

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Раджабов Камаломжид Урбанович

Должность: Исполнительный директор

Дата подписания: 2023.08.21:06:03

Уникальный программный ключ:

088d9d84706d89075c4a5aa1678d7c4c996222db



**МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ**  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
**«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ –**  
**МСХА имени К.А. ТИМИРЯЗЕВА»**  
**(ФГБОУ ВО РГАУ - МСХА имени К.А. Тимирязева)**

Институт садоводства и ландшафтной архитектуры  
Кафедра ботаники, селекции и семеноводства садовых растений

УТВЕРЖДАЮ:

Директор института садоводства и  
ландшафтной архитектуры

Раджабов А.К.

“21” августа 2023 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**  
**Б1.В.06.09 Основы машинного обучения в биологии**

для подготовки бакалавров

ФГОС ВО

Направление 35.04.05 Садоводство

Направленность (программа) «Биотехнология и селекция растений»

Курс: 4

Семестр: 8

Форма обучения: очная

Год начала подготовки: 2023

Регистрационный номер \_\_\_\_\_

Москва, 2023

Разработчики(и): С.Г. Монахос, д.с.-х.н., профессор

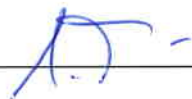
Д.Д. Лисовая, ассистент

  
\_\_\_\_\_  
(подпись)  
\_\_\_\_\_  
(подпись)

«21» августа 2023 г.

Рецензент: Монахос Г.Ф., к.с.-х.н., ст.н.с.

(ФИО, ученая степень, ученое звание)

  
\_\_\_\_\_


«21» августа 2023 г.

Программа составлена в соответствии с требованиями ФГОС ВО, по направлению подготовки 35.03.05 Садоводство и учебного плана.

Программа обсуждена на заседании кафедры ботаники, селекции и семеноводства садовых растений, протокол №15 от «21» августа 2023 г.

Зав. кафедрой С.Г. Монахос, д.с.-х.н., профессор

(ФИО, ученая степень, ученое звание)

  
\_\_\_\_\_  
(подпись)

«21» августа 2023 г.

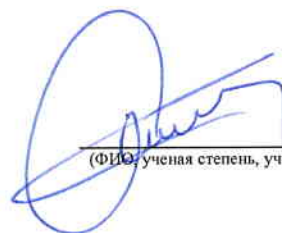
**Согласовано:**

Председатель учебно-методической комиссии факультета Маланкина Е.Л., д.с.-х.н.

  
\_\_\_\_\_  
(ФИО, ученая степень, ученое звание) (подпись)

«21» августа 2023 г.

Заведующий выпускающей кафедрой С.Г. Монахос, д.с.-х.н., профессор

  
\_\_\_\_\_  
(ФИО, ученая степень, ученое звание) (подпись)

«21» августа 2023 г.

Зав. Отделом комплектования ЦНБ

  
\_\_\_\_\_  
(подпись)

**Бумажный экземпляр ПП, электронные варианты ПП и оценочных материалов получены:**

Методический отдел УМУ \_\_\_\_\_ «\_\_» \_\_\_\_\_ 20\_\_ г

## СОДЕРЖАНИЕ

<b>1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ .....</b>	<b>4</b>
<b>2. МЕСТО ДИСЦИПЛИНЫ В УЧЕБНОМ ПРОЦЕССЕ.....</b>	<b>4</b>
<b>3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ «ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ В БИОЛОГИИ», СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.....</b>	<b>5</b>
<b>4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ .....</b>	<b>5</b>
4.1 РАСПРЕДЕЛЕНИЕ ТРУДОЁМКОСТИ ДИСЦИПЛИНЫ ПО ВИДАМ РАБОТ .....	5
ПО СЕМЕСТРАМ .....	5
4.2 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ .....	8
4.3 ПРАКТИЧЕСКИЕ ЗАНЯТИЯ .....	10
<b>5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ.....</b>	<b>13</b>
<b>6. ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ.....</b>	<b>14</b>
6.1. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ И НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ.....	14
<b>7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ .....</b>	<b>17</b>
7.1 ОСНОВНАЯ ЛИТЕРАТУРА.....	17
7.2 ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА .....	17
<b>8. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫХ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ) .....</b>	<b>17</b>
<b>9. ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ .....</b>	<b>17</b>
<b>10. ОПИСАНИЕ МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ .....</b>	<b>18</b>
<b>11. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ СТУДЕНТАМ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ.....</b>	<b>18</b>
Виды и формы отработки пропущенных занятий.....	19
<b>12. МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПРЕПОДАВАТЕЛЯМ ПО ОРГАНИЗАЦИИ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ .....</b>	<b>19</b>

# АННОТАЦИЯ

## рабочей программы учебной дисциплины

### Б1.В.06.09 «Основы машинного обучения в биологии»

для подготовки магистра по направлению 35.03.05 Садоводство  
направленности «Селекция, генетика и биотехнология садовых культур»

**Цель освоения дисциплины:** изучение современных математических методов машинного обучения, предназначенных для анализа данных и построения предсказательных моделей.

**Место дисциплины в учебном плане:** дисциплина включена в часть, формируемую участниками образовательных отношений, учебного плана по направлению подготовки 35.03.05 Садоводство

**Требования к результатам освоения дисциплины:** в результате освоения дисциплины формируются следующие компетенции: 1 профессиональную компетенцию ПКос-2 (ПКос-2.1; ПКос-2.2; ПКос-2.3).

**Краткое содержание дисциплины:** Дисциплина "Основы машинного обучения в биологии" представляет собой курс, направленный на обучение студентов использованию современных математических методов машинного обучения, предназначенных для анализа данных и построения предсказательных моделей. Курс охватывает основные концепции и методы анализа данных, включая обработку, визуализацию и статистический анализ биологических данных.

**Общая трудоемкость дисциплины:** 144/4 (часы/зач. ед.)

**Промежуточный контроль:** Зачет

#### 1. Цель освоения дисциплины

Цель данной дисциплины заключается в формировании у студентов теоретических знаний и практических навыков по основам машинного обучения, овладение студентами инструментарием, моделями и методами машинного обучения, а также приобретение навыков исследователя данных (data scientist) и разработчика математических моделей, методов и алгоритмов анализа данных.

#### 2. Место дисциплины в учебном процессе

Дисциплина «Основы машинного обучения в биологии» включена в часть профессионального цикла, формируемую участниками образовательных отношений. Реализация в дисциплине «Основы машинного обучения в биологии» требований ФГОС ВО, ОПОП и Учебного плана по направлению 35.03.05 Садоводство для подготовки бакалавров направленности «Селекция, генетика и биотехнология садовых культур».

Предшествующими курсами, на которых непосредственно базируется дисциплина «Основы машинного обучения в биологии», являются «Математика и математическая статистика», «Цифровые технологии в АПК», «Искусственный интеллект в АПК».

Дисциплина «Основы машинного обучения в биологии» является основополагающей для изучения следующих дисциплин: «Основы биоинформатики в селекции растений», «Компьютерная графика в садоводстве».

Данная дисциплина представляет собой курс, который призван обучить студентов использованию теоретических знаний и практических навыков по основам машинного обучения, овладение студентами инструментарием, моделями и методами машинного обучения, а также приобретение навыков исследователя данных (data scientist) и разработчика математических моделей, методов и алгоритмов анализа данных.

Рабочая программа дисциплины «Основы машинного обучения в биологии» для инвалидов и лиц с ограниченными возможностями здоровья разрабатывается индивидуально с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся.

### **3. Перечень планируемых результатов обучения по дисциплине «Основы машинного обучения в биологии», соотнесенных с планируемыми результатами освоения образовательной программы**

Изучение данной учебной дисциплины направлено на формирование у обучающихся компетенций, представленных в таблице 1.

## **4. Структура и содержание дисциплины**

### **4.1 Распределение трудоёмкости дисциплины по видам работ по семестрам**

Общая трудоёмкость дисциплины составляет 2 зач.ед. (72 часов), их распределение по видам работ семестрам представлено в таблице 2.

Таблица 1

Требования к результатам освоения учебной дисциплины

№ п/п	Код компетенции	Содержание компетенции (или её части)	Индикаторы компетенций	В результате изучения учебной дисциплины обучающиеся должны:		
				знать	уметь	владеть
1.	ПКос-2	Способен самостоятельно в качестве руководителя или члена коллектива организовывать и управлять производственной и научно-исследовательской деятельностью в избранной и смежных предметных областях	ПКос-2.1 правила и методики анализа результатов научных исследований, специализированное программное обеспечение, способы обработки полученных эмпирических данных и их интерпретации	информационные ресурсы, посвященные применению методов машинного обучения	пользоваться средой Jupyter Notebook для быстрого применения алгоритмов машинного обучения работать с различными, визуализировать их	базовым инструментарием для коллективной разработки алгоритмов машинного обучения
			ПКос-2.2 анализировать получаемую полевую и лабораторную биологическую информацию с использованием современной вычислительной техники; систематизировать экспериментальные данные; обобщать полученные результаты в контексте работы в накопленных в науке знаний; получать новые Досто-	принципы работы составных структур данных, способы векторизации вычислений для ускорения расчетов	оценивать простые метрики качества работы алгоритмов классификации и восстановления регрессии	навыками создания интерактивных отчетов для задач машинного обучения

				верные факты на основе наблюдений, опытов, научного анализа эмпирических данных; представлять результаты научных исследований; нести ответственность за качество выполняемых работ	
				ПКос-2.3 навыками подготовки научных публикаций, отчетов, обзоров, патентов и докладов; участия в организации и проведении научных семинаров и конференций; статистическими методами сравнения полученных экспериментальных данных и определения закономерностей; способностью формулировать выводы и практические рекомендации на основе репрезентативных и оригинальных результатов исследований	
			подходы к ускорению работы и улучшению схожимости методов машинного обучения		
					осуществлять поиск по документации специализированных библиотек машинного обучения
					навыками интеграции различных библиотек для решения комплексных задач обработки данных

## ОЧНАЯ ФОРМА ОБУЧЕНИЯ

Таблица 2а

### Распределение трудоёмкости дисциплины по видам работ по семестрам

Вид учебной работы	Трудоёмкость
	час.
<b>Общая трудоёмкость</b> дисциплины по учебному плану	<b>72</b>
<b>1. Контактная работа:</b>	<b>48,25</b>
<b>Аудиторная работа</b>	<b>48</b>
<i>в том числе:</i>	
<i>лекции (Л)</i>	12
<i>лабораторные работы (ЛР)</i>	24
<i>практические занятия (ПЗ)</i>	12/4
<i>контактная работа на промежуточном контроле (КРА)</i>	0,25
<b>2. Самостоятельная работа (СРС)</b>	<b>23,75</b>
<i>самостоятельное изучение разделов, самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к лабораторным и практическим занятиям, коллоквиумам и т.д.)</i>	23,75
Вид промежуточного контроля:	Зачет

## 4.2 Содержание дисциплины

Таблица 3

### Тематический план учебной дисциплины

Наименование разделов и тем дисциплин (укрупнённо)	Всего	Аудиторная работа				Внеаудиторная работа СР
		Л	ПЗ/С	ЛР	ПКР	
<b>Основы машинного обучения в биологии</b>	<b>72</b>	<b>12</b>	<b>12/4</b>	<b>24</b>	<b>0,25</b>	<b>23,75</b>
<b>Раздел 1. Типы задач. Метрические классификаторы. Алгоритмы кластеризации.</b>	<b>24</b>	<b>4</b>	<b>4</b>	<b>8</b>	<b>0,25</b>	<b>8</b>
Тема 1. Типы задач машинного обучения.	9	2	1	2	-	4
Тема 2. Метрические классификаторы.	6	1	1	2	-	2
Тема 3. Алгоритмы кластеризации.	9	1	2	4	-	2
<b>Раздел 2. Деревья решений, линейные классификаторы. Нейронные сети</b>	<b>24</b>	<b>4</b>	<b>4</b>	<b>8</b>	<b>-</b>	<b>8</b>
Тема 1. Деревья решений.	11	2	1	4	-	4
Тема 2. Линейные методы.	7	1	2	2	-	2
Тема 3. Нейронные сети и глубокое обучение.	6	1	1	2	-	2
<b>Раздел 3. Регрессионный анализ, Ансамблевые методы. Стохастический поиск.</b>	<b>24</b>	<b>4</b>	<b>4</b>	<b>7</b>	<b>-</b>	<b>7,75</b>
Тема 1. Регрессионный анализ	12,75	2	2	5	-	2,75
Тема 2. Ансамблевые методы	6	1	1	2	-	2
Тема 3. Стохастический поиск	6	1	2	-	-	3



Наименование разделов и тем дисциплин (укрупнённо)	Всего	Аудиторная работа				Внеаудиторная работа СР
		Л	ПЗ/С	ЛР	ПКР	
Контактная работа на промежуточном контроле (КРА)	0,25	-	-		0,25	-
<b>Итого по дисциплине</b>	72	12	24		2,4	70

### **Раздел 1. Типы задач. Метрические классификаторы. Алгоритмы кластеризации.**

Тема 1. Типы задач машинного обучения.

Предмет и задачи машинного обучения и анализа данных. Основные принципы, задачи и подходы, использование в различных областях науки и промышленности. Основные этапы эволюции алгоритмов машинного обучения.

Тема 2. Метрические классификаторы.

Общий вид метрического классификатора. Алгоритм К ближайших соседей. Алгоритмы отбора эталонов.

Тема 3. Алгоритмы кластеризации.

Типы кластерных структур. Функционал качества кластеризации. EM-алгоритм. Метод k-средних. Иерархическая кластеризация. Формула Ланса-Уильямса. Быстрая агломеративная кластеризация.

### **Раздел 2. Деревья решений, линейные классификаторы. Нейронные сети**

Тема 1. Деревья решений.

Правила и анализ качества (точность, полнота). Анализ с помощью ROC кривой. Алгоритм построения деревьев решений. Критерий информационного выигрыша и критерий Джини. Леса решающих деревьев.

Тема 2. Линейные методы.

Постановка задач линейной регрессии и линейной классификации. Метод наименьших квадратов в матричной форме. Аналитическое решение. Мультиколлинеарность и плохая обусловленность ковариационной матрицы. Метод лассо. Метод стохастического градиента. Улучшение сходимости метода SGD. Метод опорных векторов. Линейно разделимые выборки. Двойственная задача. Нелинейные обобщения. Возможные виды ядер.

Тема 3. Нейронные сети и глубокое обучение.

Логистическая регрессия. Градиентный спуск. Нейронные сети и алгоритм обратного распространения градиента. Глубокое обучение, свертки и пулинг.

### **Раздел 3. Регрессионный анализ, Ансамблевые методы. Стохастический поиск.**

Тема 1. Регрессионный анализ.

Линейная регрессия. Полиномиальная регрессия. Смещение и дисперсия. Гребневая регрессия.

Тема 2. Ансамблевые методы.

Голосование. Бутстраппинг. Бустинг, адаптивный бустинг, градиентный бустинг.

Тема 3. Стохастический поиск.

Монте-Карло поиск. Алгоритм симулированного отжига. Генетический алгоритм.

### 4.3 Практические занятия

Таблица 4

#### Содержание практических занятий и контрольные мероприятия

№ п/п	Название раздела, темы	№ и название лекций/ практических/ семинарских занятий	Формируемые компетенции	Вид контрольного мероприятия	Кол-во часов
<b>Раздел 1. Типы задач. Метрические классификаторы. Алгоритмы кластеризации.</b>			ПКос-2, ПКос-3	устный опрос контрольная работа 1	<b>16</b>
1	Тема 1. Типы задач машинного обучения.	Лекционное занятие №1. Типы задач машинного обучения.	ПКос-2, ПКос-3		2
		Практическое занятие №1. Основные принципы, задачи и подходы, использование в различных областях науки и индустрии	ПКос-2, ПКос-3	устный опрос	1
		Лабораторное занятие №1. Первичный анализ данных с библиотекой Pandas	ПКос-2, ПКос-3		2
2	Тема 2. Метрические классификаторы.	Лекционное занятие №2. Метрические классификаторы.	ПКос-2, ПКос-3		1
		Практическое занятие №2. Визуальный анализ данных с библиотеками Seaborn и Matplotlib	ПКос-2, ПКос-3	устный опрос	1
		Лабораторное занятие №2. Обобщенный метрический классификатор. Виды метрик. Метод ближайшего соседа. Алгоритм k-ближайших соседей. Взвешенная версия алгоритма k-ближайших соседей. Метод окна Парзена.	ПКос-2, ПКос-3		1
3	Тема 3. Алгоритмы кластеризации.	Лекционное занятие №3. Алгоритмы кластеризации	ПКос-2, ПКос-3		2
		Практическое занятие №3.	ПКос-2, ПКос-3	устный опрос	2

№ п/п	Название раздела, темы	№ и название лекций/ практических/ семинарских занятий	Формируемые компетенции	Вид контрольного мероприятия	Кол-во часов
		Обучение без учителя. Метод главных компонент. Кластеризация.			
		Лабораторное занятие №3. Основы анализа временных рядов	ПКос-2, ПКос-3		4
		Контрольная работа по темам № 1-3	ПКос-2, ПКос-3	контрольная работа	
<b>Раздел 2. Деревья решений, линейные классификаторы. Нейронные сети</b>			ПКос-2, ПКос-3	устный опрос контрольная работа 2	<b>16</b>
4	Тема 1. Деревья решений.	Лекционное занятие №4. Деревья решений.	ПКос-2, ПКос-3		2
		Практическое занятие №4. Разведывательный анализ данных в R. Базовые графики в R	ПКос-2, ПКос-3	устный опрос	1
		Лабораторное занятие №4. Деревья решений, логистическая регрессия и случайный лес.	ПКос-2, ПКос-3		4
5	Тема 2. Линейные методы.	Лекционное занятие №5. Линейные классификаторы	ПКос-2, ПКос-3		1
		Практическое занятие №5. Линейные модели классификации и регрессии	ПКос-2, ПКос-3	устный опрос	2
		Лабораторное занятие №5. Метод лассо. Метод стохастического градиента. Улучшение сходимости метода SGD. Метод опорных векторов.	ПКос-2, ПКос-3		2
6	Тема 3. Нейронные сети и глубокое обучение.	Лекционное занятие №6. Нейронные сети и глубокое обучение.	ПКос-2, ПКос-3		1
		Практическое занятие №6. Нейронные сети и алгоритм обратного распространения градиента.	ПКос-2, ПКос-3	устный опрос	1
		Лабораторное занятие №6. Глубокое обучение, свертки и пулинг	ПКос-2, ПКос-3		2
		Контрольная работа по темам № 1-3	ПКос-2, ПКос-3	контрольная работа 2	

№ п/п	Название раздела, темы	№ и название лекций/ практических/ семинарских занятий	Формируемые компетенции	Вид контрольного мероприятия	Кол-во часов
<b>Раздел 3. Регрессионный анализ, Ансамблевые методы. Стохастический поиск.</b>			ПКос-2, ПКос-3	устный опрос контрольная работа 3	<b>17</b>
7	Тема 1. Регрессионный анализ	Лекционное занятие №7. Регрессионный анализ.	ПКос-2, ПКос-3		2
		Практическое занятие №7. Логистическая регрессия и случайный лес.	ПКос-2, ПКос-3	устный опрос	2
		Лабораторное занятие №7. Линейная регрессия, Lasso и RF-регрессия.	ПКос-2, ПКос-3		5
8	Тема 2. Ансамблевые методы	Лекционное занятие №8. Ансамблевые методы.	ПКос-2, ПКос-3		1
		Практическое занятие №8. Голосование. Бутстраппинг.	ПКос-2, ПКос-3	устный опрос	1
		Лабораторное занятие №8. Бустинг, адаптивный бустинг, градиентный бустинг.	ПКос-2, ПКос-3		2
9	Тема 3. Стохастический поиск	Лекционное занятие №9. Стохастический поиск.	ПКос-2, ПКос-3		1
		Практическое занятие №8. Монте-Карло поиск. Алгоритм симулированного отжига. Генетический алгоритм.	ПКос-2, ПКос-3	устный опрос	2
		Контрольная работа по темам № 1-3	ПКос-2, ПКос-3	контрольная работа 3	

Таблица 5

**Перечень вопросов для самостоятельного изучения дисциплины**

№ п/п	Название раздела, темы	Перечень рассматриваемых вопросов для самостоятельного изучения
<b>Раздел 1. Типы задач. Метрические классификаторы. Алгоритмы кластеризации.</b>		
1.	Тема 1. Типы задач машинного обучения.	Общая постановка задачи машинного обучения. Обучение с учителем и без учителя, обучение с подкреплением. Задачи классификации, восстановления регрессии, предсказания. Модели алгоритмов. Признаки. Типы признаков. Понятие функционала качества. Вероятностная постановка задачи. Оценка обобщающей способности. Проблема переобучения. Критерии оценки качества работы алгоритмов машинного обучения. ROC-кривые. Примеры практических задач машинного обучения. ПКос-2, ПКос-3

№ п/п	Название раздела, темы	Перечень рассматриваемых вопросов для самостоятельного изучения
2.	Тема 2. Метрические классификаторы.	Обобщенный метрический классификатор. Виды метрик. Метод ближайшего соседа. Алгоритм k-ближайших соседей. Взвешенная версия алгоритма k-ближайших соседей. Метод окна Парзена. Метод потенциальных функций. Понятие эталона. Отступы и классификация объектов. ПКос-2, ПКос-3
3.	Тема 3. Алгоритмы кластеризации.	Алгоритмы кластеризации с фиксированным количеством кластеров. Алгоритмы кластеризации по плотности. Иерархическая кластеризация. ПКос-2, ПКос-3
<b>Раздел 2. Деревья решений, линейные классификаторы. Нейронные сети</b>		
1.	Тема 1. Деревья решений.	Правила и анализ качества (точность, полнота). Анализ с помощью ROC кривой. Алгоритм построения деревьев решений. Критерий информационного выигрыша и критерий Джини. Леса решающих деревьев. ПКос-2, ПКос-3
2.	Тема 2. Линейные методы.	Постановка задач линейной регрессии и линейной классификации. Метод наименьших квадратов в матричной форме. Аналитическое решение. Мультиколлинеарность и плохая обусловленность ковариационной матрицы. Гребневая регрессия. Метод лассо. Линейные классификаторы. Метод стохастического градиента. Улучшение сходимости метода SGD. Логистическая регрессия. Метод опорных векторов. Линейно разделимые выборки. Двойственная задача. Нелинейные обобщения. Возможные виды ядер. ПКос-2, ПКос-3
3.	Тема 3. Нейронные сети и глубокое обучение.	Логистическая регрессия. Градиентный спуск. Нейронные сети и алгоритм обратного распространения градиента. Глубокое обучение, свертки и пулинг. ПКос-2, ПКос-3
<b>Раздел 3. Регрессионный анализ, Ансамблевые методы. Стохастический поиск.</b>		
1.	Тема 1. Регрессионный анализ	Линейная регрессия. Полиномиальная регрессия. Смещение и дисперсия. Гребневая регрессия. ПКос-2, ПКос-3
2.	Тема 2. Ансамблевые методы	Голосование. Бутстраппинг. Бустинг, адаптивный бустинг, градиентный бустинг. ПКос-2, ПКос-3
3.	Тема 3. Стохастический поиск	Монте-Карло поиск. Алгоритм симулированного отжига. Генетический алгоритм. ПКос-2, ПКос-3

## 5. Образовательные технологии

Таблица 6

### Применение активных и интерактивных образовательных технологий

№ п/п	Тема и форма занятия		Наименование используемых активных и интерактивных образовательных технологий (форм обучения)
1.	Тема 1. Типы задач машинного обучения.	ПЗ	Интерактивная форма: мастер-класс
2.	Тема 3. Алгоритмы кластеризации.	ПЗ	Интерактивная форма: мастер-класс

## 6. Текущий контроль успеваемости и промежуточная аттестация по итогам освоения дисциплины

### 6.1. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков и (или) опыта деятельности

Вопросы для подготовки к контрольным мероприятиям (текущий контроль)

#### Устный опрос

1. Постановка задачи машинного обучения. Обучение с учителем и без учителя. Типы признаков
2. Типы задач. Линейные модели
3. Примеры задач обучения без учителя (3-5 примеров)
4. Примеры задач обучения с учителем (3-5 примеров)
5. Функционал качества. Минимизация эмпирического риска
6. Переобучение. Методы валидации моделей. Кросс-валидация
7. Метрические методы классификации. Виды расстояний
8. Метод ближайшего соседа. Метод k ближайших соседей. Взвешенный метод k ближайших соседей
9. Отбор эталонных объектов. Понятие отступа. Классификация объектов
10. Регрессия по соседним объектам. Окно Парзена
11. Линейная регрессия. Постановка задачи. Матричная формулировка. Точное решение
12. Регуляризация в линейной регрессии. Виды регуляризаторов
13. Гребневая регрессия. Лассо-регрессия
14. Итеративная версия решения задачи линейной регрессии без регуляризации и с регуляризацией. Метод градиентного спуска
15. Вероятностная модель данных. Максимум правдоподобия
16. Линейный классификатор. Отступы. Функционал качества
17. Метод стохастического градиента (SGD). Преимущества и недостатки
18. Методы улучшения сходимости SGD (метод моментов, метод Нестерова)
19. Методы улучшения сходимости SGD (AdaGrad, RMSProp, Adam)
20. Логистическая регрессия. Простой байесовский классификатор. Сигмоидальная функция
21. Метод опорных векторов (SVM). Постановка задачи для линейно разделимой выборки без выбросов
22. Отступы в SVM. Учет выбросов. Формулировка теоремы Куна-Таккера. Двойственная задача
23. Классификация объектов в SVM. Постановка задачи через двойственные переменные  $\lambda$
24. Нелинейные ядра в методе SVM. Примеры ядер. Способы их построения
25. Основные метрики качества алгоритмов. ROC-кривые
26. Методы кластеризации. Типы кластерных структур
27. Функционал качества кластеризации.

28. EM-алгоритм в кластерном анализе
29. Метод k-средних
30. Иерархическая кластеризация. Формула Ланса-Уильямса
31. Быстрая агломеративная кластеризация.

### **Контрольная работа №1**

- Задача 1. Реализуйте алгоритм kNN классификации по k ближайшим соседям, используя простое евклидовое расстояние.
- Задача 2. Реализуйте алгоритм k-means для кластеризации на 2-4 кластера.
- Задача 3. Реализуйте алгоритм DBSCAN, найдите параметры для кластеризации на 4 кластера.

### **Контрольная работа №2**

- Задача 1. Реализуйте алгоритмы построения дерева с критерием информационного выигрыша и критерием Джини и определению класса по мажоритарному классу в листе. Найдите оптимальную глубину дерева в обоих случаях (в отрезке 2-10).
- Задача 2. Примените метод SVM (например, из библиотеки sklearn) для датасета blobs2. Визуализируйте результат (разбиение плоскости и опорные вектора) при разных вариантах ядер (линейное; полиномиальное степеней 2,3,5; RBF).

### **Контрольная работа №3**

- Задача 1. Реализуйте алгоритм логистической регрессии со стохастическим градиентным спуском, обучите его на датасете spambase\_old (train) и проверьте на датасете spambase\_new (val). Получите ROC кривые для вариантов без нормировки и с нормировкой признаков.
- Задача 2. Модифицируйте модель из задачи 3, заменив последний нейрон на 10 нейронов, и реализовав мультиклассовую классификацию с softmax в качестве решающей функции и кросс-энтропией в качестве функции потерь и обучите на подготовленном датасете mnist.

### **Перечень вопросов, выносимых на промежуточную аттестацию (Зачет)**

1. Препроцессинг. Масштабирование. Нормировка. Полиномиальные признаки. One-hot encoding.
2. Кластеризация. kMeans, MeanShift, DBSCAN, Affinity Propagation.
3. Смещение и дисперсия (bias and variance). Понятие средней гипотезы.
4. Ансамблевые методы. Soft and Hard Voting. Bagging. Случайные леса. AdaBoost.
5. Типы обучения: с учителем, без учителя, с подкреплением, с частичным участием учителя, активное обучение.
6. Бустинг деревьев решений.
7. Ошибка внутри и вне выборки. Ошибка обобщения. Неравенство Хёфдинга. Валидация и кросс-валидация.

8. Линейная регрессия. Полиномиальная регрессия. Гребневая регрессия.
9. Размерность Вапника-Червоненкиса. Размерность Вапника-Червоненкиса для перцептрона.
10. Логистическая регрессия. Градиентный спуск.
11. Пороговые условия. Эффективность по Парето. Precision-Recall и ROC кривые. AUC.
12. Ансамблевые методы регрессии. RANSAC. Theil-Sen. Huber.
13. Перцептрон. Перцептрон с карманом.
14. Метод опорных векторов. Постановка задачи. Формулировка и решение двойственной задачи. Типы опорных векторов. Ядра.
15. Гипотезы и дихотомии. Функция роста. Точка поломки. Доказательство полиномиальности функции роста в присутствии точки поломки.
16. Деревья решений. Информационный выигрыш, критерий Джини. Регуляризация деревьев. Небрежные решающие деревья.
17. Байесовский классификатор. Типы оценки распределений признаков (Gaussian, Bernoulli, Multinomial). EM алгоритм.
18. Нейронные сети. Перцептрон Розенблатта. Функции активации. Обратное распространение градиента. Softmax.
19. Стохастическая оптимизация. Hill Climb. Отжиг. Генетический алгоритм.
20. Метрические классификаторы. kNN. WkNN. Отбор эталонов. DROP5. Kdtree

## 6.2. Описание показателей и критериев контроля успеваемости, описание шкал оценивания

### Балльно-рейтинговая система оценки

Объем рейтинга составляет: за текущий контроль - 30% от нормативного рейтинга дисциплины, за рубежный контроль - 30% от нормативного рейтинга дисциплины и за итоговый контроль - 40% от нормативного рейтинга дисциплины.

Таблица 7

#### Система рейтинговой оценки

Оценочные средства	Баллы			
	Устный опрос	0	2	4
Контрольная работа	0-4	5-6	7-8	9-10
Зачет	0-8	9-13	14-17	18-20
<b>Оценка</b>	<b>Неуд.</b>	<b>Удовл.</b>	<b>Хорошо</b>	<b>Отлично</b>
Посещение лекций и практических занятий				
Посещаемость	≤85%	86-88%	89-91%	92-100%
Баллы	0	10	20	30



Посещаемость рассчитывается, как отношение числа пропущенных занятий к общему числу занятий.

### **Максимальное число баллов – 100**

Для допуска к сдаче зачета с оценкой по дисциплине необходимо:

- фактический рейтинг семестрового контроля должен составлять более 50% от нормативного рейтинга семестрового контроля для дисциплины ( $R_{\text{факт.сем}} > 50\%R_{\text{норм семестр}}$ ), т.е. должен быть достигнут пороговый рейтинг;
- должен быть выполнен объем аудиторных занятий (включая посещение лекций), предусмотренный учебным планом.

### **Рейтинговый балл, выставляемый студенту**

Рейтинговый балл (в % от макс. балла за дисциплину)	Оценка по традиционной шкале
85,1-100%	Отлично
65,1 – 85 %	Хорошо
60,1 – 65 %	Удовлетворительно
Менее 60 %	Неудовлетворительно

## **7. Учебно-методическое и информационное обеспечение дисциплины**

### **7.1 Основная литература**

1. Никитина, Т. П. Программирование. Основы Python / Т. П. Никитина, Л. В. Королев. — Санкт-Петербург : Лань, 2023. — 156 с. — ISBN 978-5-507-45283-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/302714>
2. Рагимханова, Г. С. Программирование на Python : учебное пособие / Г. С. Рагимханова. — Махачкала : ДГПУ, 2022. — 126 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/330071>

### **7.2 Дополнительная литература**

1. Статистика / К. Н. Горпинченко, Е. В. Кремьянская, А. М. Ляховецкий [и др.]. — Санкт-Петербург : Лань, 2023. — 156 с. — ISBN 978-5-507-46528-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/339743>.
2. Волкова, Н. А. Элементы математики и статистики / Н. А. Волкова, Н. Ю. Кропачева, Е. Г. Михайлова. — 2-е изд., стер. — Санкт-Петербург : Лань, 2023. — 128 с. — ISBN 978-5-507-46535-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/310265>
3. Язык программирования PYTHON : справочник / Д. М. БИЗЛИИ. - Киев : ДиаСофт, 2000. - 326 с.

## **8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)**

1. <http://www.scopus.com/> (Электронная база данных)
2. <http://elibrary.ru/> (Научная электронная библиотека elibrary)

## 9. Перечень программного обеспечения и информационных справочных систем

Таблица 8

### Перечень программного обеспечения

№ п/п	Наименование раздела учебной дисциплины	Наименование программы	Тип программы	Автор	Год разработки
1	Коммерческие программное обеспечение и информационно справочные системы не используются				

## 10. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Процесс изучения дисциплины обеспечен аудиторией, оборудованной персональными компьютерами, мультимедийными средствами для демонстрации презентаций и доступом к информационно-телекоммуникационной сети «Интернет».

Таблица 9

### Сведения об обеспеченности специализированными аудиториями, кабинетами, лабораториями

Наименование специальных* помещений и помещений для самостоятельной работы (№ учебного корпуса, № аудитории)	Оснащенность специальных помещений и помещений для самостоятельной работы**
1	2
Центральная научная библиотека имени Н.И. Железнова, Читальные залы библиотеки	Столы, стулья, учебная литература
Общежитие №5 Комната для самоподготовки	Столы, стулья

## 11. Методические рекомендации студентам по освоению дисциплины

Основной объем материала студенту необходимо освоить самостоятельно в соответствии с темами для самостоятельной подготовки из таблицы 5. Студенты должны соблюдать дисциплину, вовремя приходить на занятия, представлять на проверку домашнюю работу, готовиться к проверочным и контрольным работам, предусмотренным курсом, проявлять активность на занятиях. Важное место в образовательном процессе занимает самостоятельная работа студентов. Для организации самостоятельной работы студентов по курсу используются современные информационные технологии: размещенные в сетевом доступе комплексы учебных и учебно-методических материалов (программа, список рекомендуемой литературы и информационных ресурсов, задания для самоконтроля), свободный доступ к сети «Интернет» для работы с молекулярными базами данных.

## Виды и формы отработки пропущенных занятий

Студент, пропустивший занятия обязан предоставить и защитить реферат по пропущенной теме.

### 12. Методические рекомендации преподавателям по организации обучения по дисциплине

Педагог, проводящий занятия, должен обладать высокой квалификацией и опытом. Необходимо разбираться в нюансах работы, чтобы при необходимости была возможность исправить ошибку студента. Для успешного освоения предмета необходимо периодически организовывать обсуждения и дискуссии по темам дисциплины.

Все практические работы носят строго профессиональный характер. Навыки, полученные при выполнении этих работ, пригодятся студенту на всех этапах обучения, при подготовке выпускной работы магистра и в профессиональной деятельности.

При преподавании курса необходимо ориентироваться на современные образовательные технологии путем использования группового способа обучения на практических занятиях, разбора конкретных ситуаций и интерактивного обсуждения результатов исследовательских учебных работ. Реализация современного подхода должна обеспечиваться широким использованием активных интерактивных форм проведения занятий, посещение профильных научно-исследовательских учреждений и повысить интерес к изучению дисциплины. Задачей преподавателя является приведение максимального количества позитивных примеров учреждений и специалистов добившихся высоких результатов в своих отраслях биотехнологии, для стимулирования интереса студентов к углубленному изучению данных дисциплин.

#### Программу разработал (и):

Лисовая Д.Д., ассистент

Монахос С.Г., д.с.-х.н., профессор

  
\_\_\_\_\_  
(подпись)

  
\_\_\_\_\_  
(подпись)

## РЕЦЕНЗИЯ

на рабочую программу дисциплины «Основы машинного обучения в биологии» ОПОП ВО по направлению 35.03.05 Садоводство, направленность «Селекция, генетика и биотехнология садовых культур» (квалификация выпускника – бакалавр).

Монахосом Григорием Федоровичем, генеральным директором ООО «Селекционная станция им. Н.Н.Тимофеева», кандидатом сельскохозяйственных наук, старшим научным сотрудником (далее по тексту рецензент), проведена рецензия рабочей программы дисциплины «Основы машинного обучения в биологии» ОПОП ВО по направлению 35.03.05 Садоводство, направленность «Селекция, генетика и биотехнология садовых культур» (магистратура) разработанной в ФГБОУ ВО «Российский государственный аграрный университет – МСХА имени К.А. Тимирязева», на кафедре ботаники, селекции и семеноводства садовых растений (*разработчики – Лисовая Дарья Дмитриевна, ассистент, Монахос Сократ Григорьевич, заведующий кафедрой, д. с.-х.н., профессор*).

Рассмотрев представленные на рецензию материалы, рецензент пришел к следующим выводам:

1. Предъявленная рабочая программа дисциплины «Основы машинного обучения в биологии» (далее по тексту Программа) соответствует требованиям ФГОС ВО по направлению 35.03.05 Садоводство. Программа содержит все основные разделы, соответствует требованиям к нормативно-методическим документам.

2. Представленная в Программе актуальность учебной дисциплины в рамках реализации ОПОП ВО не подлежит сомнению – дисциплина относится к части, формируемой участниками образовательных отношений – Б1.

3. Представленные в Программе цели дисциплины соответствуют требованиям ФГОС ВО направления 35.03.05 Садоводство.

4. В соответствии с Программой за дисциплиной «Основы машинного обучения в биологии» закреплена **2 компетенции**. Дисциплина «Основы машинного обучения в биологии» и представленная Программа способна реализовать их в объявленных требованиях.

**Результаты обучения**, представленные в Программе в категориях знать, уметь, владеть соответствуют специфике и содержанию дисциплины и демонстрируют возможность получения заявленных результатов.

5. Общая трудоёмкость дисциплины «Основы машинного обучения в биологии» составляет **4 зачётных единицы (144 часов)**.

6. Информация о взаимосвязи изучаемых дисциплин и вопросам исключения дублирования в содержании дисциплин соответствует действительности. Дисциплина «Основы машинного обучения в биологии» взаимосвязана с другими дисциплинами ОПОП ВО и Учебного плана по направлению 35.03.05 Садоводство и возможность дублирования в содержании отсутствует.

7. Представленная Программа предполагает использование современных образовательных технологий, используемые при реализации различных видов учебной работы. Формы образовательных технологий соответствуют специфике дисциплины.

8. Программа дисциплины «Основы машинного обучения в биологии» предполагает **3 часа** занятий в интерактивной форме.

9. Виды, содержание и трудоёмкость самостоятельной работы студентов, представленные в Программе, соответствуют требованиям к подготовке выпускников, содержащимся во ФГОС ВО направления 35.03.05 Садоводство.

10. Представленные и описанные в Программе формы текущей оценки знаний (опрос, как в форме обсуждения отдельных вопросов, так и выступления и участие в дискуссиях, диспутах и аудиторных заданиях), соответствуют специфике дисциплины и требованиям к выпускникам.

Форма промежуточного контроля знаний студентов, предусмотренная Программой, осуществляется в форме **Зачета**, что соответствует статусу дисциплины, как дисциплины учебного цикла – Б1 ФГОС ВО направления 35.03.05 Садоводство.

11. Формы оценки знаний, представленные в Программе, соответствуют специфике дисциплины и требованиям к выпускникам.

12. Учебно-методическое обеспечение дисциплины представлено: основной литературой – **1 источника** (базовый учебник), дополнительной литературой – **2 наименования**, Интернет-ресурсы – **2 источника** и соответствует требованиям ФГОС ВО направления 35.03.05 Садоводство.

13. Материально-техническое обеспечение дисциплины соответствует специфике дисциплины «Основы машинного обучения в биологии» и обеспечивает использование современных образовательных, в том числе интерактивных методов обучения.

14. Методические рекомендации студентам и методические рекомендации преподавателям по организации обучения по дисциплине дают представление о специфике обучения по дисциплине «Основы машинного обучения в биологии».

#### **ОБЩИЕ ВЫВОДЫ**

На основании проведенной рецензии можно сделать заключение, что характер, структура и содержание рабочей программы дисциплины «Основы машинного обучения в биологии» ОПОП ВО по направлению 35.03.05 Садоводство, направленность «Основы машинного обучения в биологии» (квалификация выпускника – магистр), разработанная Лисовой Дарьей Дмитриевной, ассистентом и Монахосом Сократом Григорьевичем, заведующим кафедрой, д.с.-х.н., профессором соответствует требованиям ФГОС ВО, современным требованиям экономики, рынка труда и позволит при её реализации успешно обеспечить формирование заявленных компетенций.

Рецензент: Монахос Григорий Федорович, генеральный директор ООО «Селекционная станция им. Н.Н.Тимофеева», кандидат сельскохозяйственных наук, старший научный сотрудник \_\_\_\_\_ «20» августа 2023 г.  
(подпись)

УТВЕРЖДАЮ:

Директор института (наименование)  
/Декан факультета (наименование)

«\_\_» \_\_\_\_\_ 201\_\_ г.

**Лист актуализации рабочей программы дисциплины<sup>1</sup>**

«\_\_\_\_\_»  
индекс по учебному плану, наименование

для подготовки бакалавров/ специалистов/ бакалавров

Направление: {шифр – название} \_\_\_\_\_

Направленность: \_\_\_\_\_

Форма обучения \_\_\_\_\_

Год начала подготовки: \_\_\_\_\_

Курс \_\_\_\_\_

Семестр \_\_\_\_\_

<sup>2</sup>а) В рабочую программу не вносятся изменения. Программа актуализирована для 20\_\_ г. начала подготовки.

б) В рабочую программу вносятся следующие изменения (указать на какой год начала подготовки):

- 1) .....
- 2) .....
- 3) .....

Разработчик (и): \_\_\_\_\_  
(ФИО, ученая степень, ученое звание) «\_\_» \_\_\_\_\_ 201\_\_ г.

Рабочая программа пересмотрена и одобрена на заседании кафедры \_\_\_\_\_  
\_\_\_\_\_ протокол № \_\_\_\_\_ от «\_\_» \_\_\_\_\_ 201\_\_ г.

Заведующий кафедрой \_\_\_\_\_

**Лист актуализации принят на хранение:**

Заведующий выпускающей кафедрой (наименование) \_\_\_\_\_ «\_\_» \_\_\_\_\_ 201\_\_ г.

Методический отдел УМУ: \_\_\_\_\_ «\_\_» \_\_\_\_\_ 201\_\_ г.

<sup>1</sup> Рабочая программа дисциплины актуализируется ежегодно перед началом нового учебного года.

<sup>2</sup> Разработчик выбирает один из представленных вариантов.