заданий</i>	<i>30</i>	<i>Количество выполненных заданий от 85% до 100%</i>
	<i>15</i>	<i>Количество выполненных заданий от 55% до 84%</i>
	<i>0</i>	<i>Количество выполненных заданий менее 55%</i>
Итого максимально:	100	

5.5. Описание дополнительных материалов и оборудования, необходимых для выполнения проверочных заданий (*при необходимости*).

Для решения задач открытого типа (ПКЗ, КР), тестовых заданий студенту разрешается использование калькулятора; программ для работы с электронными таблицами для обработки, анализа и визуализации данных. Для построения интеллект-карты и моделей в различных нотациях студенту можно использовать любой соответствующий онлайн-инструмент.

6. Формы промежуточной аттестации, критерии и шкала оценивания, типовые оценочные материалы по дисциплине

6.1. Промежуточная аттестация проводится в форме зачета с оценкой.

Зачет проводится в устной форме. Обучающийся получает экзаменационный билет с вариантами 2-х заданий различного типа. На выполнение заданий дается 20-30 минут. По завершении подготовки необходимо представить ответы, подробно изложив ход выполнения задания, сделать выводы (*при необходимости*).

При реализации промежуточной аттестации в ЭО/ДОТ могут быть использованы следующие формы: устно в ДОТ - в форме обоснованных ответов на задания различного типа; письменно в СДО - в форме письменного решения заданий различного типа; тестирование в СДО.

6.2. Типовые оценочные материалы промежуточной аттестации.

Вопросы для подготовки к зачету с оценкой.

1. Термин Big Data, цель и главные характеристики больших данных.
2. Принцип трех "V".
3. Структурированные, неструктурированные и квази-структурированные данные.
4. Business Intelligence.
5. Цели и задачи Data Science.
6. Что такое жизненный цикл аналитики данных.

7. Технология Apache Hadoop.
8. Технология MapReduce, достоинства и недостатки.
9. Файловая система HDFS.
10. Хранилища NoSQL главные отличия от обычных БД.
11. Теорема CAP и ее следствия.
12. Amazon S3.
13. Язык R.
14. Основные задачи, требования и типы визуализации.
15. Data Mining и основные типы.
16. Классификация
17. Кластеризация
18. Задачи регрессии.
19. Оценка качества обучения модели.

Типовые задания для зачета

1. Почему стохастический градиентный спуск работает быстрее, чем классический градиентный спуск?
2. Какие существуют методы обхода локальных минимумов?
3. Чем, по вашему мнению, должен определяться размер mini-batch для реализации градиентного спуска в конкретном случае?
4. Обучите свёрточную нейронную сеть разными алгоритмами оптимизации для решения задачи из текста лабораторной работы. Какой алгоритм оптимизации для настройки весов свёрточной сети сработает лучше в этом случае и почему?

Типовые проверочные задания для самоподготовки обучающегося к промежуточной аттестации:

ТИП ЗАДАНИЯ	СЦЕНАРИИ ВЫПОЛНЕНИЯ	ТИПОВЫЕ ЗАДАНИЯ	
Задание закрытого типа с выбором одного правильного ответа из нескольких вариантов предложенных	1.Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа (например, 3 или В).	Что не является целью процесса Business Intelligence? А) Интерпретация большого количества данных; Б) Моделирование исходов различных вариантов действий; В) Модификация существующего программного обеспечения ; Г) Отслеживание результатов решений.	
Задание закрытого типа на установление	1.Внимательно прочитать текст задания и понять, что в качестве ответа	Установить взаимно однозначное соответствие (a2, в1,c3)	
		Формула	действие

соответствия	ожидаются пары элементов. 2. Внимательно прочитать оба списка: список 1 – вопросы, утверждения, факты, понятия и т.д.; список 2 – утверждения, свойства объектов и т.д. 3. Сопоставить элементы списка 1 с элементами списка 2, сформировать пары элементов. 4. Записать попарно буквы и цифры (в зависимости от задания) вариантов ответа (например, А1 или Б4).	А) Модели классификации описывают ...	1) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров в функциональн
		В) Модели последовательностей описывают ...	2) правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов
		С) к описательным моделям относятся следующие модели данных:	3) регрессивные, кластеризации, исключений, итоговые и ассоциации
Задание закрытого типа с выбором нескольких правильных ответов из нескольких вариантов предложенных	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается несколько правильных ответов из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать несколько правильных ответов. 4. Записать только номера (или буквы) выбранного варианта ответа (например, 1 4 или а f).	Для анализа больших данных используют методы, включающие в себя 1 машинное обучение 2. пересказ текста 3. обработку естественного языка (NLP) 4. графовый анализ.	
Задание закрытого типа на установление последовательности	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается последовательность элементов. 2. Внимательно прочитать предложенные варианты ответа. 3. Построить верную последовательность из предложенных элементов. 4. Записать буквы/цифры (в зависимости от задания) вариантов ответа в нужной	Укажите последовательность действий при анализе больших данных: а) интерпретация результатов; б) сбор данных; с) анализ данных; д) обработка данных;	

	последовательности (например, БВА или 135).	
Задание комбинированного типа с выбором одного правильного ответа из предложенных и обоснованием выбора	1. Внимательно прочитать текст задания и понять, что в качестве ответа ожидается только один из предложенных вариантов. 2. Внимательно прочитать предложенные варианты ответа. 3. Выбрать один верный ответ. 4. Записать только номер (или букву) выбранного варианта ответа. 5. Записать аргументы, обосновывающие выбор ответа (например, 4 текст обоснования).	Алгоритм какого типа реализует функция randomForest() ... а. кластеризации б. классификации в. регрессии г. визуализации в виде деревьев
Задание открытого типа с развернутым ответом	1. Внимательно прочитать текст задания и понять суть вопроса. 2. Продумать логику и полноту ответа. 3. Записать ответ, используя четкие компактные формулировки. 4. В случае расчетной задачи, записать решение и ответ	Что является целью кластеризации?

6.3. Критерии и шкала оценивания на основе БРС.

Критерии и балльная шкала определяются преподавателем

КРИТЕРИИ ОЦЕНИВАНИЯ	РЕЗУЛЬТАТ В БАЛЛАХ
Дан полный, в логической последовательности развернутый ответ на поставленный вопрос, где он продемонстрировал знания предмета в полном объеме учебной программы, достаточно глубоко осмысливает дисциплину, самостоятельно, и исчерпывающе отвечает на дополнительные вопросы, приводит собственные примеры по проблематике поставленного вопроса, решил предложенные практические задания без ошибок	40
Дан развернутый ответ на поставленный вопрос, где студент демонстрирует знания, приобретенные на лекционных и семинарских занятиях, а также полученные посредством изучения обязательных учебных материалов по курсу, дает аргументированные ответы, приводит примеры, в ответе присутствует свободное владение монологической речью,	30-39

логичность и последовательность ответа. Однако допускается неточность в ответе. Решил предложенные практические задания с небольшими неточностями.	
Дан ответ, свидетельствующий в основном о знании процессов изучаемой дисциплины, отличающийся недостаточной глубиной и полнотой раскрытия темы, знанием основных вопросов теории, слабо сформированными навыками анализа явлений, процессов, недостаточным умением давать аргументированные ответы и приводить примеры, недостаточно свободным владением монологической речью, логичностью и последовательностью ответа. Допускается несколько ошибок в содержании ответа и решении практических заданий.	20-29
Дан ответ, который содержит ряд серьезных неточностей, обнаруживающий незнание процессов изучаемой предметной области, отличающийся неглубоким раскрытием темы, незнанием основных вопросов теории, несформированными навыками анализа явлений, процессов, неумением давать аргументированные ответы, слабым владением монологической речью, отсутствием логичности и последовательности. Выводы поверхностны. Решение практических заданий не выполнено, т.е. студент не способен ответить на вопросы даже при дополнительных наводящих вопросах преподавателя.	0-19

6.4. Описание дополнительных материалов и оборудования, необходимых для выполнения проверочных заданий (*при необходимости*).

Для решения задач открытого типа (кейсов, ПКЗ), тестовых заданий студенту разрешается использование калькулятора; программ для работы с электронными таблицами для обработки, анализа и визуализации данных. Для построения интеллект-карты и моделей в различных нотациях студенту можно использовать любой соответствующий онлайн-инструмент.

7. Методические материалы по освоению дисциплины (модуля)

Для изучения основных вопросов образовательной программы необходимо конспектировать материалы лекций, работать с рекомендованной преподавателем литературой, а также ресурсами информационно-телекоммуникационной сети «Интернет». Для приобретения навыков активного использования знаний полезно обсуждать плановые и возникающие вопросы, а также решаемые задачи на практических занятиях. Чтобы легче и

прочнее усвоить материал следует постоянно использовать конкретные примеры, сравнения из уже полученных областей наук.

Для закрепления изученного материала даны вопросы по каждой теме дисциплины, на которые следует самостоятельно найти ответы.

Важной составной частью учебного процесса в вузе являются практические занятия. Практические занятия проводятся главным образом по дисциплинам, требующим закрепления навыков решения задач, и помогают студентам глубже усвоить учебный материал, приобрести умения применять принципы системного подхода к решению разнообразных задач, определять и оценивать ресурсы и существующие ограничения разного рода проектов.

При подготовке к практическим занятиям необходимо проанализировать конспект лекции, ознакомиться с рекомендованной литературой по соответствующей теме, осуществить подготовку по рекомендованным в рабочей программе вопросам для обсуждения темы, выполнить домашнее задание (при необходимости).

Необходимо помнить, что на лекции обычно рассматривается не весь материал, а только его часть. Остальная его часть восполняется в процессе самостоятельной работы. В связи с этим работа с рекомендованной литературой обязательна. Особое внимание при этом необходимо обратить на содержание основных положений и выводов, объяснение явлений и фактов, уяснение практического приложения рассматриваемых теоретических вопросов. В процессе этой работы студент должен стремиться понять и запомнить основные положения рассматриваемого материала, примеры, поясняющие его, а также разобраться в иллюстративном материале. В процессе подготовки к занятиям рекомендуется взаимное обсуждение материала, во время которого закрепляются знания, а также приобретает практика в изложении и разъяснении полученных знаний, развивается речь. При необходимости следует обращаться за консультацией к преподавателю (в том числе по электронной почте). Планируя консультацию, необходимо хорошо продумать вопросы, которые требуют разъяснения. Заканчивать подготовку следует составлением плана (конспекта) по изучаемому материалу (вопросу). Это позволяет составить концентрированное, сжатое представление по изучаемым вопросам. Записи имеют первостепенное значение для самостоятельной работы студентов. Они помогают понять построение изучаемого материала, выделить основные положения, проследить их логику. Кроме того, ведение записей способствует превращению чтения в активный процесс, мобилизует, наряду со зрительной, и моторную память. Следует помнить: у студента, систематически ведущего записи, создается свой индивидуальный фонд методических материалов для быстрого повторения изученных вопросов, для мобилизации

накопленных знаний. Особенно важны и полезны записи тогда, когда в них находят отражение мысли, возникшие при самостоятельной работе.

После изучения базовых тем курса проводится текущий контроль знаний студентов в виде опроса или письменного тестирования. Типовые тесты и задания по темам дисциплины приведены в специальном разделе данной рабочей программы.

Подготовка к текущему и промежуточному контролю предполагает изучение представленных вопросов к зачету, работу над тестами, представленными в данной рабочей программе, выполнение семестровой проектной работы по применению системного подхода и методов системного анализа к выбранной системе.

Работа в малых группах – это одна из самых популярных форм проведения занятий, так как она дает всем обучающимся (в том числе и стеснительным) возможность участвовать в работе, практиковать навыки сотрудничества, межличностного общения (в частности, умение активно слушать, вырабатывать общее мнение, разрешать возникающие разногласия). Цель данной формы проведения занятий: продемонстрировать сходство или различия определенных явлений, выработать стратегию или разработать план, выяснить отношение различных групп участников к одному и тому же вопросу. В ходе этой работы дополнительно решаются следующие задачи: развитие навыков общения и взаимодействия в группе, формирование ценностно-ориентационного единства группы, поощрение к гибкой смене социальных ролей в зависимости от ситуации.

Группа студентов делится на несколько малых групп. Количество групп определяется числом творческих заданий, которые будут обсуждаться в процессе занятия. Малые группы формируются либо по желанию студентов, либо по родственной тематике для обсуждения. Каждая малая группа обсуждает творческое задание в течение отведенного времени. Основным этапом – проведение обсуждения творческого задания. Заслушиваются суждения, предлагаемые каждой малой группой по творческому заданию. Преподаватель дает оценочное суждение и работе малых групп, по решению творческих заданий, и эффективности предложенных путей решения.

В качестве самостоятельной работы студентами выполняется семестровая работа по применению системного подхода и методов системного анализа к выбранной системе по всем темам. Рекомендуется выбрать организационно-техническую систему. Перед выполнением задания по теме 1 выбранную систему необходимо согласовать с преподавателем. При выполнении заданий по темам могут использоваться представленные студентом материалы по предыдущим темам. Выполненная семестровая работа представляется

студентом на открытой защите на промежуточной аттестации.

8. Учебная литература и ресурсы информационно-телекоммуникационной сети Интернет

8.1. Основная литература

1. Кухаренко Б.Г. Интеллектуальные системы и технологии: учебное пособие/ сост. Кухаренко Б.Г. - Москва: Альтаир: МГАВТ, 2023. – 115с. Текст: электронный. - URL: https://biblioclub.ru/index.php?page=book_red&id=429758&sr=1 (дата обращения: 27.04.2025)
2. Наука о данных и аналитика больших объемов данных: методические рекомендации / составители А. С. Копырин, Д. И. Попов. – Сочи: СГУ, 2023. – 32 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/417155> (дата обращения: 04.08.2025). – Текст: электронный.
3. Железнов, М. М. Методы и технологии обработки больших данных: учеб.-метод. пособие / М. М. Железнов. – Москва: МИСИ – МГСУ, 2023. – 46 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/145102> (дата обращения: 04.08.2025).
4. Лебедев, А. С. Методы Big Data: учеб.-метод. пособие / А. С. Лебедев, Ш. Г. Магомедов. – Москва: РТУ МИРЭА, 2022. – 91 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/182452> (дата обращения: 04.08.2025). – Текст : электронный.

8.2. Дополнительная литература

1. Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 2-е изд., стер. — Санкт-Петербург : Лань, 2022. — 188 с. — ISBN 978-5- 8114-9690-7. — Текст : электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/198599>. — Режим доступа: для авториз. пользователей. (дата обращения: 04.08.2025).
2. Просто BIG DATA . — Санкт-Петербург : Страта, 2021. — 148 с. — ISBN 978-5-907127- 29-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/141885>. — Режим доступа: для авториз. пользователей. (дата обращения: 04.08.2025).

3. Радченко, И. А. Технологии и инфраструктура Big Data : учебное пособие / И. А. Радченко, И. Н. Николаев. — Санкт-Петербург : НИУ ИТМО, 2020. — 52 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/136430>— Режим доступа: для авториз. Пользователей (дата обращения: 04.08.2025).

8.3. Нормативные правовые документы и иная правовая информация

Не используются

8.4 Интернет-ресурсы

Обучающимся обеспечен доступ к материалам курса в СДО Академии <http://lms.ranepa.ru>, а так же через сайт научной библиотеки к следующим подписным электронным ресурсам:

Русскоязычные ресурсы

- Электронные учебники электронно-библиотечной системы (ЭБС) «Айбукс»
- Электронные учебники электронно-библиотечной системы (ЭБС) «Лань»
- Электронные учебники электронно-библиотечной системы (ЭБС) «ZNANIUM.COM»
- Электронные учебники электронно-библиотечной системы (ЭБС) «BOOK.RU»
- Электронные учебники электронно-библиотечной системы (ЭБС) «IPRSMART»

9. Материально-техническая база, информационные технологии, программное обеспечение и информационные справочные системы

№ п/п	Наименование
1.	Специализированные залы для проведения лекций, оснащенные персональным компьютером/ноутбуком и мультимедийным проектором
2.	Аудитории и компьютерные классы, оборудованные посадочными местами и персональными компьютерами с выходом в Интернет для проведения практических занятий
3.	«МТС Линк» — российская платформа для онлайн-коммуникаций и совместной работы команд ; «Яндекс Телемост» — сервис для видеоконференций от Яндекса; Я-мессенджер
4.	Технические средства обучения: персональные компьютеры;

	программные средства, обеспечивающие просмотр видеофайлов в форматах MPEG-4, DivX, RMVB, WMV; программы для работы с электронными таблицами для обработки, анализа и визуализации данных; соответствующие онлайн-инструменты.
5.	Научная библиотека (в т.ч. электронные информационные ресурсы научной библиотеки)
6.	СДО Академии https://lms.ranepa.ru/