



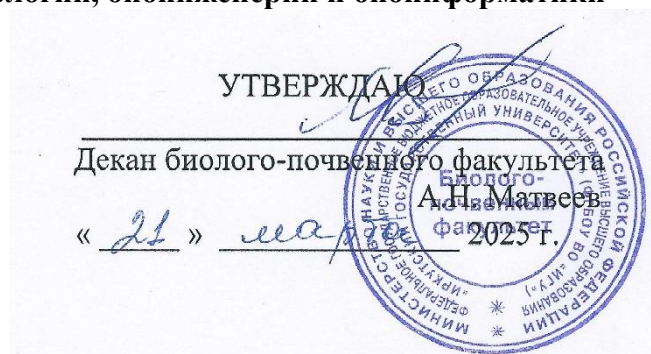
МИНОБРНАУКИ РОССИИ

федеральное государственное бюджетное образовательное учреждение
высшего образования

«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ФГБОУ ВО «ИГУ»

Кафедра физико-химической биологии, биоинженерии и биоинформатики



Рабочая программа дисциплины

Наименование дисциплины: Б1.О.37 **«ИСКУСТВЕННЫЙ ИНТЕЛЛЕКТ И
МАШИННОЕ ОБУЧЕНИЕ»**

Направление подготовки: 06.05.01 «Биоинженерия и биоинформатика»

Направленность (профиль) подготовки: «Биоинженерия и биоинформатика»

Квалификация выпускника: биоинженер и биоинформатик

Форма обучения: очная с элементами электронного обучения и дистанционных образовательных технологий

Согласовано с УМК биолого-почвенного
факультета

Протокол № 5 от 21 марта 2025 г.

Председатель А.Н. Матвеев

Рекомендовано кафедрой физико-химической
биологии, биоинженерии и биоинформатики

Протокол № 12 от 19 марта 2025 г.

Зав. кафедрой В.П. Саловарова

Иркутск 2025 г.

Содержание

	стр.
I. Цель и задачи дисциплины	3
II. Место дисциплины в структуре ОПОП	3
III. Требования к результатам освоения дисциплины	3
IV. Содержание и структура дисциплины	7
4.1 Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов	
4.2 План внеаудиторной самостоятельной работы обучающихся по дисциплине	
4.3 Содержание учебного материала	
4.3.1 Перечень семинарских, практических занятий и лабораторных работ	
4.3.2. Перечень тем (вопросов), выносимых на самостоятельное изучение в рамках самостоятельной работы студентов	
4.4. Методические указания по организации самостоятельной работы студентов	
4.5. Примерная тематика курсовых работ (проектов)	
V. Учебно-методическое и информационное обеспечение дисциплины	14
а) перечень литературы	
б) периодические издания	
в) список авторских методических разработок	
г) базы данных, поисково-справочные и информационные системы.....	
VI. Материально-техническое обеспечение дисциплины	17
6.1. Учебно-лабораторное оборудование	
6.2. Программное обеспечение	
6.3. Технические и электронные средства обучения	
VII. Образовательные технологии	20
VIII. Оценочные материалы для текущего контроля и промежуточной аттестации	21

I. Цель и задачи дисциплины:

Цель: является формирование системного представления, первичных знаний, умений и навыков по основам знаний и области машинного обучения как основному направлению построения систем; искусственного интеллекта. Дать представление практические знания в области использования методов машинного обучения и искусственного интеллекта в области анализ бельгийских данных и предсказания поведения биосистем.

Задачи:

- сформировать способность использовать современные информационные технологии и программные средства, при решении задач профессиональной деятельности с использованием интеллектуальных систем.
- сформировать способность разрабатывать модели биологички процессов и компонентов интеллектуальных систем, включая модели баз знаний;
- изучить спектр алгоритмов и методов машинного обучения и искусственного интеллекта их применением при решении практических задач по моделированию биосистем с помощью библиотек для различных языков программирования и онлайн ресурсов.
- сформировать способность проводить концептуальное, функциональное и логическое проектирование и разрабатывать архитектуру, прототип, дизайн интеллектуальных систем различного масштаба и сложности;
- сформировать способность выполнять работы по разработке, модификации, тестированию, развертыванию, эксплуатации и сопровождению интеллектуальных систем при исследовании различных процессов в области биологии и экологии.
- сформировать способности производить обработку запросов на анализ биологических данных и биосистем с помощью методов машинного обучения и искусственного интеллекта.

II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

2.1. Учебная дисциплина Б1.О.37 **«Искусственный интеллект и машинное обучение»** относится к обязательной части программы, изучается на 3 курсе в 5 и 6 семестрах.

2.2. Для изучения данной учебной дисциплины необходимы знания, умения и навыки, формируемые предшествующими дисциплинами: «Основы программирования», «Математика», «Физика», «Информатика», «Иностранный язык», «Специальные главы математики», «Математический анализ».

2.3. Перечень последующих учебных дисциплин, для которых необходимы знания, умения и навыки, формируемые данной учебной дисциплиной: «Структурно-функциональная биоинформатика», «Молекулярная филогенетика», «Геномный и метагеномный анализ», «ДНК метабаркодинг», выполнение ВКР.

III. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Процесс освоения дисциплины направлен на формирование компетенций в соответствии с ФГОС ВО и ОП ВО по данному направлению подготовки 06.05.01 «Биоинженерия и биоинформатика»:

ОПК-2: Способен использовать специализированные знания фундаментальных разделов математики, физики, химии и биологии для проведения исследований в области биоинженерии, биоинформатики и смежных дисциплин (модулей);

ОПК-3: Способен проводить экспериментальную работу с организмами и клетками, использовать физико-химические методы исследования макромолекул, математические методы обработки результатов биологических исследований.

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

Компетенция	Индикаторы компетенций	Результаты обучения
<i>ОПК-2</i> Способен использовать специализированные знания фундаментальных разделов математики, физики, химии и биологии для проведения исследований в области биоинженерии, биоинформатики и смежных дисциплин (модулей);	<i>ИДК ОПК-2.1</i> Демонстрирует специализированные знания в области фундаментальных разделов математики, физики, химии, биологии и перспективы междисциплинарных исследований	Знать: литературу по теме, владеть наивками, анализа информации сети «интернет» для поиска и освоения новых методов анализа данных, информационных технологий и алгоритмов машинного обучения. Уметь: выбирать оптимальные методы, алгоритмы и программы для решения задач в области анализа биологической информации в геномики, эволюционной биологии и экологии с помощью машинного обучения. Владеть: методами построение сложных алгоритмов на основе технологий машинного обучения.
	<i>ИДК ОПК-2.2</i> Умеет использовать навыки проведения исследований в области биоинженерии, биоинформатики с учетом специализированных фундаментальных знаний	Знать: классификацию алгоритмов, основные типы алгоритмов, синтаксис в области бутсрепа, алгоритмов на основе показателей правдоподобия и машинного обучения. Уметь: анализировать входные и выходные данные разрабатываемого алгоритма, производить отладку и тестирование разработанных алгоритмов для анализа данных в эволюционной биологии, геномики и экологии с применением машинного обучения. Владеть: навыками анализа сложных данных в различных отраслях биологии и биоинформатики с применением машинного обучения.
	<i>ИДК ОПК-2.3</i> Владеет методами химии, физики и математического моделирования для проведения исследований в области биоинженерии, биоинформатики	Знать: классификацию основных типов алгоритмов анализа сложных данных и применять изученные алгоритмы для создания сложных конвейеров анализа данных. Уметь: осуществлять интерпретацию результатов математических расчетов с применением всех изученных типов алгоритмов машинного обучения. Владеть: методами анализа комплексных биологических данных в эволюционной биологии, геномики и экологии
<i>ОПК-3</i> Способен проводить экспериментальную работу с организмами и	<i>ИДК ОПК-3.1</i> Проводит экспериментальную работу с организмами и	Знать: основные математические понятия и методы, применимые для анализа биологички систем и биологических данных с применением алгоритмов с применением

клетками, использовать физико-химические методы исследования макромолекул, математические методы обработки результатов биологических исследований;	клетками с использованием физико-химических методов исследования макромолекул	машинного обучения. Уметь: адекватно выбрать математический метод и алгоритмы для описания биологической системы и биологического процесса в эволюционной биологии геномики, биоинформатики и экологии с применением машинного обучения. Владеть: основными принципами формализации сложных конвейеров анализа данных с применением всех рассмотренных алгоритмов машинного обучения.
	<i>ИДК ОПК-3.2</i> Демонстрирует практические навыки математических методов обработки результатов экспериментальных исследований	Знать: цель, основные задачи и области применения алгоритмов биоинформатики в рамках направления подготовки. Уметь: формализовать процесс обработки данных в геномики, эволюционной биологии, экологии и других биологических дисциплинах в виде конвейеров различных вычислительных алгоритмов машинного обучения. Владеть: методами применения разработанных алгоритмов и конвейеров анализа данных при исследовании биологических процессов и биосистем с применением машинного обучения.
	<i>ИДК ОПК-3.3</i> Владеет опытом применения методов для исследования макромолекул, обработки результатов биологических исследований, прогнозирования перспектив и социальных последствий своей профессиональной деятельности.	Знать: спектр алгоритмов машинного обучения, применяемых при анализе биомолекул. Уметь: выбирать спектр программ и алгоритмов, использующих машинное обучение для исследования биомоллекул для решения поставленных биологических задач. Владеть: методами реконструкции структур биомоллекул с применением методов машинного обучения.

IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Объем дисциплины составляет 7 зачетных единицы, 252 часов.

Из них реализуется с использованием электронного обучения и дистанционных образовательных технологий 28 часов.

Форма промежуточной аттестации: зачет.

4.1 Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов

№ п/п	Раздел дисциплины/тема	Семестр	Всего часов	Из них практическая подготовка обучающихся	Виды учебной работы, включая самостоятельную работу обучающихся , практическую подготовку и трудоемкость (в часах)				Форма текущего контроля успеваемости/ Форма промежуточной аттестации (по семестрам)
					Контактная работа преподавателя с обучающимися			Самостоятельн ая работа	
					Лекция	Семинар/ Практическое, лабораторное занятие/	Консультация		
1	2	3	4	5	6	7	8	9	10
1	Тема 1.Введение, что такое машинное обучение.	5	12	4	2	4		6	Опрос КСР
2	Тема 2. Обучение с учителем. Ленная регрессия.	5	10	4	2	4		4	Опрос КСР
3	Тема 3. Обучение с учителем. Линейная классификация.	5	12	4	2	4		6	Опрос КСР
4	Тема 4. Обучение с учителем. Метод k ближайших соседей.	5	10	4	2	4		4	Опрос КСР
5	Тема 5. Обучение с учителем. Деревья решений.	5	12	4	2	4		6	Опрос КСР
6	Тема 6. Обучение с учителем. Ансамбли	5	10	4	2	4		4	Опрос

	деревьев решений.								КСР
7	Тема 7. Обучение с учителем. Метод опорных векторов.	5	10	4	2	4		4	Опрос КСР
8	Тема 8. Обучение с учителем. Наивный Байесовский классификатор	5	10	4	2	4		4	Опрос КСР
9	Тема 9. Обучение с учителем. Нейронные сети (глубокое обучение).	5	12	4	2	4		6	Опрос КСР
10	Тема 10. Обучение без учителя. Типы машинного обучения без учителя.	6	14	4	2	4		8	Опрос КСР
11	Тема 11. Обучение без учителя. Предварительная обработка и масштабирование данных.	6	14	4	2	4		8	Опрос КСР
12	Тема 12. Обучение без учителя. Анализ методом главных компонент.	6	14	4	2	4		8	Опрос КСР
13	Тема 13. Обучение без учителя. Кластеризация k средних.	6	14	4	2	4		8	Опрос КСР
14	Тема 14. Обучение без учителя. Иерархическая кластеризация.	6	14	4	2	4		8	Опрос КСР
15	Тема 15. Обучение без учителя. Оценки достоверности кластеризации.	6	14	4	2	4		8	Опрос КСР
16	Тема 16. Оценка и улучшения качества моделей машинного обучения	6	14	4	2	4		8	Опрос КСР
14	Тема 17. Объединение алгоритмов в цепочки и конвейеры.	6	17	4	2	4	1	10	Опрос КСР

4.2 План внеаудиторной самостоятельной работы обучающихся по дисциплине

Семестр	Название раздела, темы	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение самостоятельной работы
		Вид самостоятельной работы	Сроки выполнения	Трудоемкость (час.)		
5	Тема 1. Введение, что такое машинное обучение.	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Подготовка доклада	1	6	Доклад	Раздел 5 а-г
5	Тема 2. Обучение с учителем. Линейная регрессия.	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашних заданий по теме линейная регрессия в машинном обучении.	2	4	Письменная работа КСР	- « -
5	Тема 3. Обучение с учителем. Линейная классификация.	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашних заданий по теме линейная классификация в машинном обучении.	4	6	Письменная работа КСР	- « -
5	Тема 4. Обучение с учителем. Метод k ближайших соседей.	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашних заданий по использованию метода k ближайших соседей в машинном обучении..	5	4	Письменная работа КСР	- « -
5	Тема 5. Обучение с учителем. Деревья решений.	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашних заданий по использованию деревьев машинном обучении.	6	6	Письменная работа КСР	- « -
5	Тема 6. Обучение с учителем. Ансамбли деревьев решений.	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашних заданий по использованию ансамблей деревьев машинном обучении.	8	4	Письменная работа КСР	- « -
5	Тема 7. Обучение с учителем. Метод опорных векторов.	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашних заданий по использованию метода опорных векторов в машинном обучении.	9	4	Письменная работа КСР	- « -

Семестр	Название раздела, темы	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение самостоятельной работы
		Вид самостоятельной работы	Сроки выполнения	Трудоемкость (час.)		
5	Тема 8. Обучение с учителем. Наивный Байесовский классификатор	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашних заданий по использованию наивного Байесовского классификатора.	10	4	Письменная работа КСР	- « -
5	Тема 9. Обучение с учителем. Нейронные сети (глубокое обучение).	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашних заданий по построению нейронных сетей.	11	6	Письменная работа КСР	- « -
6	Тема 10. Обучение без учителя. Типы машинного обучения без учителя.	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам.	1	8	Доклад	- « -
6	Тема 11. Обучение без учителя. Предварительная обработка и масштабирование данных.	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашних заданий по использованию методов масштабирования данных.	2	8	Письменная работа КСР	- « -
6	Тема 12. Обучение без учителя. Анализ методом главных компонент.	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашних заданий с использованием метода главных компонент в машинном обучении.	4	8	Письменная работа КСР	- « -
6	Тема 13. Обучение без учителя. Кластеризация k средних.	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашнего задания по использованию алгоритма кластеризации k средних.	5	8	Письменная работа КСР	- « -

Семестр	Название раздела, темы	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение самостоятельной работы
		Вид самостоятельной работы	Сроки выполнения	Трудоемкость (час.)		
6	Тема 14. Обучение без учителя. Иерархическая кластеризация.	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашних заданий использованию методов иерархической кластеризации при анализе данных.	6	8	Письменная работа КСР	- « -
6	Тема 15. Обучение без учителя. Оценки достоверности кластеризации.	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашних задание по анализу достоверности результатов кластеризации.	8	8	Письменная работа КСР	- « -
6	Тема 16. Оценка и улучшения качества моделей машинного обучения	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашних задание по анализу качества моделей машинного обучения для анализа заданных массивов данных..	9	8	Письменная работа КСР	- « -
6	Тема 17. Объединение алгоритмов в цепочки и конвейеры.	1. Разбор темы лекции и практического занятия. 2. Подготовка по контрольным вопросам. 3. Решение домашних задание разработке простейшего конвейера по анализу данным методами машинного обучения.	11	10	Письменная работа КСР	- « -
Общий объем самостоятельной работы по дисциплине (час) – 110						
Из них объем самостоятельной работы с использованием электронного обучения и дистанционных образовательных технологий 110 часов.						

4.3 Содержание учебного материала

Тема 1. Введение, что такое машинное обучение. В рамках темы даются понятие «машинному обучению» как области искусственного интеллекта, которая изучает методы и алгоритмы, позволяющие вычислительным машинам при использовании специализированного программного обеспечения "обучаться" на данных и выполнять задачи построения моделей, для предсказания поведения биосистем, принятия решения и формулировки выводов.

Тема 2. Обучение с учителем. Ленная регрессия. В рамках темы рассматривается один из самых простых и популярных методов обучения с учителем, используемый для решения задач регрессии. Регрессия — это задача предсказания непрерывной целевой переменной на основе входных признаков. Рассматриваются примеры использования метода при анализе реальных биологических и экологических данных, при реализации с помощью библиотек языков программирования R или Python.

Тема 3. Обучение с учителем. Линейная классификация. В рамках темы рассматривается один из методов обучения с учителем, который используется для задач классификации. Изучается линейная классификация как конструируемая модель пытается разделить данные на классы с помощью линейной границы (прямой или плоскости) в двухмерном или многомерном пространстве. Рассматриваются примеры использования метода при анализе реальных биологических и экологических данных, при реализации с помощью библиотек языков программирования R или Python.

Тема 4. Обучение с учителем. Метод k ближайших соседей. В рамках темы рассматривается алгоритм обучения с учителем, используемый для задач классификации и регрессии. Изучается идея алгоритма которая заключается в том, чтобы предсказать значение целевой переменной для нового объекта на основе значений его ближайших соседей в обучающих данных. . Рассматриваются примеры использования метода при анализе реальных биологических и экологических данных, при реализации с помощью библиотек языков программирования R или Python.

Тема 5. Обучение с учителем. Деревья решений. В рамках темы рассматривается алгоритм обучения с учителем, используемый для задач классификации и регрессии. Изучается принцип работы алгоритма, основанный построении иерархической структуры, где каждый внутренний узел представляет собой условие на основе признаков, каждая ветвь — результат этого условия, а каждый лист — предсказанное значение (класс или численное значение). Рассматриваются примеры использования метода

при анализе реальных биологических и экологических данных, при реализации с помощью библиотек языков программирования R или Python.

Тема 6. Обучение с учителем. Ансамбли деревьев решений. В рамках темы рассматривается метод обучения с учителем деревья решений и ансамбли деревьев решений как методы, которые широко используются для задач классификации и регрессии. Рассматриваются ансамбли деревьев решений как алгоритм, объединяющий несколько деревьев для улучшения точности и устойчивости модели. Рассматриваются примеры использования метода при анализе реальных биологических и экологических данных, при реализации с помощью библиотек языков программирования R или Python.

Тема 7. Обучение с учителем. Метод опорных векторов. В рамках темы рассматривается метод обучения с учителем - метод опорных векторов (Support Vector Machines, SVM), используемый для задач классификации и регрессии. Изучается основная идея SVM, которая заключается в нахождении гиперплоскости, которая максимально разделяет классы в пространстве признаков. Рассматриваются примеры использования метода при анализе реальных биологических и экологических данных, при реализации с помощью библиотек языков программирования R или Python.

Тема 8. Обучение с учителем. Наивный Байесовский классификатор. В рамках темы рассматривается метод обучения с учителем - наивный Байесовский классификатор как алгоритм машинного обучения, основанный на теореме Байеса. Алгоритм широко используется для задач классификации, особенно в текстовой аналитике – в биологии для анализа текстов белгийских последовательностей (ДНК, РНК и белков). Рассматриваются примеры использования метода при анализе реальных биологических и экологических данных, при реализации с помощью библиотек языков программирования R или Python.

Тема 9. Обучение с учителем. Нейронные сети (глубокое обучение). В рамках темы рассматриваются нейронные сети как один из алгоритмов глубокого обучения. Изучаются вопросы нейронных сетей для решения сложных задач, таких как классификация, регрессия, обработка изображений, обработка биологических текстов. Нейронные сети рассматриваются как системы, состоящие из множества слоев, которые позволяют модели автоматически извлекать иерархические признаки из данных. Рассматриваются примеры использования метода при анализе реальных биологических и экологических данных, при реализации с помощью библиотек языков программирования R или Python.

Тема 10. Обучение без учителя. Типы машинного обучения без учителя. В рамках темы дается общая характеристика классам алгоритмов машинного

обучения машинного обучения, в котором модель обучается на данных без явных меток или целевых переменных. В отличие от обучения с учителем, где модель учится предсказывать выходные значения на основе входных данных, в обучении без учителя цель состоит в том, чтобы обнаружить скрытые структуры, закономерности или группировки в данных.

Тема 11. Обучение без учителя. Предварительная обработка и масштабирование данных. В рамках темы рассматриваются различные алгоритмы масштабирования данных, работа которых заключается в приведении признаков к единой шкале изменения переменных. Рассматриваются вопросы как масштабирование помогает ускорить сходимость алгоритмов, улучшить их производительность и избежать доминирования признаков с большими значениями. Рассматриваются примеры реализаций алгоритмов масштабирования с помощью библиотек языков программирования R или Python.

Тема 12. Обучение без учителя. Анализ методом главных компонент. В рамках темы рассматривается статистический метод — метод главных компонент (PCA, Principal Component Analysis), используемый для уменьшения размерности данных, который помогает выявить основные направления (компоненты), объясняющие наибольшую вариацию в данных. Изучаются основные этапы реализации метода для анализа биологических данных. Рассматриваются примеры реализаций метода главных компонент с помощью библиотек языков программирования R или Python.

Тема 13. Обучение без учителя. Кластеризация k средних. В рамках темы изучается кластеризация методом k-средних (k-means clustering) — это один алгоритмов обучения без учителя, используемый для группировки данных на основе их сходства. Изучается основная цель метода кластеризации k-средних — разделить данные на k кластеров, где каждый кластер представляет собой группу объектов, близких друг к другу в пространстве признаков. Рассматриваются примеры реализаций метода для анализа данных биологии и экологии с помощью библиотек языков программирования R или Python.

Тема 14. Обучение без учителя. Иерархическая кластеризация. В рамках темы рассматривается иерархическая кластеризация — это метод обучения без учителя, который строит иерархию кластеров, представляя данные в виде древовидной структуры (дендрограммы). Изучаются преимущества и недостатки метода при анализе различных типов данных. Рассматриваются примеры реализаций различных методов иерархической кластеризации для анализа данных биологии и экологии с помощью библиотек языков программирования R или Python.

Тема 15. Обучение без учителя. Оценки достоверности кластеризации. В рамках темы рассматриваются различные подходы и алгоритмы, которые применяются для оценки достоверности и качества кластеризации с использованием различных индексов. Рассматриваются примеры реализаций различных методов оценки качества кластеризации при анализа данных биологии и экологии с помощью библиотек языков программирования R или Python.

Тема 16. Оценка и улучшения качества моделей машинного обучения. В рамках темы рассматриваются различные методы и метрики (среднеквадратичная ошибка, средняя абсолютная ошибка, коэффициент детерминации, индекс силуэта, кросс-валидация) оценки качества моделей машинного обучения, которые напрямую влияют на эффективность их применения. Рассматриваются примеры реализаций различных метрик оценки качества моделей машинного обучения при анализа данных биологии и экологии с помощью библиотек языков программирования R или Python.

Тема 17. Объединение алгоритмов в цепочки и конвейеры. В рамках темы рассматривается комплексный подход к анализу в машинном обучении, который позволяет комбинировать несколько этапов обработки данных и моделирования в единый процесс. Это особенно полезно для автоматизации и упрощения работы с данными, а также для повышения воспроизводимости и эффективности моделей. Рассматриваются примеры реализаций конвейеров для при анализа данных биологии и экологии с помощью библиотек языков программирования R или Python.

4.3.1. Перечень семинарских, практических занятий и лабораторных работ

№ п/п	№ раздела и темы	Наименование семинаров, практических и лабораторных работ	Трудоемкость (час.)		Оценочные средства	Формируемые компетенции (индикаторы)*
			Всего часов	Из них практическая подготовка		
1	2	3	4	5	6	7
1	Тема 1	Решение задач по теме введение в языки программирования R и Python как инструменты реализации методов машинного обучения	4	4	Опрос КСР	ОПК-2 ИДК ОПК-2.1 ИДК ОПК-2.2 ИДК ОПК-2.3 ОПК-3 ИДК ОПК-3.1 ИДК ОПК-3.2 ИДК ОПК-3.3
2	Тема 2	Решение задач по теме логистическая регрессия	4	4	Опрос КСР	ОПК-2 ИДК ОПК-2.1 ИДК ОПК-2.2 ИДК ОПК-2.3

						ОПК-3 ИДК ОПК-3.1 ИДК ОПК-3.2 ИДК ОПК-3.3
3	Тема 3	Решение задач по теме линейная классификация.	4	4	Опрос КСР	ОПК-2 ИДК ОПК-2.1 ИДК ОПК-2.2 ИДК ОПК-2.3 ОПК-3 ИДК ОПК-3.1 ИДК ОПК-3.2 ИДК ОПК-3.3
4	Тема 4	Решение задач по теме метод k ближайших соседей.	4	4	Опрос КСР	ОПК-2 ИДК ОПК-2.1 ИДК ОПК-2.2 ИДК ОПК-2.3 ОПК-3 ИДК ОПК-3.1 ИДК ОПК-3.2 ИДК ОПК-3.3
5	Тема 5	Решение задач по теме анализ данных с помощью дерева решений.	4	4	Опрос КСР	ОПК-2 ИДК ОПК-2.1 ИДК ОПК-2.2 ИДК ОПК-2.3 ОПК-3 ИДК ОПК-3.1 ИДК ОПК-3.2 ИДК ОПК-3.3
6	Тема 6	Решение задач по теме анализ данных с помощью ансамблей деревьев решений	4	4	Опрос КСР	ОПК-2 ИДК ОПК-2.1 ИДК ОПК-2.2 ИДК ОПК-2.3 ОПК-3 ИДК ОПК-3.1 ИДК ОПК-3.2 ИДК ОПК-3.3
7	Тема 7	Решение задач по теме анализ данных методом опорных векторов.	4	4	Опрос КСР	ОПК-2 ИДК ОПК-2.1 ИДК ОПК-2.2 ИДК ОПК-2.3 ОПК-3 ИДК ОПК-3.1 ИДК ОПК-3.2 ИДК ОПК-3.3
8	Тема 8	Решение задач по теме наивный Байесовский классификатор	4	4	Опрос КСР	ОПК-2 ИДК ОПК-2.1 ИДК ОПК-2.2 ИДК ОПК-2.3 ОПК-3

						ИДК ОПК-3.1 ИДК ОПК-3.2 ИДК ОПК-3.3
9	Тема 9	Решение задач по теме конструирование нейронных сетей для анализа биологических данных	4	4	Опрос КСР	ОПК-2 ИДК ОПК-2.1 ИДК ОПК-2.2 ИДК ОПК-2.3 ОПК-3 ИДК ОПК-3.1 ИДК ОПК-3.2 ИДК ОПК-3.3
10	Тема 10	Решение задач по теме введение в языки программирования R и Python как инструменты реализации методов машинного обучения без учителя	4	4	Опрос КСР	ОПК-2 ИДК ОПК-2.1 ИДК ОПК-2.2 ИДК ОПК-2.3 ОПК-3 ИДК ОПК-3.1 ИДК ОПК-3.2 ИДК ОПК-3.3
11	Тема 11	Решение задач по теме масштабирование данных.	4	4	Опрос КСР	ОПК-2 ИДК ОПК-2.1 ИДК ОПК-2.2 ИДК ОПК-2.3 ОПК-3 ИДК ОПК-3.1 ИДК ОПК-3.2 ИДК ОПК-3.3
12	Тема 12	Решение задач по теме анализ методом главных компонент.	4	4	Опрос КСР	ОПК-2 ИДК ОПК-2.1 ИДК ОПК-2.2 ИДК ОПК-2.3 ОПК-3 ИДК ОПК-3.1 ИДК ОПК-3.2 ИДК ОПК-3.3
13	Тема 13	Решение задач по теме кластеризация k средних.	4	4	Опрос КСР	ОПК-2 ИДК ОПК-2.1 ИДК ОПК-2.2 ИДК ОПК-2.3 ОПК-3 ИДК ОПК-3.1 ИДК ОПК-3.2 ИДК ОПК-3.3
14	Тема 14	Решение задач по теме иерархическая кластеризация.	4	4	Опрос КСР	ОПК-2 ИДК ОПК-2.1 ИДК ОПК-2.2 ИДК ОПК-2.3 ОПК-3 ИДК ОПК-3.1

						<i>ИДК ОПК-3.2</i> <i>ИДК ОПК-3.3</i>
15	Тема 15	Решение задач по теме оценка достоверности кластеризации.	4	4	Опрос КСР	ОПК-2 <i>ИДК ОПК-2.1</i> <i>ИДК ОПК-2.2</i> <i>ИДК ОПК-2.3</i> ОПК-3 <i>ИДК ОПК-3.1</i> <i>ИДК ОПК-3.2</i> <i>ИДК ОПК-3.3</i>
16	Тема 16	Решение задач по теме оценка и улучшения качества моделей машинного обучения	4	4	Опрос КСР	ОПК-2 <i>ИДК ОПК-2.1</i> <i>ИДК ОПК-2.2</i> <i>ИДК ОПК-2.3</i> ОПК-3 <i>ИДК ОПК-3.1</i> <i>ИДК ОПК-3.2</i> <i>ИДК ОПК-3.3</i>
17	Тема 17	Решение задач по теме разработка цепочек и конвейеров для анализа данных методами машинного обучения.	4	4	Опрос КСР	ОПК-2 <i>ИДК ОПК-2.1</i> <i>ИДК ОПК-2.2</i> <i>ИДК ОПК-2.3</i> ОПК-3 <i>ИДК ОПК-3.1</i> <i>ИДК ОПК-3.2</i> <i>ИДК ОПК-3.3</i>

4.3.2. Перечень тем (вопросов), выносимых на самостоятельное изучение студентами в рамках самостоятельной работы (СРС)

№ п/п	Тема	Задание	Формируемая компетенция	ИДК
1.	Тема 9. Обучение с учителем. Нейронные сети (глубокое обучение).	Самостоятельное изучение темы – использование нейронных сетей для распознавания образов и идентификации биологических видов.	ОПК-2 ОПК-3	<i>ИДК ОПК-2.1</i> <i>ИДК ОПК-2.2</i> <i>ИДК ОПК-2.3</i> <i>ИДК ОПК-3.1</i> <i>ИДК ОПК-3.2</i> <i>ИДК ОПК-3.3</i>

4.4. Методические указания по организации самостоятельной работы студентов

Самостоятельная работа студентов является составной частью учебного процесса и имеет целью закрепление и углубление полученных знаний и навыков, поиск и приобретение новых знаний, а также выполнение учебных заданий, подготовку к предстоящим занятиям, и экзамену по предмету.

Для организации самостоятельной работы по дисциплине «Алгоритмы биоинформатики» используются следующие формы самостоятельной учебной работы:

- Работа по изучению темы с использованием материалов практического занятия.
- Подбор, изучение, анализ рекомендованной литературы.
- Изучения тем занятий, вынесенных на самостоятельное изучение, подготовка

устных докладов по темам.

- Самостоятельное изучение программного обеспечения для выполнения задач по использованию различных алгоритмов в анализе биологических данных.

- Самостоятельное решения домашних задач по анализу биологических данных на основе опыта, полученного на практических занятиях.

- Подготовка письменных отчетов по решению домашних задач.

- Подготовка к зачету.

Письменный отчет по решению домашних заданий – это отчет о выполнении домашнего задания по темам дисциплины, содержащий следующую информацию:

- Ф.И.О. номер группы студента;

- номер задания;

- формулировка задания;

- список программного обеспечения и интернет сервисов и баз данных, применяемых для решения задания с указанием параметров для запуска;

- описание результат решения задания с приведением таблиц и рисунков в соответствии с формулировкой задания.

Критерий оценки отчета по решению домашнего задания:

- Оценка «зачтено». Задание выполнено правильно и в полном объеме, все таблицы и графики согласно формулировке задания предоставлены в отчете.

- Оценка «не зачтено». Задание выполнено не правильно или не в полном объеме, вопрошается на переделку и доработку.

Устный доклад – это сообщение в течение 10-15 мин, в котором студент в лаконичной форме должен изложить материал по соответствующей теме, придерживаясь следующего плана: введение, основная часть, заключение. Доклад сопровождается презентацией, отражающей основные положения по соответствующей теме, включающей наглядные материалы (схемы, таблицы, фото и т.д.). По окончании доклада студенту задают вопросы, как преподаватель, так и студенты, на которые докладчик должен дать исчерпывающие ответы.

Критерии оценивания устного доклада:

- Оценка «отлично». В докладе полностью раскрыта тема, проанализировано современное состояние вопроса; студент свободно владеет материалом, излагает его логично, последовательно, лаконично, соблюдая основные правила культуры речи. Доклад сопровождается презентацией, которая отражает основные положения доклада, презентация составлена грамотно с соблюдением общих требований, правил шрифтового оформления, подачи графического материала, имеются ссылки на приведенные фото, рисунки, схемы и т.д., приводится список использованной литературы. При обсуждении доклада студент дает исчерпывающие, аргументированные, корректные ответы на вопросы.

- Оценка «хорошо». Тема раскрыта, приведено достаточное количество материала, но при этом материал в недостаточной степени проанализирован автором. Презентация не в полной степени соответствует общим требованиям. Ответы студента не на все вопросы являются исчерпывающими и аргументированными.

- Оценка «удовлетворительно». Тема раскрыта не полно, материал приведен как простая констатация фактов, не проанализирован, студент показывает поверхностные знания. Презентация частично соответствует установленным требованиям. При обсуждении доклада студент не всегда дает правильные, исчерпывающие ответы на задаваемые вопросы.

- Оценка «неудовлетворительно». Тема доклада не раскрыта, скудный объем приведенных материалов; презентация отсутствует. При обсуждении доклада студент не дает ответы или они не соответствуют заданным вопросам.

4.5. Примерная тематика курсовых работ (проектов): не предусмотрены учебным планом.

V. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

а) перечень литературы

1. Леск А. Введение в биоинформатику : пер. с англ. / А. М. Леск ; ред.: А. А. Миронов, В. К. Швядаса. - М. : Бином. Лаборатория знаний, 2009. - 318 с. - ISBN 978-5-94774-501-6 (8 экз.)
2. Приставка А. А. Большой практикум по биоинженерии и биоинформатике [Текст] : учеб.-метод. пособие : в 3 ч. / А. А. Приставка, В. П. Саловарова - Иркутск : Изд-во ИГУ, 2013. - Ч. 1 : Белки. - 2013. - 121 с. - ISBN 978-5-9624-0962-7 (69 экз.)
3. Белькова Н.Л. Большой практикум по биоинженерии и биоинформатике [Текст] : учеб.-метод. пособие : в 3 ч. / Н. Л. Белькова. - Иркутск : Изд-во ИГУ, 2013. - ISBN 978-5-9624-0956-6. Ч. 2 : Нуклеиновые кислоты. - 2014. - 155 с. - ISBN 978-5-9624-1184-2 (39 экз.)

б) дополнительная литература

1. Игнасимуту С. Основы биоинформатики / С. Игнасимуту ; пер. с англ. А. А. Чумичкин. - Ижевск : Регулярная и хаотическая динамика : Ин-т компьютер. исслед., 2007. - 316 с. - ISBN 978-5-93972-620-7 (1 экз.)
2. Каменская М.А. Информационная биология / М.А. Каменская. - М.: Академия, 2006. - 361 с. - ISBN 5-7695-2580-0 (8 экз.)
3. Компьютеры и суперкомпьютеры в биологии / Под ред. В.Д. Лахно, М.Н. Устинин. - Москва-Ижевск: Институт компьютерных исследований, 2002. - 528 с. - ISBN 5-93972-188-5 (2 экз.)
4. Математические методы для анализа последовательностей ДНК. / Под ред. М.С. Уотермена, перевод с англ. - М.: Мир, 1999. - 349 с. - ISBN 5030025200 (1 экз.)
5. Паун Г. ДНК-компьютер. Новая парадигма вычислений / Г. Паун, Г. Розенберг, А. Саломаа ; Пер. с англ. Д. С. Ананичева, И. С. Киселевой, О. Б. Финогеновой, ред. М. В. Волков. - М. : Мир, 2004. - 527 с. - ISBN 5-03-003480-3 (1 экз.)
6. Структура и функционирование белков: применение методов биоинформатики / пер. с англ.: В. Н. Новоселецкий, Е. Д. Балицкая, Т. В. Науменкова ; ред. В. Н. Новоселецкий. - М. : УРСС : Ленанд, 2014. - 414 с. - ISBN 978-5-9710-0842-2. - ISBN 978-5-453-00057-9 (1 экз.)
7. Шипунов А. Б., Балдин Е. М., Волкова П.А., и др. Наглядная статистика. Используем R! Издательство: ДМК Пресс, 2014 – 300 с. Книга доступна в свободном доступе по ссылке: <http://ashipunov.info/shipunov/school/books/rbook.pdf>

в) периодические издания

1. <https://www.matbio.org/> - сайт журнала «Математическая биология и биоинформатика». Содержит большое количество статей в pdf – формате.

2. <https://journal.r-project.org/> - сайт журнала по статистическим методам на R, «The R Journal».

г) базы данных, информационно-справочные и поисковые системы

1. <http://dmb.biophys.msu.ru> - Информационная система «Динамические модели в биологии», рассчитанная на широкий круг пользователей, включает в себя гипертекстовые документы и реляционные базы данных и обеспечивает унифицированный доступ к разнообразной информации по данной предметной области.
2. <http://www.jcbi.ru/> - сайт объединенного центра вычислительной биологии и биоинформатики
3. <http://mathmod.aspu.ru/> - Сайт совместной лаборатории Института математических проблем биологии Российской академии наук и Астраханского государственного университета
4. <http://www.exponenta.ru/> - образовательный математический сайт
5. <http://www.library.biophys.msu.ru/MathMod/BM.HTML> - книга Г.Ю. Ризниченко «Биология математическая»
6. <http://nature.web.ru/db/msg.html?mid=1156624&uri=index.htm> - Бейли Н.. Математика в биологии и медицине. – М.: Мир, 1970.
7. <http://www.biometrica.tomsk.ru/> - электронный журнал «Биометрика» для медиков и биологов – сторонников доказательной биомедицины. Содержит большое количество статей и иных материалов, посвященных математическим моделям в биологии.
8. <http://www.library.biophys.msu.ru/FominBerk/main.htm> - Фомин С.В., Беркинблит М.Б. Математические проблемы в биологии. - М.: Гаука, 1973. - 200 с.
9. <https://www.elibrary.ru> – электронная библиотека научных статей, монографии и материалов конференций, выпущенных Российскими учеными.
10. <https://pubmed.ncbi.nlm.nih.gov/> - международная база данных научных статей и монографий, посвященная различным вопросам биологии.
11. <https://apps.webofknowledge.com> – международная база данных, индексирующая научные публикации в высокорейтинговых изданиях

VI. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебно-лабораторное оборудование:

- Аудитория для проведения занятий лабораторного типа. Компьютерный класс (учебная аудитория). Аудитория оборудована: специализированной (учебной) мебелью на 20 посадочных мест, доской меловой; оборудована техническими средствами обучения: Системный блок PentiumG850, Монитор BenQ G252HDA-1 шт.; Системный блокAthlon 2 X2 250, Монитор BenQ G252HDA – 8 шт.; Системный блок PentiumD 3.0GHz, Монитор Samsung 740N – 3 шт.; Моноблок IRU T2105P – 2 шт.; Системный блок Pentium G3250, Монитор BenQG955 – 1 шт.; Системный блок Pentium G3250, Монитор BenQ GL2250 – 1 шт.; Системный блок Pentium G3250, Монитор Samsung T200 HD – 1 шт.; Системный блок Pentium G3250, Монитор Samsung T190N – 1 шт.; Системный блок Pentium G3250, Монитор Samsung 740N – 1 шт.; Проектор BenQ MX503; экран ScreenVtdiaEcot. С неограниченным доступом к сети Интернет и обеспечением доступа в электронную информационно-образовательную среду организации, учебно-наглядными пособиями, обеспечивающими тематические иллюстрации по дисциплине «Моделирование и программирование биопроцессов» в количестве 8 шт., презентации по каждой теме программы.

- Компьютерный класс (учебная аудитория) для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, организации самостоятельной работы. Аудитория оборудована: специализированной (учебной) мебелью на 20 посадочных мест, доской меловой; оборудована техническими средствами обучения: Системный блок PentiumG850, Монитор BenQ G252HDA-1 шт.; Системный блок Athlon 2 X2 250, Монитор BenQ G252HDA – 8 шт.; Системный блок PentiumD 3.0GHz, Монитор Samsung 740N – 3 шт.; Моноблок IRU T2105P – 2 шт.; Системный блок Pentium G3250, Монитор BenQG955 – 1 шт.; Системный блок Pentium G3250, Монитор BenQ GL2250 – 1 шт.; Системный блок Pentium G3250, Монитор Samsung T200 HD – 1 шт.; Системный блок Pentium G3250, Монитор Samsung T190N – 1 шт.; Системный блок Pentium G3250, Монитор Samsung 740N – 1 шт.; с неограниченным доступом к сети Интернет; Проектор BenQ MX503; экран ScreenVtdiaEcot. Ноутбук Lenovo G580 – 1 шт. С неограниченным доступом к сети Интернет.

- Помещения для хранения и профилактического обслуживания учебного оборудования. Аудитория оборудована: специализированной мебелью на 11 посадочных мест; Шкаф для документов - 3 шт.; Сейф – 1 шт.; Шкаф-купе - 2 шт.; Принтер цв. Canon LBR-5050 Laser Printer; Принтер Canon LBP-3010; Ноутбук Lenovo G580 – 1 шт.

6.2. Программное обеспечение:

DreamSpark Premium Electronic Software Delivery (3 years) Renewal (Windows 10 Education 32/64-bit (Russian) - Microsoft Imagine, Windows 7 Professional with Service Pack 1 32/64-bit (English) - Microsoft Imagine, Windows Server 2008 Enterprise and Standard without Hyper-V with SP2 32/64-bit (English) - Microsoft Imagine, Access 2016 32/64-bit (Russian) - Microsoft Imagine, Access 2010 32/64-bit (Russian) - Microsoft Imagine). Договор №03-016-14 от 30.10.2014г.

Kaspersky Endpoint Security для бизнеса - Стандартный Russian Edition. 250-499. Форус Контракт №04-114-16 от 14 ноября 2016г KES. Счет №РСЦЗ-000147 и АКТ от 23 ноября 2016г Лиц. №1В08161103014721370444.

Microsoft Office Enterprise 2007 Russian Academic OPEN No Level. Номер Лицензии Microsoft 43364238.

Microsoft Windows XP Professional Russian Upgrade Academic OPEN No Level. Номер Лицензии Microsoft 41059241.

Office 365 профессиональный плюс для учащихся. Номер заказа: 36dde53d-7cdb-4cad-a87f-29b2a19c463e.

6.3. Технические и электронные средства:

Презентации по всем темам курса.

VII. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

При реализации различных видов учебной работы дисциплины используются как стандартные методы обучения, так и интерактивные формы проведения занятий.

Стандартные методы обучения:

1. Информационная лекция.
2. Практические занятия, предназначенные для освоения студентами базовых методов анализа данных в физико-химической экологии.
3. Самостоятельная работа студентов.
4. Консультации преподавателя.
5. Подготовка ответов на контрольные вопросы.
6. Обучения с применением интерактивных форм образовательных технологий.
7. Кейс-метод – обучение в контексте моделируемой ситуации, воспроизводящей реальные условия научной деятельности (разбор конкретных ситуаций).

8. Информационно-коммуникационные образовательные технологии – лекция-визуализация, представление результатов деятельности (рефератов и отчетов) с использованием специализированных программных сред.
9. Интернет-технология – задействование образовательного портала ИГУ - educa.isu.ru для предоставления письменных отчетов по домашним работам.

Наименование тем занятий с использованием активных форм обучения:

№	Тема занятия	Вид занятия	Форма / Методы интерактивного обучения	Кол-во часов
1	Тема 1. Введение, что такое машинное обучение.	Подготовка доклада	Загрузка презентации для контроля на образовательный портал ИГУ educa.isu.ru	6
2	Тема 2. Обучение с учителем. Логистическая регрессия.	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ educa.isu.ru	4
3	Тема 3. Обучение с учителем. Линейная классификация.	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ educa.isu.ru	6
4	Тема 4. Обучение с учителем. Метод k ближайших соседей.	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ educa.isu.ru	4
5	Тема 5. Обучение с учителем. Деревья решений.	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ educa.isu.ru	6
6	Тема 6. Обучение с учителем. Метод опорных векторов.	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ educa.isu.ru	4
7	Тема 7. Обучение с учителем. Деревья решений, ансамбли деревьев решений.	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ educa.isu.ru	4
8	Тема 8. Обучение с учителем. Наивный Байесовский классификатор	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ	4

			educa.isu.ru	
9	Тема 9. Обучение с учителем. Нейронные сети (глубокое обучение).	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ educa.isu.ru	6
	Тема 10. Обучение без учителя. Типы машинного обучения без учителя.	Подготовка доклада	Загрузка презентации для контроля на образовательный портал ИГУ educa.isu.ru	8
	Тема 11. Обучение без учителя. Предварительная обработка и масштабирование данных.	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ educa.isu.ru	8
	Тема 12. Обучение без учителя. Анализ методом главных компонент.	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ educa.isu.ru	8
	Тема 13. Обучение без учителя. Кластеризация k средних.	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ educa.isu.ru	8
	Тема 14. Обучение без учителя. Иерархическая кластеризация.	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ educa.isu.ru	8
	Тема 15. Обучение без учителя. Оценки достоверности кластеризации.	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ educa.isu.ru	8
	Тема 16. Оценка и улучшения качества моделей машинного обучения	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ educa.isu.ru	10
	Тема 17. Объединение алгоритмов в цепочки и конвейеры.	самостоятельная работа	Загрузка задания для контроля на образовательный портал ИГУ educa.isu.ru	10
Итого часов				110

VIII. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

В рамках дисциплины «Искусственный интеллект и машинное обучение» используются следующие формы текущего контроля:

- устный опрос;
- устный доклад по теме;
- письменная работа по решению домашних заданий;

Фонд оценочных средств включает:

- перечень тем докладов по темам дисциплины;
- вопросы для самостоятельного изучения (срс);
- задачи для самостоятельного домашнего решения;
- вопросы и билеты для зачета.

Назначение оценочных средств: выявить сформированность компетенций ОПК-6, ОПК-7 (см. п. III).

Перечень темы устных докладов

1. Тема 1. Использование методов обучения с учителем для анализа биологических последовательностей (ДНК, РНК, белки).
2. Тема 10. Использование методов машинного обучения без учителя в биомедицине для диагностики заболеваний.

Перечень вопросов для самостоятельного изучения

1. Самостоятельное изучение темы – использование нейронных сетей для распознавания образов и идентификации биологических видов.

Перечень домашних задания для самостоятельного выполнения

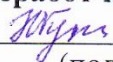
- Тема 1. Решение задач по теме введение в языки программирования R и Python как инструменты реализации методов машинного обучения
- Тема 2. Решение задач по теме линейная регрессия
- Тема 3. Решение задач по теме линейная классификация.
- Тема 4. Решение задач по теме метод k ближайших соседей.
- Тема 5. Решение задач по теме анализ данных с помощью дерева решений.
- Тема 6. Решение задач по теме анализ данных с помощью ансамблей деревьев решений
- Тема 7. Решение задач по теме анализ данных методом опорных векторов.
- Тема 8. Решение задач по теме наивный Байесовский классификатор
- Тема 9. Решение задач по теме конструирование нейронных сетей для анализа биологических данных
- Тема 10. Решение задач по теме введение в языки программирования R и Python как инструменты реализации методов машинного обучения без учителя
- Тема 11. Решение задач по теме масштабирование данных.
- Тема 12. Решение задач по теме анализ методом главных компонент.
- Тема 13. Решение задач по теме кластеризация k средних.
- Тема 14. Решение задач по теме иерархическая кластеризация.
- Тема 15. Решение задач по теме оценка достоверности кластеризации.
- Тема 16. Решение задач по теме оценка и улучшения качества моделей машинного обучения
- Тема 17. Решение задач по теме разработка цепочек и конвейеров для анализа данных методами машинного обучения.

Примерный перечень вопросов к экзамену

1. Что такое машинное обучение? Какие основные типы задач оно решает?
2. В чем разница между обучением с учителем (supervised learning) и без учителя (unsupervised learning)?
3. Какие основные этапы процесса машинного обучения?

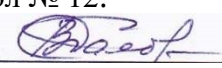
4. Что такое переобучение (overfitting) и как его избежать?
5. Что такое регуляризация и зачем она нужна?
6. Какие методы используются для обработки пропущенных значений в данных?
7. Как обрабатывать выбросы (outliers) в данных?
8. Что такое масштабирование данных? Какие методы масштабирования вы знаете?
9. Как кодировать категориальные признаки? В чем разница между One-Hot Encoding и Label Encoding?
10. Что такое дисбаланс классов и как с ним бороться?
11. Какие метрики используются для оценки моделей классификации?
12. Что такое ROC-AUC и как интерпретировать его значение?
13. Какие метрики используются для оценки моделей регрессии?
14. Что такое кросс-валидация и зачем она нужна?
15. Как анализировать ошибки модели с помощью confusion matrix?
16. Какие алгоритмы машинного обучения вы знаете? В чем их основные различия?
17. Что такое гиперпараметры модели? Как их настраивать?
18. В чем разница между Grid Search и Random Search?
19. Что такое ансамбли моделей? Какие типы ансамблей вы знаете?
20. Как работает метод случайного леса (Random Forest)?
21. Как выбрать подходящую модель для конкретной задачи?
22. Какие библиотеки для машинного обучения вы знаете? В чем их особенности?
23. Как визуализировать результаты работы модели?
24. Как интерпретировать важность признаков (feature importance)?
25. Какие подходы используются для улучшения качества модели?
26. Как предобработать данные для задачи классификации?
27. Как построить конвейер для задачи регрессии?
28. Как оценить качество модели кластеризации?
29. Как использовать конвейеры для работы с текстовыми данными?
30. Как применить ансамбли моделей для улучшения качества предсказаний?
31. Что такое функция потерь (loss function)? Какие функции потерь вы знаете?
32. Как работает градиентный спуск (gradient descent)?
33. В чем разница между байесовскими и частотными подходами в машинном обучении?
34. Что такое бутстрап (bootstrap) и как он используется в машинном обучении?
35. Как работает метод главных компонент (PCA)?

Разработчик:

 _____ доцент Букин Ю.С.
(подпись)

Программа составлена в соответствии с требованиями ФГОС ВО по направлению 06.03.01 Биология

Программа рассмотрена на заседании кафедры физико-химической биологии, биоинженерии и биоинформатики 19.03.2025 г. протокол № 12.

Зав. кафедрой, д.б.н., профессор В.П. Саловарова 

Настоящая программа, не может быть воспроизведена ни в какой форме без предварительного письменного разрешения кафедры-разработчика программы.