

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
Факультет компьютерных технологий и прикладной математики

УТВЕРЖДАЮ:

Проректор по учебной работе,  
качеству образования – первый  
проректор

\_\_\_\_\_ Хагуров Т.А.

*подпись*

« 29 » августа 2025 г.

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

Б1. В.12 Машинное обучение

Направление подготовки 02.03.01 Математика и компьютерные науки

Профиль Искусственный интеллект и аналитика данных

Форма обучения очная

Квалификация бакалавр

Краснодар 2025

Рабочая программа дисциплины «Машинное обучение» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) по направлению подготовки 02.03.01 Математика и компьютерные науки

Программу составил(и):  
Н.О. Чубырь, канд. физ.-мат. наук, доцент

подпись

Рабочая программа дисциплины утверждена на заседании центра  
искусственного интеллекта  
протокол № 01 «28» августа 2025 г.  
Руководитель центра ИИ Коваленко А.В.

подпись

Утверждена на заседании учебно-методической комиссии факультета  
компьютерных технологий и прикладной математики  
протокол № 01 «28» августа 2025 г.  
Председатель УМК факультета Коваленко А.В.

подпись

Рецензенты:  
Мостовой Евгений Викторович, генеральный директор ООО «Портал-Юг»,  
e-mail: mostovoy@portal-yug.ru

Луценко Евгений Вениаминович, доктор экономических наук, кандидат технических наук, профессор кафедры компьютерных технологий и систем Федерального государственного бюджетное образовательное учреждение высшего образования «Кубанский государственный аграрный университет имени И.Т. Трубилина», e-mail: prof.lutsenko@gmail.com

## **1 Цели и задачи изучения дисциплины (модуля)**

### **1.1 Цель освоения дисциплины**

**Цели** изучения дисциплины определены государственным образовательным стандартом высшего образования и соотнесены с общими целями ООП ВО по направлению подготовки 09.03.03 Прикладная информатика, в рамках которой преподается дисциплина.

Цели дисциплины Машинное обучение:

- познакомить студентов с основными разделами искусственного интеллекта;
- научить студентов правильно выбирать методы решения задач ИИ в соответствии с поставленной задачей;
- научить студентов проводить предварительный анализ данных и подготовку данных для дальнейшего использования в задачах машинного обучения с помощью языков R и Python.

### **1.2 Задачи дисциплины:**

- изучить базовые понятия машинного обучения;
- познакомить студентов с основными этапами анализа и подготовки обучающих данных;
- изучить основные алгоритмы классического машинного обучения;
- изучить библиотеки, необходимые при работе с классическим машинным обучением на Python (Seaborn, Scikit learn, Matplotlib, Pandas) и с помощью языка R.

### **1.3 Место дисциплины (модуля) в структуре образовательной программы**

Дисциплина «Машинное обучение» относится к вариативной части Блока 1 «Дисциплины (модули)» учебного плана. Данная дисциплина («Машинное обучение») тесно связана с дисциплинами: Математический анализ, Векторная алгебра, Курс теории вероятностей, Математические модели нейронных сетей.

### **1.4 Профессиональные роли в структуре образовательной программы**

Роль 1: Data Analyst (Аналитик данных)

Задачи:

1. Статистический анализ, визуализация данных, предварительная обработка.
2. Создание прогнозных моделей
3. Построение аналитических моделей для поддержки бизнес-решений.

Роль 2: MLOps (Специалист по эксплуатации ИИ)

Задачи:

1. DevOps для ML.
2. Автоматизация, мониторинг ML-систем.
3. Операционное управление жизненным циклом ML-моделей.

Роль 3: AI PM (Менеджер проектов ИИ)

Задачи:

1. Управление ИИ-проектами от идеи до внедрения
2. Анализ бизнес-требований и постановка задач
3. Оценка эффективности и ROI ИИ-решений

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

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

Код и наименование индикатора*	Результаты обучения по дисциплине
<b>ПК-4</b> Способность анализировать цифровой след в соответствии с моделью деятельности человека (группы людей) и информационно-коммуникационной системой (ИКС) для выявления закономерностей, прогнозирования поведения и принятия управленческих решений	
<b>ПК-4.3</b> Способен провести оценку этических и правовых аспектов работы с цифровым следом	Выявляет потенциальные этические и правовые проблемы во всех этапах жизненного цикла ML-проекта. Проводит аудит набора данных на предмет наличия скрытых смещений, репрезентативности выборки и законности источников их получения
<b>ML-3</b> Способен применять классические алгоритмы машинного обучения с пониманием их математических основ и областей применения	
<b>ML-3.2</b> Эффективно применяет классические методы и модели машинного обучения для обеспечения достижимости функциональных характеристик систем ИИ	Применяет методы байесовской классификации и ансамблевых методов МО (бэггинг, бустинг, стэкинг моделей), а также производных от них (случайные леса, градиентный бустинг на деревьях). Применяет классические методы МО для временных рядов (ARIMA, экспоненциальное сглаживание, линейная регрессия с лагами)
	Владеет инструментами оценки качества моделей ранжирования и сравнения ранжирующих моделей между собой. Владет методами обучения типа pairwise и listwise. Знает и применяет на практике различные архитектуры ранжированного поиска (одно-двух-трехстадийное ранжирование)
<b>ML-4</b> Способен применять методы обучения без учителя для анализа структуры данных и выявления скрытых закономерностей	
<b>ML-4.1</b> Применяет алгоритмы кластеризации и понижения размерности для решения практических задач	Владеет инструментами очистки данных и предварительной подготовки данных методами понижения размерности и визуализации для анализа данных.
<b>ML-4.2</b> Выявляет аномалии и применяет методы поиска ассоциативных правил	Настраивает и применяет алгоритмы обнаружения аномалий (статистические методы isolation forest one-class SVM) и ассоциативного анализа (Apriori, FP-Growth) с учётом структуры и особенностей реальных данных

Результаты обучения по дисциплине достигаются в рамках осуществления всех видов контактной и самостоятельной работы обучающихся в соответствии с утвержденным учебным планом.

Индикаторы достижения компетенций считаются сформированными при достижении соответствующих им результатов обучения.

## 2. Структура и содержание дисциплины

### 2.1 Распределение трудоёмкости дисциплины по видам работ

Общая трудоёмкость дисциплины составляет 3 зачетных единицы (108 часов), их распределение по видам работ представлено в таблице

Виды работ	Всего Часов	Форма обучения
		Очная
		6 семестр (часы)
<b>Контактная работа, в том числе:</b>	<b>52,2</b>	<b>52,2</b>
<b>Аудиторные занятия (всего):</b>	<b>48</b>	<b>48</b>
занятия лекционного типа	16	16
лабораторные занятия	34	34

практические занятия	-	-
семинарские занятия	-	-
<b>Иная контактная работа:</b>	<b>4,2</b>	<b>4,2</b>
Контроль самостоятельной работы (КСР)	2	2
Промежуточная аттестация (ИКР)	0,2	0,2
<b>Самостоятельная работа, в том числе:</b>	<b>55,8</b>	<b>55,8</b>
Проработка и повторение лекционного материала и материала учебников и учебных пособий	15,8	15,8
Подготовка к текущему контролю	40	40
<b>Контроль:</b>	<b>-</b>	<b>-</b>
Подготовка к экзамену	-	-
<b>Общая трудоемкость</b>	<b>час.</b>	<b>108</b>
	<b>в том числе контактная работа</b>	<b>52,2</b>
	<b>зач. Ед</b>	<b>3</b>

## 2.2 Содержание дисциплины

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы (темы) дисциплины, изучаемые в 5 семестре (очная форма обучения)

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная Работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1.	Предобработка данных	24	4		10	10
2.	Основы классического машинного обучения	22	12			10
3.	Машинное обучение на Python и R	59,8			24	35,8
	<i>ИТОГО по разделам дисциплины</i>	<b>103,8</b>	<b>16</b>		<b>32</b>	<b>55,8</b>
	Контроль самостоятельной работы (КСР)	2				
	Промежуточная аттестация (ИКР)	0,2				
	Подготовка к экзамену	-				
	Общая трудоемкость по дисциплине	<b>108</b>	<b>16</b>		<b>34</b>	<b>55,8</b>

Примечание: Л – лекции, ПЗ – практические занятия / семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

## 2.3 Содержание разделов (тем) дисциплины

### 2.3.1 Занятия лекционного типа

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
1.	Предобработка данных	Переменные. Датафреймы. Элементы синтаксиса. Описательные статистики. Сохранение результатов. Подготовка данных для машинного обучения	Опрос
2.	Основы классического машинного обучения	Введение в ИИ, области ИИ. Виды машинного обучения. История и тренды ИИ. Подготовка данных в задачах машинного обучения. Алгоритм градиентного спуска. Задача классификации. Логистическая регрессия. Постановка задачи, входные данные и алгоритм Задача регрессии. Линейные модели. Переобучение. Метрики качества классификации. Метрики качества регрессии Метод К ближайших соседей. Постановка задачи кластеризации. Алгоритм DBSCAN. Постановка задачи кластеризации. Алгоритм k-means. Решающие деревья. Критерий Джини и энтропии Случайные леса. Ансамблевые методы: стекинг,	Т

№	Наименование раздела (темы)	Содержание раздела (темы)	Форма текущего контроля
1	2	3	4
		бэггинг, бустинг. Задача понижения размерности. Отбор признаков. Задача понижения размерности. Выделение новых признаков на основе старых. Ассоциативные правила. Алгоритм Argiori. Рекомендательные системы	
3	Машинное обучение на Python и R	Линейная и логистическая регрессия на языке R. Основы библиотек Matplotlib, Seaborn и Pandas. Линейная и логистическая регрессия на Python. Понижение размерности с помощью библиотеки Sklearn. Кластеризация на Python и R. Деревья решений и случайные леса в библиотеке Sklearn	Т

### 2.3.2 Занятия семинарского типа (практические / семинарские занятия/ лабораторные работы)

№	Наименование раздела (темы)	Наименование лабораторных работ	Форма текущего контроля
1	2	3	4
1.	Предобработка данных	Переменные. Датафреймы. Элементы синтаксиса. Описательные статистики. Сохранение результатов. Визуализация данных.	ЛР
2	Основы классического машинного обучения	Алгоритм градиентного спуска. Задача классификации. Алгоритм DBSCAN. Постановка задачи кластеризации. Алгоритм k-means. Решающие деревья. Критерий Джини и энтропии Случайные леса. Ансамблевые методы: стекнинг, бэггинг, бустинг. Задача понижения размерности. Отбор признаков. Задача понижения размерности.	ЛР
3.	Машинное обучение на Python и R	Линейная и логистическая регрессия на языке R. Основы библиотек Matplotlib, Seaborn и Pandas. Линейная и логистическая регрессия на Python. Понижение размерности с помощью библиотеки Sklearn. Кластеризация на Python и R. Деревья решений и случайные леса в библиотеке Sklearn	ЛР

Защита лабораторной работы (ЛР), выполнение курсового проекта (КП), курсовой работы (КР), расчетно-графического задания (РГЗ), написание реферата (Р), эссе (Э), коллоквиум (К), тестирование (Т), решение кейсов (РК) и т.д.

### 2.3.3 Курсовые работы не предусмотрены

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

№	Вид СРС	Перечень учебно-методического обеспечения дисциплины по выполнению самостоятельной работы
1	2	3
1	Проработка и повторение лекционного материала, материала учебной и научной литературы, подготовка к семинарским	Методические указания для подготовки к лекционным и семинарским занятиям, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г. Методические указания по выполнению самостоятельной работы,

	занятиям	утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.
2	Подготовка к лабораторным занятиям	Методические указания по выполнению лабораторных работ, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.
3	Подготовка к решению задач и тестов	Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.
4	Подготовка к текущему контролю	Методические указания по выполнению самостоятельной работы, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.
5	Подготовка докладов	Методические указания для подготовки эссе, рефератов, курсовых работ, утвержденные на заседании кафедры анализа данных и искусственного интеллекта факультета компьютерных технологий и прикладной математики ФГБОУ ВО «КубГУ», протокол №7 от 22.03.2023 г.

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

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

### **3. Образовательные технологии, применяемые при освоении дисциплины (модуля)**

В соответствии с требованиями ФГОС в программа дисциплины предусматривает использование в учебном процессе следующих образовательные технологии: чтение лекций с использованием мультимедийных технологий; метод малых групп, разбор практических задач и кейсов.

При обучении используются следующие образовательные технологии:

– Технология коммуникативного обучения – направлена на формирование коммуникативной компетентности студентов, которая является базовой, необходимой для адаптации к современным условиям межкультурной коммуникации.

– Технология разноуровневого (дифференцированного) обучения – предполагает осуществление познавательной деятельности студентов с учётом их индивидуальных способностей, возможностей и интересов, поощряя их реализовывать свой творческий потенциал. Создание и использование диагностических тестов является неотъемлемой частью данной технологии.

– Технология модульного обучения – предусматривает деление содержания дисциплины на достаточно автономные разделы (модули), интегрированные в общий курс.

– Информационно-коммуникационные технологии (ИКТ) - расширяют рамки образовательного процесса, повышая его практическую направленность, способствуют интенсификации самостоятельной работы учащихся и повышению познавательной активности. В рамках ИКТ выделяются 2 вида технологий:

– Технология использования компьютерных программ – позволяет эффективно дополнить процесс обучения языку на всех уровнях.

– Интернет-технологии – предоставляют широкие возможности для поиска информации, разработки научных проектов, ведения научных исследований.

– Технология индивидуализации обучения – помогает реализовывать личностно-ориентированный подход, учитывая индивидуальные особенности и потребности учащихся.

– Проектная технология – ориентирована на моделирование социального взаимодействия учащихся с целью решения задачи, которая определяется в рамках профессиональной подготовки, выделяя ту или иную предметную область.

– Технология обучения в сотрудничестве – реализует идею взаимного обучения, осуществляя как индивидуальную, так и коллективную ответственность за решение учебных задач.

– Игровая технология – позволяет развивать навыки рассмотрения ряда возможных способов решения проблем, активизируя мышление студентов и раскрывая личностный потенциал каждого учащегося.

– Технология развития критического мышления – способствует формированию разносторонней личности, способной критически относиться к информации, умению отбирать информацию для решения поставленной задачи.

Комплексное использование в учебном процессе всех вышеназванных технологий стимулируют личностную, интеллектуальную активность, развивают познавательные процессы, способствуют формированию компетенций, которыми должен обладать будущий специалист.

Основные виды интерактивных образовательных технологий включают в себя:

– работа в малых группах (команде) - совместная деятельность студентов в группе под руководством лидера, направленная на решение общей задачи путём творческого сложения результатов индивидуальной работы членов команды с делением полномочий и ответственности;

– проектная технология - индивидуальная или коллективная деятельность по отбору, распределению и систематизации материала по определенной теме, в результате которой составляется проект;

– анализ конкретных ситуаций - анализ реальных проблемных ситуаций, имевших место в соответствующей области профессиональной деятельности, и поиск вариантов лучших решений;

– развитие критического мышления – образовательная деятельность, направленная на развитие у студентов разумного, рефлексивного мышления, способного выдвинуть новые идеи и увидеть новые возможности.

Подход разбора конкретных задач и ситуаций широко используется как преподавателем, так и студентами во время лекций, лабораторных занятий и анализа результатов самостоятельной работы. Это обусловлено тем, что при исследовании и решении каждой конкретной задачи имеется, как правило, несколько методов, а это требует разбора и оценки целой совокупности конкретных ситуаций.

#### **Занятия, проводимые с использованием интерактивных технологий**

№	Наименование разделов (тем)	Количество часов	
		всего ауд. часов	интерактивные часы

1.	Предобработка данных	14	2
2.	Основы классического машинного обучения	12	4
3.	Машинное обучение на Python и R	24	6
	<i>Итого по дисциплине:</i>	<b>50</b>	<b>12</b>

Темы, задания и вопросы для самостоятельной работы призваны сформировать навыки поиска информации, умения самостоятельно расширять и углублять знания, полученные в ходе лекционных и практических занятий.

Подход разбора конкретных ситуаций широко используется как преподавателем, так и студентами при проведении анализа результатов самостоятельной работы.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Для лиц с ограниченными возможностями здоровья предусмотрена организация консультаций с использованием электронной почты.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

#### **4. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации**

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Машинное обучение».

Оценочные средства включает контрольные материалы для проведения **текущего контроля** в форме тестовых заданий и лабораторных работ, и **промежуточной аттестации** в форме вопросов к зачету.

Текущий контроль успеваемости осуществляется в течение семестра, в ходе повседневной учебной работы и предполагает овладение материалами лекций, литературы, программы, работу студентов в ходе проведения лабораторных занятий, а также систематическое выполнение тестовых работ, решение практических задач и иных заданий для самостоятельной работы студентов. Данный вид контроля стимулирует у студентов стремление к систематической самостоятельной работе по изучению дисциплины. Он предназначен для оценки самостоятельной работы слушателей по решению задач, выполнению лабораторных работ, подведения итогов тестирования. Оценивается также активность и качество результатов практической работы на занятиях, участие в дискуссиях, обсуждениях и т.п. Индивидуальные и групповые самостоятельные, аудиторные работы по всем темам дисциплины организованы единообразным образом. Для контроля освоения содержания дисциплины используются оценочные средства. Они направлены на определение степени сформированности компетенций.

Промежуточная аттестация студентов осуществляется в рамках завершения изучения дисциплины и позволяет определить качество усвоения изученного материала, предполагает контроль и управление процессом приобретения студентами необходимых знаний, умения и навыков, определяемых по ФГОС ВО по соответствующему направлению подготовки в качестве результатов освоения учебной дисциплины.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на зачете;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

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

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

*ПК-4.1; ПК-4.2; ПК-4.3; LC-3.1; LC-3.2; ML-3.1; ML-3.2; ML-3.3; ML-4.1; ML-4.2; ML-4.3; ML-5.1; ML-5.2; ML-6.1; ML-6.2; ML-7.1; ML-7.2; ML-8.1; ML-8.2; SS-3.1; SS-3.2; SS-3.3; MF-1.1; MF-1.2; MF-2 .1; MF-4.1*

### Структура оценочных средств для текущей и промежуточной аттестации

№ п/п	Контролируемые разделы (темы) дисциплины*	Код контролируемой компетенции (или ее части)	Наименование оценочного средства	
			Текущий контроль	Промежуточная аттестация
1	Тема 1. Переменные. Датафреймы. Элементы синтаксиса. Описательные статистики. Сохранение результатов.  Тема 2. Подготовка данных для машинного обучения	<i>ПК-4; LC-3. ML-3; ML-4; ML-5; ML-6; ML-7, ML-8; SS-3; MF-1; MF-2; MF-4</i>	<i>Тест, Лабораторная работа №1 Лабораторная работа №2</i>	<i>Вопросы к зачету 1-11</i>
2	Тема 1. Введение в ИИ, области ИИ. Виды машинного обучения. История и тренды ИИ. Подготовка данных в задачах машинного обучения.  Тема 2. Алгоритм градиентного спуска. Задача классификации	<i>ПК-4; LC-3. ML-3; ML-4; ML-5; ML-6; ML-7, ML-8; SS-3; MF-1; MF-2; MF-4</i>	<i>Лабораторная работа №3 Лабораторная работа №4, Лабораторная работа №5</i>	<i>Вопросы к зачету 12-17</i>

	<p>Тема 3. Логистическая регрессия. Постановка задачи, входные данные и алгоритм</p> <p>Задача регрессии. Линейные модели. Переобучение. Метрики качества классификации. Метрики качества регрессии</p> <p>Тема 4. Метод К ближайших соседей. Постановка задачи кластеризации. Алгоритм DBSCAN. Постановка задачи кластеризации. Алгоритм k-means. Решающие деревья. Критерий Джини и энтропии</p> <p>Тема 4. Случайные леса. Ансамблевые методы: стекинг, бэггинг, бустинг. Задача понижения размерности. Отбор признаков. Задача понижения размерности. Выделение новых признаков на основе старых. Ассоциативные правила. Алгоритм Apriori. Рекомендательные системы</p> <p>.</p>			
3	<p>Тема 1. Линейная и логистическая регрессия на языке R. Основы библиотек Matplotlib, Seaborn и Pandas.</p> <p>Тема 2. Линейная и логистическая регрессия на Python.</p> <p>Тема 2. Понижение размерности с помощью библиотеки Sklearn. Кластеризация на Python и R.</p> <p>Тема 2. Деревья решений и случайные леса в</p>	<p><i>ПК-4; LC-3. ML-3; ML-4; ML-5; ML-6; ML-7, ML-8; SS-3; MF-1; MF-2; MF-4</i></p>	<p><i>Лабораторная работа №6 Лабораторная работа №7, Лабораторная работа №8</i></p>	<p><i>Вопросы к зачету 18-23</i></p>

	библиотеке Sklearn.			
--	---------------------	--	--	--

### Показатели, критерии и шкала оценки сформированных компетенций

Соответствие **пороговому уровню** освоения компетенций планируемым результатам обучения и критериям их оценивания (оценка: **зачтено**):

**ПК-4** Способность анализировать цифровой след в соответствии с моделью деятельности человека (группы людей) и информационно-коммуникационной системой (ИКС) для выявления закономерностей, прогнозирования поведения и принятия управленческих решений

**ПК-4.3** Способен провести оценку этических и правовых аспектов работы с цифровым следом

**Знать** Основные источники цифрового следа, методы предобработки данных, форматы данных. Основы машинного обучения, классификация, кластеризация, регрессия

**Уметь** Извлекать данные из API или баз данных, очищать и трансформировать сырые данные (удаление дубликатов, кодирование категориальных признаков), визуализировать данные для первичного анализа. Строить простые модели прогнозирования, интерпретировать результаты, формулировать рекомендации на основе данных, применять методы сбора и обработки цифрового следа для построения и анализа моделей

**Владеть** Навыками работы с Python, стандартными библиотеками Python (например, Scikit-Learn, TensorFlow, PyTorch). Навыком оценки этических и правовых аспектов работы с цифровым следом.

**ML-3** Способен применять классические алгоритмы машинного обучения с пониманием их математических основ и областей применения

**ML-3.2** Эффективно применяет классические методы и модели машинного обучения для обеспечения достижимости функциональных характеристик систем ИИ

**Знать** Основные категории алгоритмов: классификация, регрессия, кластеризация, ансамбли. Преимущества и ограничения алгоритмов. Методы адаптации моделей (регуляризация, балансировка классов, подбор признаков. Метрики качества для разных задач. Методы ускорения вычислений. Статистические методы сравнения, методы интерпретации моделей.

**Уметь** Выбирать подходящие алгоритмы для конкретной задачи, реализовывать базовые модели с помощью Scikit-Learn, интерпретировать параметры модели. Сравнить алгоритмы по производительности, обосновывать выбор гиперпараметров, анализировать переобучение. Оптимизировать производительность. Проводить статистические тесты, анализировать ошибки моделей, формулировать рекомендации

- Владеть** Навыками работы с кросс-валидации, методами предобработки данных. Визуализацией результатов сравнения, метриками бизнес-эффективности
- ML-4** **Способен применять методы обучения без учителя для анализа структуры данных и выявления скрытых закономерностей**
- ML-4.1 Применяет алгоритмы кластеризации и понижения размерности для решения практических задач
- ML-4.2 Выявляет аномалии и применяет методы поиска ассоциативных правил
- Знать** Основные задачи (кластеризация, понижение размерности, поиск аномалий). Метрики для кластеризации, методы визуальной оценки..
- Уметь** Выбирать подходящие алгоритмы для конкретной задачи, подготавливать данные для обучения без учителя, настраивать гиперпараметры (например, количество кластеров K-means). Оценивать качество результатов обучения без учителя. Выяснять и объяснять аномалии, формулировать выводы на основе выявленных закономерностях
- Владеть** Навыками нормализации и стандартизации данных, работой с категориальными признаками, обработкой пропущенных значений. Навыками работы с Python, стандартными библиотеками Python (например, Scikit-Learn, TensorFlow, PyTorch). Навыками подготовки отчетов с описанием методов и результатов

**4.2 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы**

### **Пример тестирования по теме «Введение в ИИ, области ИИ»**

Какое из перечисленных понятий НЕ входит в понятие ИИ?

- а) Глубинное обучение
- б) Аналитика данных
- в) Экспертные системы
- г) Машинное обучение

Выберите верное утверждение

- а) Искусственный интеллект – это сложное понятие, не имеющее четкого определения и включающее различные области математики, информационных технологий и др.
- б) Искусственный интеллект - это робот для общения с людьми посредством текстового интерфейса, разработанный Аланом Тьюрингом в 1950 году
- в) Искусственный интеллект - это четко определенное понятие, означающее создание машины, повторяющей умственные процессы человека

Для чего может быть полезно применять ИИ в банке (несколько правильных ответов)?

- а) Автоматизация выдачи наличных средств

- б) Автоматизация работы всего персонала банка
- в) Автоматизация обработки документов
- г) Автоматизация работы с клиентами

В чем состоит тест Тьюринга?

- а) Человек получает ответы на вопросы от другого человека и от компьютера «вслепую» и должен определить, кто из собеседников – компьютер
- б) Машина должна «выжить» в сложной, искусственно заданной среде, с которой она взаимодействует посредством некоторого механизма
- в) Человеку показывают серию картин, и он должен выделить те, которые созданы машиной

Продолжите фразу:

Общий искусственный интеллект

- а) Решает разнообразные сложные интеллектуальные и творческие задачи (на сегодня не представляется возможным разработать)
- б) Решает конкретные интеллектуальные задачи (разработаны системы для различных задач)
- в) Решает четко поставленные задачи, для которых известны конкретные эффективные алгоритмы (используется повсеместно)

Продолжите фразу:

Узко-специализированный искусственный интеллект

- а) Решает разнообразные сложные интеллектуальные и творческие задачи (на сегодня не представляется возможным разработать)
- б) Решает конкретные интеллектуальные задачи (разработаны системы для различных задач)
- в) Решает четко поставленные задачи, для которых известны конкретные эффективные алгоритмы (используется повсеместно)

Продолжите фразу:

Программирование

- а) Решает разнообразные сложные интеллектуальные и творческие задачи (на сегодня не представляется возможным разработать)
- б) Решает конкретные интеллектуальные задачи (разработаны системы для различных задач)
- в) Решает четко поставленные задачи, для которых известны конкретные эффективные алгоритмы (используется повсеместно)

Выберите верное утверждение: Современный искусственный интеллект основывается на...

- а) Обучении алгоритмов, способных решать задач, аналогичные тем, что решает человек
- б) Изучении и компьютерном повторении структуры человеческого мозга
- в) Создании искусственного мозга на основе биотехнологий

### **Примеры типовых лабораторных работ Лабораторная работа по теме: «Переменные»**

**Задание**

В векторе `my_vector` отберите только те наблюдения, которые отклоняются от среднего меньше чем на одно стандартное отклонение. Сохраните эти наблюдения в новую переменную `my_vector_2`.

При этом исходный вектор `my_vector` оставьте без изменений:  
21 18 21 19 25 20 17 17 18 22 17 18 18 19 19 27 21 20 24 17 15 24 24 29 19 14 21 17 19  
18 18 20 21 21 19 19 17 21 13 17 13 23 15 23 24 16 17 25 24 22

### Лабораторная работа по теме: «Датафреймы»

#### Задание 1

В датафрейме `mtcars` создайте новую колонку (переменную) под названием `even_gear`, в которой будут единицы, если значение переменной (`gear`) четное, и нули если количество нечетное.

#### Задание 2

В датафрейме `mtcars` создать переменную – вектор `mpg_4` и сохранить в нее значения расхода топлива (`mpg`) для машин с четырьмя цилиндрами (`cyl`).

#### Задание 3

Создать новый `dataframe` под названием `mini_mtcars`, в котором будут сохранены только третья, седьмая, десятая, двенадцатая и последняя строчка датафрейма `mtcars`.

### Лабораторная работа по теме: «Элементы синтаксиса»

#### Задача №1

Создайте новую числовую переменную `new_var` в данных `mtcars`, которая содержит единицы в строчках, если в машине не меньше четырёх карбюраторов (переменная `"carb"`) или больше шести цилиндров (переменная `"cyl"`). В строчках, в которых условие не выполняется, должны стоять нули.

#### Задача №2

В переменной `my_vector` сохраните вектор из 50 чисел:  
20.67 23.34 22.65 17.11 22.1 26.32 20.39 21.04 23.78 31.11 21.13 22.44 23.21 27.02  
18.64 20.9 20.77 20.0 21.29 23.48 18.47 25.02 17.04 30.97 12.91 23.88 32.95 8.46 23.15 21.05  
20.63 19.95 17.38 29.35 24.43 23.66 18.32 30.13 19.36 19.67 24.23 20.82 18.21 9.91 21.45  
18.04 18.31 17.18 10.99 10.06

Решите задачу используя конструкцию:

```
if () {  
  } else {  
  }  
}
```

Если среднее значение вектора `my_vector` больше 20, в переменную `result` сохраните "My mean is great", если среднее значение `my_vector` меньше или равно 20 то в переменную `result` сохраните строку "My mean is not so great".

#### Задача №3

В этой задаче от вас потребуется узнать некоторую информацию о типах данных в R самостоятельно! Встроенные в R данные `AirPassengers` - это новый для нас формат данных типа `Time-Series`. Изучите структуру этих данных, прежде чем начать решение задачи! Например напишите команды:

```
> ?AirPassengers # справка о данных  
> str(AirPassengers) # структура данных
```

Во встроенных в R данных `AirPassengers` хранится 144 значения (количество пассажиров в месяц) с 1949 по 1960 год. Данные `Time-Series` очень похожи на вектор по своей структуре, например мы можем обратиться к любому из 144 элементов используя уже знакомую нам индексацию `AirPassengers[1]` или `AirPassengers[56]`.

Можно вообще перевести исходные данные в вектор при помощи команды `as.vector(AirPassengers)` и продолжить с ними работу как с вектором.

И так ваша задача создать переменную `good_months` и сохранить в нее число пассажиров только в тех месяцах, в которых это число больше, чем показатель в предыдущем месяце.

Важный момент! В R оператор : для создания последовательности имеет приоритет над арифметическими действиями. Таким образом, если у вас есть переменная  $i$ , равная 10, и вы хотите создать вектор от 1 до  $i - 1$ , воспользуйтесь скобками, чтобы указать последовательность действий.

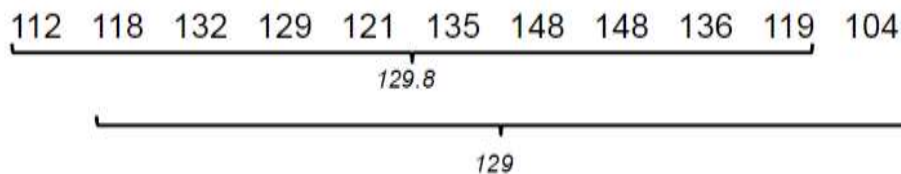
```
> i <- 10
> 1 : i - 1 # так мы создадим последовательность от 1 до 10, а потом вычтем
единицу из каждого элемента
[1] 0 1 2 3 4 5 6 7 8 9
> 1 : (i - 1) # а вот так мы создадим последовательность от 1 до i - 1, то есть от 1 до
9.
[1] 1 2 3 4 5 6 7 8 9
```

#### Задача №4

Для встроенных в R данных `AirPassengers` рассчитайте скользящее среднее с интервалом сглаживания равным 10. Напечатайте получившийся результат (первым значением в выводе должно быть среднее для элементов 1:10, во втором значении - среднее для элементов 2:11 и т.д., в последнем - среднее для элементов 135 :144)

Все полученные значения средних сохраните в переменную `moving_average`.

Пример расчета для вектора из 11 элементов:



Соответственно, для наших данных из 144 наблюдений должно получиться 135 средних (первые и последние 4 средних):

```
129.8 129.0 129.0 127.3 ... 483.6 489.2 486.5 490.6
```

Если вам потребуется создать вектор `moving_average` заранее, то есть несколько способов сделать это:

1. самый простой, но не очень правильный вариант - создать пустой вектор `moving_average <- c()`
2. можно сразу создать вектор определенной длины и определенного типа: `moving_average <- numeric(135)`

Такой вариант является более предпочтительным. Также можно познакомиться с функцией `cumsum`. Подсказка: если у нас есть два вектора одинаковой длины, то если из одного вектора вычесть второй вектор, мы найдем разность для первых элементов векторов, затем для вторых и т.д.

```
> x <- c(2, 4, 7)
> y <- c(2, 3, 5)
> x - y
[1] 0 1 2
```

#### Лабораторная по теме «Основы библиотек Matplotlib, Seaborn и Pandas».

##### Задание №1

Используя данные `mtcars`, рассчитайте средний расход топлива (mpg) для автомобилей с числом лошадиных сил (hp), большим 120 и у которых вес менее 4000 фунтов.

Получившийся результат (среднее значение) сохраните в переменную `result`.

##### Задание №2

Рассчитайте стандартное отклонение переменной mpg (расход топлива), переменной disp (вместимости двигателя) у машин с автоматической и ручной коробкой передач.

Полученные результаты сохраните в переменную descriptions\_stat.

Задание №3

Воспользуйтесь данными airquality. В новую переменную сохраните подмножество исходных данных, оставив наблюдения только для месяцев 5, 6 и 7.

Рассчитайте количество непропущенных наблюдений по переменной Solar.R (солнечная радиация) в 5, 6 и 7 месяце.

Результат выполнения сохраните в переменную result1.

Задание №4

Примените функцию describe к количественным переменным данным airquality, группируя наблюдения по переменной Month. Чему равен коэффициент асимметрии (skew) переменной Wind в седьмом месяце?

Задание №5

Создайте случайный вектор my\_vector (выборка из нормального распределения) с пропущенными значениями, состоящий из 30 элементов.

Вам нужно создать новый вектор fixed\_vector, в котором все пропущенные значения вектора my\_vector будут заменены на среднее значение по имеющимся наблюдениям.

При этом исходный вектор оставьте без изменений!

Задание №6

Постройте график boxplot, используя данные airquality. По оси x отложите номер месяца, по оси y — значения переменной Temp.

На графике boxplot отдельными точками отображаются наблюдения, отклоняющиеся от 1 или 3 квартиля больше чем на полтора межквартильных размаха. Сколько таких наблюдений присутствует в июле (месяц №7)?

Задание №7

Используя данные mtcars, нужно построить Scatterplot (диаграмма рассеивания), по оси x которого будет hp, по оси y — qsec, а цветом отобразить переменную (mpg).

Задание №8

Основываясь на данных iris постройте график Scatterplot (диаграмма рассеивания), где по оси X будет отложена переменная Sepal.Length, по оси Y переменная Sepal.Width. За цвет точек будет отвечать переменная Species, а за размер точек переменная Petal.Length.

**Проверяемые компетенции комплексом практических заданий: ПК-4; LC-3. ML-3; ML-4; ML-5; ML-6; ML-7, ML-8; SS-3; MF-1; MF-2; MF-4**

**Зачетно-экзаменационные материалы для промежуточной аттестации (экзамен/зачет)**

### **Вопросы к зачёту**

1. История развития ИИ. Основные этапы.
2. Области искусственного интеллекта
3. Подготовка данных в задачах машинного обучения
4. Алгоритм градиентного спуска
5. Задача классификации. Логистическая регрессия. Постановка задачи, входные данные и алгоритм
6. Задача регрессии. Линейные модели. Переобучение
7. Метрики качества классификации

8. Метрики качества регрессии
9. Метод К ближайших соседей
10. Постановка задачи кластеризации. Алгоритм DBSCAN
11. Постановка задачи кластеризации. Алгоритм k-means
12. Решающие деревья. Критерий Джини и энтропии
13. Случайные леса
14. Ансамблевые методы: стекинг, бэггинг, бустинг.
15. Задача понижения размерности. Отбор признаков.
16. Задача понижения размерности. Выделение новых признаков на основе старых
17. Ассоциативные правила. Алгоритм Apriori
18. Рекомендательные системы
19. Подготовка данных для машинного обучения с помощью библиотеки Pandas
20. Предобработка и визуализация данных с помощью языка R
21. Визуализация данных с помощью Matplotlib и Seaborn
22. Классическое машинное обучение и подсчёт метрик с помощью библиотеки Sklearn
23. Классическое машинное обучение на R

*Проверяемые компетенции комплексом практических заданий: ПК-4; LC-3. ML-3; ML-4; ML-5; ML-6; ML-7, ML-8; SS-3; MF-1; MF-2; MF-4*

#### **4.2 Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций**

##### **Методические рекомендации, определяющие процедуры оценивания на экзамене:**

Процедура промежуточной аттестации проходит в соответствии с Положением о текущем контроле и промежуточной аттестации обучающихся ФГБОУ ВО «КубГУ».

Итоговой формой контроля сформированности компетенций у обучающихся по дисциплине является зачет. Студенты обязаны сдать экзамен в соответствии с расписанием и учебным планом.

##### **Критерии оценивания результатов обучения**

Оценка	Критерии оценивания по зачету
зачтено	оценку «зачтено» заслуживает студент, освоивший знания, умения, компетенции и теоретический материал (Разделы 1-3); выполнивший все лабораторные работы, кейсы и сдавший тестирование выше проходного балла.
незачтено	оценку «незачтено» заслуживает студент не выполнивший лабораторные работы, кейсы и не освоивший теоретический материал (разделы 1-3).

Для допуска к зачету необходимо выполнить **все лабораторные работы и практические кейсы.**

**Практические кейсы** оцениваются по:

- Корректности кода.
- Достижению целевых показателей обучения.
- Качеству отчета (анализ ошибок, визуализация).

**Методические рекомендации, определяющие процедуры оценивания лабораторных работ:**

Процедура оценивания лабораторных работ проходит в соответствии с Положением о текущем контроле и промежуточной аттестации обучающихся ФГБОУ ВО «КубГУ».

По каждой лабораторной работе оформляется отчет. Отчеты сдаются на проверку руководителю в течение курса по мере их выполнения, и защищаются студентами в установленном порядке.

При защите отчета студенту могут быть заданы вопросы и дополнительные задания по сути лабораторной работы, в том числе из списка контрольных вопросов к данной лабораторной работе. При неудовлетворительной оценке знаний студента по теме данного отчета, студент возвращается к повторному изучению соответствующих материалов, после чего допускается к повторной защите. Неудовлетворительно выполненный отчет также возвращается на доработку.

Отчет должен содержать заголовок, тему лабораторной работы, цель, задание, индивидуальную тему, описание хода выполнения работы, необходимые прикладные материалы (схемы, макеты документов и т.п.), в соответствии с требованиями к содержанию, и выводы по работе.

Оценочные средства для инвалидов и лиц с ограниченными возможностями здоровья выбираются с учетом их индивидуальных психофизических особенностей.

– при необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на зачете;

– при проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями;

– при необходимости для обучающихся с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

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

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

### **4.3. Методические указания по организации вычислительной инфраструктуры**

#### **Условия применения:**

- Курс рассчитан на студентов 3-го года обучения (6 семестр).
- Наличие доступа к вычислительным ресурсам (GitLab, Google Colab или Yandex DataSphere, JupyterHub, VSCode, среда Matlab).
- Разработаны лабораторные работы;
- Инфраструктура для приёма задач (gitlab, CI/CD) согласована с лабораторными работами.

## **Цели, задачи и ожидаемые результаты**

Цели организации вычислительной инфраструктуры:

- дать начальное представление о работе в IT инфраструктуре (приучить пользоваться гитом, jupyter-ноутбуками).

Задачи преподавателя:

- Организация регистрации студентов в Google Colab и Yandex DataSphere
- Создание учетных записей студентов в gitlab вуза;
- GitLab Runner;
- Работа над шаблонным репозиторием лабораторных работ;
- Написание инструкции для студентов.

Ожидаемые результаты студентов:

- начальное представление о работе в IT инфраструктуре (гит, нейминг).

## **Порядок реализации**

Задача №1: Организация регистрации студентов в Google Colab и Yandex DataSphere

Задача №2: Создание учетных записей студентов в gitlab вуза

Задача №3: GitLab Runner.

Для выполнения CI/CD пайплайна был настроен GitLab Runner на удалённой виртуальной машине с ОС Ubuntu 24.04.

Последовательность настройки включала следующие шаги:

- Настройка системы – установка необходимых компонентов, таких, как Docker.
- Установка GitLab Runner по официальной инструкции.
- Регистрация Runner для частного сервера GitLab.

Задача №4: Работа над шаблонным репозиторием лабораторных работ

Этот репозиторий служит основой для всех новых лабораторных заданий и содержит преднастроенную структуру проекта и CI-конфигурации.

Ключевые файлы и их назначение:

.gitlab-ci.yml — основной конфигурационный файл CI/CD. Определяет среду выполнения (образ Docker), команды для сборки.

Dockerfile – файл для запуска докер-образа.

Задача №5: Написание инструкции для студентов

README.md – инструкции к использованию шаблонного репозитория.

Вся структура максимально адаптирована для копирования студентами и минимизации порога входа при выполнении лабораторных

## **4.4. Методические указания по организации лабораторных работ по дисциплине**

**Условия применения:**

- Курс рассчитан на студентов 3-го года обучения.
- Наличие доступа к вычислительным ресурсам (GitLab) и к GPU/CPU.
- Разработана инфраструктура для приёма задач (Gitlab, CI/CD) и согласована с лабораторными работами и настроена на всех студентов образовательной программы;
- Техническое обеспечение Python, JavaScript, C++, Java.
- Использование открытых датасетов и библиотек.

**Цели, задачи и ожидаемые результаты**

Цели:

- Закрепление теоретических знаний на практике.
- Развитие навыков определения алгоритма решения конкретной задачи машинного обучения.
- Подготовка к решению реальных задач в индустрии.

Задачи:

1. Обеспечить студентов структурированными лабораторными работами.
2. Предоставить доступ к необходимому программному обеспечению.
3. Организовать проверку и обратную связь по выполненным работам.

Ожидаемые результаты:

- Умение применять машинное обучение на практике.
- Навыки работы с языками программирования: Python, JavaScript, C++, Java.
- Опыт решения практических задач машинного обучения в различных отраслях.

### **Порядок реализации.**

**Задача №1** Обеспечение студентов структурированными лабораторными работами.

1) Определение тем:

- Линейная регрессия.
- Логистическая регрессия и классификация
- Кластеризация (K-means, DBSCAN).
- Нейронные сети
- Понижение размерности
- Обучение с подкреплением.
- Обработка естественного языка.
- Ансамбли

2) Разработка заданий:

- Пошаговые инструкции.
- Алгоритмы.
- Контрольные вопросы.

### ***Разработка заданий для лабораторной работы «Кластеризация (K-means, DBSCAN).***

Цель: Изучить методы кластеризации на практике, сравнение алгоритмов K-means, DBSCAN, оценка качества кластеризации.

1. Пошаговые инструкции

Шаг1 Подготовка среды данных (установка библиотек, импорт библиотек, загрузка данных, нормализация данных)

Шаг 2 Кластеризация K-means

Шаг 3 Кластеризация **DBSCAN**

Шаг 4 Сравнение методов

#### **Алгоритм метода K-means.**

1. Выбирается число кластеров K
2. Случайно инициализируются центроиды (центры кластеров)
3. На каждом шаге:
  - каждая точка приписывается к ближайшему центроиду
  - центроиды пересчитываются как среднее точек в кластере
4. Процесс повторяется, пока центроиды не стабилизируются

#### **Алгоритм метода DBSCAN**

1. Инициализация. Выбираются параметры  $\epsilon$  и  $\text{min\_samples}$  и все точки помечаются как непосещенные

2. Поиск соседей. Для каждой точки  $p$  в данных:

- 1) если  $p$  уже посещена, то пропускаем
- 2) если  $p$  уже посещена, то помечаем как посещенную
- 3) находим все точки в  $\epsilon$ -окрестности  $p$  (включая саму  $p$ )

-если число соседей больше либо равно  $\text{min\_samples}$ , то  $p$ - core point, создаем

новый

кластер

- иначе  $p$  – шум (но позже может стать *border point*)
- 3. Если  $p$  — *core point*:
  1. Все точки в её  $\epsilon$ -окрестности добавляются в кластер.
  2. Для каждой новой точки  $q$  в кластере:
    - Если  $q$  не посещена  $\rightarrow$  рекурсивно проверяем её соседей.
    - Если  $q$  — *core point*  $\rightarrow$  добавляем её соседей в кластер.
- 4. Завершение. Все точки, не вошедшие в кластеры, считаются шумом

### Контрольные вопросы

1. Какие типы задач решает кластеризация? Приведите примеры.
2. Какие основные подходы к кластеризации вы знаете?
3. Какие метрики используются для оценки качества кластеризации?
4. Как выбрать оптимальное число кластеров в K-means? Объясните \*метод локтя\*.
5. Какие недостатки у K-means? В каких случаях он плохо работает?
6. Какие параметры есть у DBSCAN? Как они влияют на результат?
7. Почему DBSCAN лучше K-means для кластеров произвольной формы?

### Критерии оценки

*Отлично*: Полное выполнение всех шагов, проведение сравнительного анализа и ответы на все вопросы.

*Хорошо*: Корректная работа алгоритма, но при ответах на вопросы получены незначительные замечания

*Удовлетворительно*: Алгоритм работает с ошибками, устраняемыми в процессе приема задачи, при ответах на вопросы получен ряд замечаний.

*Неудовлетворительно*: Невыполнение ключевых этапов.

#### 3) Подготовка датасетов:

- Использование открытых данных Kaggle для решения индивидуальных задач.
- Генерация синтетических индивидуальных задач при необходимости.

Задача №2: доступ к необходимым вычислительным ресурсам (в п.4.3 РПД)

Задача №3: Организация проверки и обратной связи по выполненным работам.

### Порядок проверки корректности

Чек-лист для проверки лабораторных работ:

1. Выполнение заданий:
  - Код запускается без ошибок.
  - Достигнуто целевое качество модели
2. Качество кода:
  - Соблюдение PEP-8.
  - Наличие комментариев.
3. Отчет:
  - инструкция по работе с гитом с подробным описанием именования методов и коммитов;
  - описание хода работы.
4. Анализ результатов. Своевременность:
  - Работа сдана в установленный срок.

Критерии оценки:

*Отлично*: Полное выполнение всех заданий, качественный код и отчет.

*Хорошо*: Незначительные недочеты в коде или отчете.

*Удовлетворительно:* Выполнены базовые задания, но с ошибками.

*Неудовлетворительно:* Критические ошибки или невыполнение работы.

#### **4.5. Методические указания по организации проектной деятельности**

##### **Условия применения:**

- Курс рассчитан на студентов 3-го года обучения,
- Общее время на проект – не более 16 часов на каждого студента.
- Имеется доступ к кейсам промышленных партнеров;

##### **Цели, задачи и ожидаемые результаты**

Цели организации вычислительной инфраструктуры:

- дать начальное представление о реальных задачах, решаемых с помощью нейронных сетей и возникающих проблемах.

Задачи преподавателя:

- сбор кейсов промышленных партнеров;
- сбор кейсов преподавателей практиков и лабораторий в вузе;
- формирование ТЗ на проект на основе кейсов;
- разработка системы учёта результатов проекта в итоговой зачетной оценке

Ожидаемые результаты студентов:

- начальное представление о реальных задачах, решаемых с помощью машинного обучения и возникающих проблемах, готовое решение с обоснованием выбора метода, презентация результатов.

##### **Порядок реализации:**

###### **1. Пример проекта:**

**Тема:** Индивидуальная оценка кредитоспособности клиента на основе поведенческих данных.

**Задание:**

- Разработать ML-модель, оценивающую вероятность дефолта по нестандартным поведенческим признакам.
- **Критерии оценки:**
- Эффективность решения.
- Качество кода и отчетности.

##### **Порядок проверки корректности:**

Чек-лист:

- Набор кейсов промышленных партнеров – 4 шт
- Наличие репозитория с кодом.
- Защита проекта (презентация, ответы на вопросы).

#### **5. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)**

##### **5.1 Основная литература:**

1. Остроух, А. В. Системы искусственного интеллекта : монография / А. В. Остроух, Н. Е. Суркова. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 228 с. — ISBN 978-5-8114-8519-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/176662> (дата обращения: 18.08.2021). — Режим доступа: для авториз. пользователей.

2. Буховец, А. Г. Алгоритмы вычислительной статистики в системе R : учебное пособие / А. Г. Буховец, П. В. Москалев. — 2-е изд., перераб. и доп. — Санкт-Петербург : Лань, 2022. — 160 с. — ISBN 978-5-8114-1802-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/212195>. — Режим доступа: для авториз. пользователей.

3. Бессмертный, И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2024. — 243 с. — (Высшее образование). — ISBN 978-5-534-01042-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/537001> (дата обращения: 30.05.2024).

4. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — Москва : Издательство Юрайт, 2024. — 85 с. — (Высшее образование). — ISBN 978-5-534-15561-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/544780> (дата обращения: 30.05.2024).

### 5.2 Дополнительная литература

1. Kadurin, Artur, et al. "The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology." *Oncotarget* 8.7 (2016): 10883.
2. Kadurin, Artur, et al. "druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico." *Molecular pharmaceutics* 14.9 (2017): 3098-3104.
3. Polykovskiy, Daniil, et al. "Molecular sets (MOSES): a benchmarking platform for molecular generation models." *Frontiers in pharmacology* 11 (2020): 565644.
4. Khrabrov, Kuzma, et al. " $\nabla^2$  DFT: A Universal Quantum Chemistry Dataset of Drug-Like Molecules and a Benchmark for Neural Network Potentials." *Advances in Neural Information Processing Systems* 37 (2024): 36869-36889.
5. Polykovskiy, Daniil, et al. "Entangled conditional adversarial autoencoder for de novo drug discovery." *Molecular pharmaceutics* 15.10 (2018): 4398-4405.
6. Николенко, Сергей, Кадури, Артур и Архангельская Екатерина. Глубокое обучение. "Издательский дом Питер", 2017.
7. Kirillova, E.; Kovalenko, A.; Urtenov, M. Study of the current–voltage characteristics of membrane systems using neural networks. *AppliedMath* 2025, 5, 10. <https://doi.org/10.3390/appliedmath5010010>,
8. Anand Subramoney, et al. Efficient recurrent architectures through activity sparsity and sparse back-propagation through time / Published as a conference paper at ICLR 2023, Конференции А\* <https://openreview.net/forum?id=ElUrNM9U8c#discussion>
9. Veronika Ganeeva, , et al. Lost in Translation: Chemical Language Models and the Misunderstanding of Molecule Structures / Findings of the Association for Computational Linguistics: EMNLP 2024, pages 12994–13013, Конференции А\* <https://aclanthology.org/2024.findings-emnlp.760/>
10. Andrey Sakhovskiy, et al. Biomedical Entity Representation with Graph-Augmented Multi-Objective Transformer / Findings of the Association for Computational Linguistics: NAACL 2024, pages 4626–4643, Конференции А\* <https://aclanthology.org/2024.findings-naacl.288/>

Для освоения дисциплины инвалидами и лицами с ограниченными возможностями здоровья имеются издания в электронном виде в электронно-библиотечных системах «Лань».

### 5.3. Периодическая литература

1. Базы данных компании «Ист Вью» <http://dlib.eastview.com>
2. Электронная библиотека GREBENNIKON.RU <https://grebennikon.ru/>

#### 5.4. Интернет-ресурсы, в том числе современные профессиональные базы данных и информационные справочные системы

*Электронно-библиотечные системы (ЭБС):*

1. ЭБС «ЮРАЙТ» <https://urait.ru/>
2. ЭБС «УНИВЕРСИТЕТСКАЯ БИБЛИОТЕКА ОНЛАЙН»

[www.biblioclub.ru](http://www.biblioclub.ru)

3. ЭБС «BOOK.ru» <https://www.book.ru>
4. ЭБС «ZNANIUM.COM» [www.znanium.com](http://www.znanium.com)
5. ЭБС «ЛАНЬ» <https://e.lanbook.com>

*Профессиональные базы данных:*

1. Web of Science (WoS) <http://webofscience.com/>
2. Scopus <http://www.scopus.com/>
3. ScienceDirect [www.sciencedirect.com](http://www.sciencedirect.com)
4. Журналы издательства Wiley <https://onlinelibrary.wiley.com/>
5. Научная электронная библиотека (НЭБ) <http://www.elibrary.ru/>
6. Полнотекстовые архивы ведущих западных научных журналов на Российской платформе научных журналов НЭИКОН <http://archive.neicon.ru>
7. Национальная электронная библиотека (доступ к Электронной библиотеке диссертаций Российской государственной библиотеки (РГБ) <https://rusneb.ru/>
8. Президентская библиотека им. Б.Н. Ельцина <https://www.prilib.ru/>
9. Электронная коллекция Оксфордского Российского Фонда <https://ebookcentral.proquest.com/lib/kubanstate/home.action>
10. Springer Journals <https://link.springer.com/>
11. Nature Journals <https://www.nature.com/siteindex/index.html>
12. Springer Nature Protocols and Methods <https://experiments.springernature.com/sources/springer-protocols>
13. Springer Materials <http://materials.springer.com/>
14. zbMath <https://zbmath.org/>
15. Nano Database <https://nano.nature.com/>
16. Springer eBooks: <https://link.springer.com/>
17. "Лекториум ТВ" <http://www.lektorium.tv/>
18. Университетская информационная система РОССИЯ <http://uisrussia.msu.ru>

*Информационные справочные системы:*

1. Консультант Плюс - справочная правовая система (доступ по локальной сети с компьютеров библиотеки)

*Ресурсы свободного доступа:*

1. Американская патентная база данных <http://www.uspto.gov/patft/>
2. Полные тексты канадских диссертаций <http://www.nlc-bnc.ca/thesescanada/>
3. КиберЛенинка (<http://cyberleninka.ru/>);
4. Министерство науки и высшего образования Российской Федерации <https://www.minobrnauki.gov.ru/>;
5. Федеральный портал "Российское образование" <http://www.edu.ru/>;
6. Информационная система "Единое окно доступа к образовательным ресурсам" <http://window.edu.ru/>;
7. Единая коллекция цифровых образовательных ресурсов <http://school-collection.edu.ru/> .
8. Федеральный центр информационно-образовательных ресурсов (<http://fcior.edu.ru/>);

9. Проект Государственного института русского языка имени А.С. Пушкина "Образование на русском" <https://pushkininstitute.ru/>;
10. Справочно-информационный портал "Русский язык" <http://gramota.ru/>;
11. Служба тематических толковых словарей <http://www.glossary.ru/>;
12. Словари и энциклопедии <http://dic.academic.ru/>;
13. Образовательный портал "Учеба" <http://www.ucheba.com/>;
14. Законопроект "Об образовании в Российской Федерации". Вопросы и ответы [http://xn--273--84d1f.xn--p1ai/voprosy\\_i\\_otvety](http://xn--273--84d1f.xn--p1ai/voprosy_i_otvety)

*Собственные электронные образовательные и информационные ресурсы КубГУ*

1. Электронный каталог Научной библиотеки КубГУ <http://megapro.kubsu.ru/MegaPro/Web>
2. Электронная библиотека трудов ученых КубГУ <http://megapro.kubsu.ru/MegaPro/UserEntry?Action=ToDb&idb=6>
3. Среда модульного динамического обучения <http://moodle.kubsu.ru>
4. База учебных планов, учебно-методических комплексов, публикаций и конференций <http://infoneeds.kubsu.ru/>
5. Библиотека информационных ресурсов кафедры информационных образовательных технологий <http://mschool.kubsu.ru;>
6. Электронный архив документов КубГУ <http://docspace.kubsu.ru/>
7. Электронные образовательные ресурсы кафедры информационных систем и технологий в образовании КубГУ и научно-методического журнала "ШКОЛЬНЫЕ ГОДЫ" <http://icdau.kubsu.ru/>

#### **6. Методические указания для обучающихся по освоению дисциплины**

По курсу предусмотрено проведение лекционных занятий, на которых дается основной теоретический материал, лабораторных занятий, позволяющих студентам в полной мере ознакомиться с понятиями дисциплины и освоиться в решении практических задач. Важнейшим этапом курса является самостоятельная работа. Целью самостоятельной работы бакалавра является углубление знаний, полученных в результате аудиторных занятий. Вырабатываются навыки самостоятельной работы. Закрепляются опыт и знания, полученные во время лабораторных занятий. Самостоятельная работа студентов в ходе изучения дисциплины состоит в выполнении индивидуальных заданий (отчет в электронной форме), подготовки теоретического материала к лабораторным занятиям, на основе конспектов лекций и учебной литературы, согласно календарному плану и подготовки теоретического материала к тестовому опросу, зачету и экзамену, согласно вопросам к экзамену.

Указания по оформлению работ:

- работа на лабораторных занятиях выполняется за компьютером в компьютерном классе;
- оформление индивидуальных заданий (отчетов) желательно в виде файлов в формате word.

Итогом самостоятельной работы студента является отчет, в котором на оригинальной таблице исходных данных студент самостоятельно проводит анализ данных всеми изученными в рамках курса методами и, представляет его на проверку в электронном виде.

Проверка индивидуальных заданий по темам, разобранным на лабораторных занятиях, осуществляется через неделю на текущем лабораторном занятии, либо в течение недели после этого занятия на консультации.

Для разъяснения непонятных вопросов лектором и ассистентом еженедельно проводятся консультации, о времени которых группы извещаются заранее.

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья большое значение имеет индивидуальная учебная работа (консультации) – дополнительное разъяснение учебного материала.

Индивидуальные консультации по предмету являются важным фактором, способствующим индивидуализации обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или лицом с ограниченными возможностями здоровья.

Подход, определяющий установление соответствия кейсов ИП и УГТ (5-7), позволяет четко соотносить этапы развития технологии с вовлеченностью партнера и снижать риски при переходе от лабораторных испытаний к промышленному внедрению.

### **Примеры кейсов от ПАО «Сбербанк»**

#### **Кейс №1. Индивидуальная оценка кредитоспособности клиента на основе поведенческих данных**

##### **Описание:**

Современный кредитный скоринг выходит за рамки финансовых данных. Необходимо исследовать, как поведенческие и цифровые следы (частота входа в мобильный банк, способы оплаты, география, время отклика) влияют на персональную оценку риска.

##### **Цель:**

Разработать ML-модель, оценивающую вероятность дефолта по нестандартным поведенческим признакам (возможно — с explainable AI).

##### **Ожидаемый результат:**

Прототип скоринговой модели, которая, помимо стандартных данных, учитывает цифровой профиль клиента и объясняет решения (SHAP, LIME и др.).

#### **Кейс №2. Генерация сценариев фишинговых писем для обучения сотрудников**

##### **Описание:**

Банк проводит киберучения, включая рассылку тестовых фишинговых писем сотрудникам для повышения их устойчивости к социальным атакам. Проект предполагает использование генеративной модели для создания реалистичных фишинговых писем различных типов (поддельные счета, HR-запросы, ИТ-поддержка).

##### **Цель:**

Создать генератор, способный на основе заданных параметров (тема, стиль, уровень угрозы) создавать тексты фишинга для тренировок.

##### **Ожидаемый результат:**

Набор разнообразных примеров фишинга и оценка их эффективности по реакции сотрудников, а также классификация моделей угроз.

### **Кейсы от «АВАЛАБ» (лаборатория искусственного интеллекта Сбербанка)**

#### **Кейс №1. Модель прогнозирования сроков сдачи объектов на основе текстовых и визуальных данных**

##### **Описание:**

Девелоперская компания ведёт аналитический архив по срокам строительства. С помощью мультимодальных моделей (текстовые отчёты + фото стройки) можно прогнозировать вероятность отклонения от графика сдачи.

##### **Цель:**

Разработать модель, которая по текущему статусу объекта (фото, отчёт СМР) оценивает риски задержек.

### **Ожидаемый результат:**

Прототип, который показывает вероятность отклонений и даёт текстовые пояснения (основанные на распознанных признаках — «не завершены фасадные работы», «монтаж инженерии не начат»).

### **Кейс №2. Анализ обращений клиентов и CRM-переписки**

#### **Описание:**

В службе клиентского сервиса застройщика ежедневно обрабатываются десятки обращений (e-mail, звонки, мессенджеры). Требуется реализовать систему семантического анализа и классификации NLU: выявлять суть обращений, уровень удовлетворенности, отслеживать повторяющиеся запросы.

#### **Цель:**

Автоматизировать первичный разбор и маршрутизацию запросов по тематике (сдача объекта, отделка, документы, жалоба и т.д.).

#### **Ожидаемый результат:**

Прототип, который выделяет суть обращений и формирует дашборд по текущим «болям» клиентов.

Для студентов с ограниченными возможностями здоровья предусмотрены дополнительные индивидуальные консультации, на которых преподаватель подробно разъясняет сложные аспекты дисциплины, помогает адаптировать практические задания и обеспечивает специальные условия для освоения методов работы с системами искусственного интеллекта. Индивидуальный подход позволяет таким студентам полноценно участвовать в учебном процессе и достигать требуемых результатов обучения.

## **7. Перечень информационных технологий, используемых при осуществлении образовательного процесса**

### **7.1 Перечень информационно-коммуникационных технологий**

1. Облачные платформы и сервисы  
cloud.ru, YandexCloud, AWS/GCP/Azure – облачные вычисления
2. Системы управления версиями и коллаборации  
Git/GitHub/GitLab – контроль версий кода и совместная разработка
3. Система управления обучением Moodle – сдача работ
4. Проверка индивидуальных заданий и консультирование посредством электронной почты.
5. Использование электронных презентаций при проведении лекционных и лабораторных занятий.
6. Использование математических пакетов при выполнении индивидуальных заданий.
7. Справочно-правовая система «Консультант Плюс» <http://www.consultant.ru>
8. Портал открытых данных Российской Федерации <https://data.gov.ru>
9. База открытых данных Министерства труда и социальной защиты РФ <https://rosmintrud.ru/opendata>
10. База данных Научной электронной библиотеки eLIBRARY.RU <https://elibrary.ru/>
11. База данных Всероссийского института научной и технической информации (ВИНИТИ) РАН <http://www2.viniti.ru/>
12. Базы данных и аналитические публикации «Университетская информационная система РОССИЯ» <https://uisrussia.msu.ru/>
13. Полная математическая база данных zbMATH <https://zbmath.org/>

14. [www.statlab.kubsu.ru](http://www.statlab.kubsu.ru)
15. <http://www.statsoft.ru/home/textbook/default.htm>
16. <http://statsoft.ru/solutions/>

## 7.2 Перечень необходимого программного обеспечения.

1. Операционная система MS Windows.
2. Интегрированное офисное приложение MS Office.
3. Программное обеспечение для организации управляемого коллективного и безопасного доступа в Интернет.
4. Языки программирования: Python
5. Google Colab – облачная среда для выполнения кода на Python с GPU/TPU

## 8. Материально-техническое обеспечение по дисциплине (модулю)

Виртуальные машины, кластер Managed Kubernetes и ресурсы GPU в облаке предоставляется индустриальным партнером ПАО «Сбербанк»:

№	Продукт	Параметры продукта	Кол-во	Кол-во конфигураций	Ед. изм.
1	Виртуальная машина	Виртуальная машина 10% vCPU 2 vCPU 4 RAM	1	60	Шт
		ОС Ubuntu 22.04	1		Шт
		Системный диск SSD	1		Шт
			10		Гб
		Аренда публичного IP	1		Шт
2	Виртуальная машина с GPU	Виртуальная машина с GPU NVIDIA® Tesla® V100 2 GPU 8 vCPU 128 Гб RAM	1	1	Шт
		ОС Ubuntu_24.04	1		Шт
		Системный диск SSD	1		Шт
			2000		Гб
		Диск SSD	1		Шт
			4096		Гб
		Диск SSD	1		Шт
	4096	Гб			
3	K8S	Аренда публичного IP	1	1	Шт
		Master node 8 vCPU 16 RAM	1		Шт
		Worker node 10% доля 4 vCPU 32 RAM	5		Шт
		Worker node SSD-NVME	64		Гб
	Аренда публичного IP	1	Шт		
4	ML Inference Instance Type GPU	Время работы в месяц	40	1	Ч
		Инстанс 8 x NVIDIA® H100 NVLink PCIe 160 vCPU 1520 GB RAM	1		Шт
		Количество запросов к ML-моделям	1		Млн. Шт

		Кэш ML-моделей	160	Гб
5	LLM	Токены GigaChat 2 Max	50	Млн. Шт
		Токены Embeddings	400	Млн. Шт

Дополнительные облачные ресурсы предоставляются технологическим партнером Yandex Cloud.

№	Вид работ	Материально-техническое обеспечение дисциплины (модуля) и оснащенность
	Лекционные занятия	Лекционная аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) и соответствующим программным обеспечением (ПО), доска Ауд. 129, 131, 301б, 305, 307
	Лабораторные занятия	Лаборатория, укомплектованная техническими средствами обучения – компьютерами с соответствующим программным обеспечением, маркерная доска. Ауд. 101, 106, 106а
	Групповые (индивидуальные) консультации	Аудитория, укомплектованная маркерной доской и оснащенная компьютером. Ауд. 129
	Текущий контроль, промежуточная аттестация	Аудитория, укомплектованная маркерной доской и оснащенная компьютером. Ауд. 129
	Самостоятельная работа	Кабинет для самостоятельной работы, оснащенный компьютерной техникой с возможностью подключения к сети «Интернет», программой экранного увеличения и обеспеченный доступом в электронную информационно-образовательную среду университета. 102-А и читальный зал

Примечание: Конкретизация аудиторий и их оснащение определяется ОПОП.