

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Астраханский государственный университет имени В. Н. Татищева»
(Астраханский государственный университет им. В. Н. Татищева)

СОГЛАСОВАНО
Руководитель ОПОП

Р.Ю. Демина
«05» мая 2025 г.

УТВЕРЖДАЮ
И.о. заведующего кафедрой
информационной безопасности

В.А. Черкасова
«05» мая 2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Введение в методы искусственного интеллекта

наименование

Составитель(-и)	Мартьянова А.Е., доцент, к.т.н., доцент кафедры информационной безопасности
Направление подготовки	10.03.01 Информационная безопасность
Направленность (профиль) ОПОП	«Организация и технологии защиты информации» (в сфере информационных и коммуникационных технологий)
Квалификация (степень)	бакалавр
Форма обучения	очно-заочная
Год приема	2023
Курс	4
Семестр	7

Астрахань, 2025

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

1.1. Целью дисциплины (модуля) Введение в методы искусственного интеллекта является формирование у студентов практических навыков в области систем искусственного интеллекта и принятия решений, изучение технологий, используемых при конструировании интеллектуальных систем в области информационной безопасности.

1.2. Задачи курса (модуля) Введение в методы искусственного интеллекта:

- ознакомление с приемами практического применения методов искусственного интеллекта;
- изучение принципов построения интеллектуальных систем;
- получение начальных навыков использования технологий искусственного интеллекта;
- ознакомление с программным обеспечением, используемым для построения интеллектуальных систем.

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОПОП

2.1. Учебная дисциплина (модуль) Б1.Б.09.01 «Введение в методы искусственного интеллекта» входит в обязательную часть учебного плана приема 2023 года направления подготовки 10.03.01 Информационная безопасность.

2.2. Для изучения данной учебной дисциплины (модуля) необходимы следующие знания, умения, навыки и (или) опыт деятельности, формируемые предшествующими учебными дисциплинами:

Информатика.

Основы программирования.

Знания:

- основных принципов алгоритмизации; основные методы обработки данных;
- этапов разработки программ и методы автоматизации программирования;
- основных понятий и методов технологий программирования;
- конструкций языка высокого уровня;

Умения:

- самостоятельно работать на ПЭВМ с соблюдением основных принципов работы;
 - осуществлять декомпозицию решения задачи и составлять алгоритмы отдельных его частей в соответствии с современной технологией программирования;
 - применять основные операторы, общие для всех языков программирования;
 - использовать отладчик как средство изучения и тестирования программ;
- работать с ресурсами компьютера программными средствами;

Навыки:

- навыками разработки и отладки программ на языках высокого уровня, навыками оптимизации программного кода.

2.3 Последующие учебные дисциплины (модули) и (или) практики, для которых необходимы знания, умения, навыки, формируемые данной учебной дисциплиной (модулем):

Дисциплина «Введение в методы искусственного интеллекта» поможет студентам при освоении дисциплины «Системы искусственного интеллекта в информационной безопасности», а также реализации задач преддипломной практики и написанию бакалаврской работы.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Процесс изучения дисциплины (модуля) направлен на формирование элементов следующих компетенций в соответствии с ФГОС ВО и ОПОП ВО по данному направлению подготовки (специальности):

обще профессиональных (ОПК):

ОПК-3. Способен использовать необходимые математические методы для решения задач профессиональной деятельности

ОПК-7. Способен использовать языки программирования и технологии разработки программных средств для решения задач профессиональной деятельности.

Таблица 1 – Декомпозиция результатов обучения

Код и наименование компетенции	Планируемые результаты обучения по дисциплине (модулю)		
	Знать (1)	Уметь (2)	Владеть (3)
ОПК-3. Способен использовать необходимые математические методы для решения задач профессиональной деятельности	ИОПК-3.1. Знать: основы математики, основные математические методы.	ИОПК-3.2. Уметь: решать стандартные профессиональные задачи с применением методов математического анализа и моделирования.	ИОПК-3.3. Владеть: навыками математического исследования объектов профессиональной деятельности.
ОПК-7. Способен использовать языки программирования и технологии разработки программных средств для решения задач профессиональной деятельности	ИОПК-7.1. Знать: основы программирования.	ИОПК-7.2. Уметь: использовать языки программирования и технологии разработки программных средств для решения задач профессиональной деятельности.	ИОПК-7.3. Владеть: навыками программирования для решения задач профессиональной деятельности.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Общая трудоемкость дисциплины (модуля) составляет 4 зачетные единицы, 144 часа. на контактную работу обучающихся с преподавателем (по видам учебных занятий) выделено 36 часов. Лекции – 18 часов, лабораторный занятия – 18 часов, самостоятельная работа – 108 часов.

Таблица 2 – Структура и содержание дисциплины (модуля)

№ п/п	Наименование раздела (темы)	Семестр	Неделя семестра	Контактная работа (в часах)	Самостоят. работа	Формы текущего контроля успеваемости (по неделям семестра) Форма

				Л	ПЗ	ЛР	КР	СР	промежуточной аттестации (по семестрам)
1	Раздел 1. Введение в интеллектуальные информационные системы	7	1-3	4		4		18	Отчет по лабораторной работе 1.
2	Раздел 2. Линейные модели. Классификация	7	4-6	4		2		26	Отчет по лабораторной работе 2.
3	Раздел 3. Кластеризация	7	7-9	2		4		20	Отчет по лабораторной работе 3.
4	Раздел 4. Искусственные нейронные сети	7	10-12	4		4		18	Отчет по лабораторной работе 4.
5	Раздел 5. Методы и подходы продвинутого машинного обучения	7	13-18	4		4		26	Отчет по лабораторной работе 5. Тест
7	ИТОГО		144	18		18		108	дифзачет

Примечание: Л – лекция; ПЗ – практическое занятие, семинар; ЛР – лабораторная работа; КР – курсовая работа; СР – самостоятельная работа.

Таблица 3 – Матрица соотнесения разделов, тем учебной дисциплины (модуля) и формируемых компетенций

Раздел, тема дисциплины (модуля)	Кол-во часов	Код компетенции		Общее количество компетенций
		ОПК-3	ОПК-7	
Раздел 1. Введение в интеллектуальные информационные системы	26	+	+	2
Раздел 2. Линейные модели. Классификация	32	+	+	2
Раздел 3. Кластеризация	26	+	+	2
Раздел 4. Искусственные нейронные сети	26	+	+	2
Раздел 5. Методы и подходы продвинутого машинного обучения	34	+	+	2
Итого	144			

Краткое содержание каждой темы дисциплины (модуля)

Раздел 1. Введение в интеллектуальные информационные системы

Общие понятия. Задачи машинного обучения. Модель и процесс машинного обучения. Введение в Python для анализа данных (NumPy, Pandas, Matplotlib, Seaborn, SciPy, Scikit-learn, TensorFlow).

Раздел 2. Линейные модели. Классификация

Линейная регрессия. Логистическая регрессия. Постановка задачи классификации. Оценки эффективности моделей классификации. Методы решения задачи классификации. Деревья решений и случайный лес. Теорема Байеса. Наивный Байесовский классификатор.

Раздел 3. Кластеризация

Постановка задачи кластерного анализа. Меры расстояний в кластеризации. Методы кластерного анализа.

Раздел 4. Искусственные нейронные сети

Модель искусственного нейрона. Нейронная сеть. Типы нейронных сетей. Архитектуры нейронных сетей.

Раздел 5. Методы и подходы продвинутого машинного обучения

Методы и подходы продвинутого машинного обучения. Практическое применение. Системы компьютерного зрения. Системы распознавания речи. Системы интеллектуального анализа текста.

5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРЕПОДАВАНИЮ И ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1. Указания для преподавателей по организации и проведению учебных занятий по дисциплине (модулю)

При подготовке к лекционным и практическим занятиям необходимо воспользоваться учебно-методической литературой из п.8. Лекции необходимо проводить с использованием презентаций, созданных в прикладном пакете Microsoft Office PowerPoint.

При подготовке к лабораторным занятиям необходимо воспользоваться учебно-методической литературой (основной и дополнительной) и информационно-справочными ресурсами из п.8.

Тестирование допускается проводить в бумажном или электронном виде на специализированных образовательных площадках (Moodle).

5.2. Указания для обучающихся по освоению дисциплины (модулю)

Таблица 4 – Содержание самостоятельной работы обучающихся

Номер раздела (темы)	Темы/вопросы, выносимые на самостоятельное изучение	Кол-во часов	Формы работы
1	Раздел 1. Введение в интеллектуальные информационные системы. Выполнение лабораторной работы 1	18	Внеаудиторная, изучение учебных пособий
2	Раздел 2. Линейные модели. Классификация. Выполнение лабораторной работы 2	26	Внеаудиторная, изучение учебных пособий
3	Раздел 3. Кластеризация. Выполнение лабораторной работы 3	20	Внеаудиторная, изучение учебных пособий
4	Раздел 4. Искусственные нейронные сети. Выполнение лабораторной работы 4	18	Внеаудиторная, изучение учебных пособий
5	Раздел 5. Методы и подходы продвинутого машинного обучения. Выполнение лабораторной работы 5.	26	Внеаудиторная, изучение учебных пособий

	Подготовка к тестированию		
--	---------------------------	--	--

5.3. Виды и формы письменных работ, предусмотренных при освоении дисциплины, выполняемые обучающимися самостоятельно – не предусмотрены.

6. ОБРАЗОВАТЕЛЬНЫЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

При реализации различных видов учебной работы по дисциплине могут использоваться электронное обучение и дистанционные образовательные технологии.

6.1. Образовательные технологии

Таблица 5 – Образовательные технологии, используемые при реализации учебных занятий

Раздел, тема дисциплины (модуля)	Форма учебного занятия		
	Лекция	Практическое занятие, семинар	Лабораторная работа
Раздел 1. Введение в интеллектуальные информационные системы	обзорная лекция	Не предусмотрено	выполнение лабораторной работы
Раздел 2. Линейные модели. Классификация	лекция - презентация	Не предусмотрено	выполнение лабораторной работы
Раздел 3. Кластеризация	лекция - презентация	Не предусмотрено	выполнение лабораторной работы
Раздел 4. Искусственные нейронные сети	лекция - презентация	Не предусмотрено	выполнение лабораторной работы
Раздел 5. Методы и подходы продвинутого машинного обучения	лекция - презентация	Не предусмотрено	выполнение лабораторной работы

Учебные занятия по дисциплине могут проводиться с применением информационно-телекоммуникационных сетей при опосредованном (на расстоянии) интерактивном взаимодействии обучающихся и преподавателя в режимах on-line в формах: видеолекций, лекций-презентаций, видеоконференции, собеседования в режиме чат, форума, чата, выполнения виртуальных практических и/или лабораторных работ и др.

Максимальный объем занятий обучающегося с применением электронных образовательных технологий не должен превышать 25%.

6.2. Информационные технологии

При организации учебной и внеучебной работы используются возможности сети Интернет, учебные пособия и литература в электронном виде, презентации. Отправка отчетов и рефератов на проверку возможна на электронный адрес (kafedra_ib_agu@mail.ru).

При реализации различных видов учебной и внеучебной работы используются следующие информационные технологии:

- использование возможностей Интернета в учебном процессе (использование информационного сайта преподавателя (рассылка заданий, предоставление выполненных работ, ответы на вопросы, ознакомление учащихся с оценками и т.д.));

- использование электронных учебников и различных сайтов (например, электронные библиотеки, журналы и т.д.) как источников информации;
- использование возможностей электронной почты преподавателя;
- использование средств представления учебной информации (электронных учебных пособий и практикумов, применение новых технологий для проведения очных (традиционных) лекций и се-минаров с использованием презентаций и т.д.);
- использование интегрированных образовательных сред, где главной составляющей являются не только применяемые технологии, но и содержательная часть, т.е. информационные ресурсы (до-ступ к мировым информационным ресурсам, на базе которых строится учебный процесс);
- использование виртуальной обучающей среды (или системы управления обучением LMS Moodle «Электронное образование») или иных информационных систем, сервисов и мессенджеров

6.3. Программное обеспечение, современные профессиональные базы данных и информационные справочные системы

6.3.1. Программное обеспечение

Наименование программного обеспечения	Назначение
Adobe Reader	Программа для просмотра электронных документов
Платформа дистанционного обучения LMS Moodle	Виртуальная обучающая среда
Mozilla FireFox	Браузер
Microsoft Office 2013,	Офисная программа
Microsoft Office Project 2013, Microsoft Office Visio 2013	
7-zip	Архиватор
Microsoft Windows 7 Professional	Операционная система
Kaspersky Endpoint Security	Средство антивирусной защиты
Microsoft Visual Studio	Среда разработки
Python	Среда разработки
Weka	Среда моделирования

6.3.2. Современные профессиональные базы данных и информационные справочные системы

- Электронный каталог Научной библиотеки АГУ на базе MARK SQL НПО «Информ-систем»: <https://library.asu.edu.ru>.
- Электронный каталог «Научные журналы АГУ»: <http://journal.asu.edu.ru/>.
- Универсальная справочно-информационная полнотекстовая база данных периодических изданий ООО «ИВИС»: <http://dlib.eastview.com/>
- Электронно-библиотечная система eLibrary. <http://elibrary.ru>
- Справочная правовая система КонсультантПлюс: <http://www.consultant.ru>
- Информационно-правовое обеспечение «Система ГАРАНТ»: <http://garant-astrakhan.ru>

7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

7.1. Паспорт фонда оценочных средств

При проведении текущего контроля и промежуточной аттестации по дисциплине (модулю) «Системы искусственного интеллекта в информационной безопасности» проверяется сформированность у обучающихся компетенций, указанных в разделе 3 настоящей программы. Этапность формирования данных компетенций в процессе освоения образовательной программы определяется последовательным освоением дисциплин (модулей) и прохождением практик, а в процессе освоения дисциплины (модуля) – последовательным достижением результатов освоения содержательно связанных между собой разделов, тем.

Таблица 6 – Соответствие разделов, тем дисциплины (модуля), результатов обучения по дисциплине (модулю) и оценочных средств

Контролируемый раздел, тема дисциплины (модуля)	Код контролируемой компетенции	Наименование оценочного средства
Раздел 1. Введение в интеллектуальные информационные системы	ОПК-3, ОПК-7	Лабораторная работа 1
Раздел 2. Линейные модели. Классификация	ОПК-3, ОПК-7	Лабораторная работа 2
Раздел 3. Кластеризация	ОПК-3, ОПК-7	Лабораторная работа 3
Раздел 4. Искусственные нейронные сети	ОПК-3, ОПК-7	Лабораторная работа 4
Раздел 5. Методы и подходы продвинутого машинного обучения	ОПК-3, ОПК-7	Лабораторная работа 5

7.2. Описание показателей и критериев оценивания компетенций, описание шкал оценивания

Таблица 7 – Показатели оценивания результатов обучения в виде знаний

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует глубокое знание теоретического материала, умение обоснованно излагать свои мысли по обсуждаемым вопросам, способность полно, правильно и аргументированно отвечать на вопросы, приводить примеры
4 «хорошо»	демонстрирует знание теоретического материала, его последовательное изложение, способность приводить примеры, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует неполное, фрагментарное знание теоретического материала, требующее наводящих вопросов преподавателя, допускает существенные ошибки в его изложении, затрудняется в приведении примеров и формулировке выводов
2 «неудовлетворительно»	демонстрирует существенные пробелы в знании теоретического материала, не способен его изложить и ответить на наводящие вопросы преподавателя, не может привести примеры

Таблица 8 – Показатели оценивания результатов обучения в виде умений и владений

Шкала оценивания	Критерии оценивания
------------------	---------------------

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы
4 «хорошо»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует отдельные, несистематизированные навыки, испытывает затруднения и допускает ошибки при выполнении заданий, выполняет задание по подсказке преподавателя, затрудняется в формулировке выводов
2 «неудовлетворительно»	не способен правильно выполнить задания

7.3. Контрольные задания или иные материалы, необходимые для оценки результатов обучения по дисциплине (модулю)

Раздел 1. Введение в интеллектуальные информационные системы Лабораторная работа № 1. Линейная регрессия

Написать свой класс линейной регрессии и протестировать его для решения задачи регрессии.

```
class MyLinearRegression:
    def __init__(self):
        self.coef_ = None
        self.intercept_ = None
    def fit(self, X, y):
        """
```

функция обучения линейной регрессии. Аналог метода fit модели LinearRegression из sklearn.

функция принимает на вход обучающую выборку (X — матрица признаков, y — массив ответов, значений целевой переменной), и подбирает коэффициенты линейной регрессии.

входящие параметры:

X: матрица размера (n, f), где n — количество элементов датасета,

f — количество признаков

y: массив размера (n,), где n — количество элементов датасета

"""

```
# переведем X и y в формат numpy array,
```

```
# чтобы было проще работать
```

```
X = np.array(X)
```

```
y = np.array(y)
```

```
# реализуйте код получения коэффициентов линейной
```

```
# регрессии по заданной обучающей выборке X и y.
```

```
# инструкция по реализации дана выше. Не забудьте про
```

```
# вектор-столбец из единиц!
```

```
<ВАШ КОД ТУТ>
```

```
# список коэффициентов, которые модель поставила
```

```

# в соответствие признакам датасета.
self.coef_ = <ВАШ КОД ТУТ>
# коэффициент — свободный член.
self.intercept_ = <ВАШ КОД ТУТ>
def predict(self, X):
    """
    функция получения предсказания линейной регрессии по входящему массиву
    признаков X.
    входящие параметры:
        X: матрица размера (n, f), где n — количество элементов датасета,
        f — количество признаков
    """
    y_pred = <ВАШ КОД ТУТ>
    return y_pred

```

Раздел 2. Линейные модели. Классификация

Лабораторная работа № 2. Логистическая регрессия.

Задание № 1. Реализация логистической регрессии с регуляризацией. Вам предстоит применить L_2 - регуляризацию к алгоритму логистической регрессии.

```

def basic_term(self, X, y, logits):
    # Вычисляет градиент логистической функции потерь по весам алгоритма
    # (исключая регуляризационное слагаемое)
    # ВАШ КОД
    grad = #ВАШ КОД. Вычислите слагаемое от логистической функции потерь
    return grad
def regularization_term(self, weights):
    # Вычисляет регуляризационное слагаемое градиента функции потерь
    # (без домножения на константу регуляризации)
    grad = #ВАШ КОД
    return grad
def grad(self, X, y, logits, weights):
    # Принимает на вход X, y, logits и вычисляет градиент логистической
    # функции потерь (включая регуляризационное слагаемое).
    # ВАШ КОД. Вычислите basic_term и regularization_term.
    grad = #ВАШ КОД. Сложите две компоненты (не забудьте домножить на C)
    return grad

```

Задание № 2. Применение созданного класса к искусственным данным.

Создайте объект класса `MyL2LogisticRegression`, рассмотрев параметр регуляризации $C = 0.01$. Обучите модель с помощью метода `fit` с параметром `max_iter=1000`.

```

clf = # ВАШ КОД. Создайте модель
losses = # ВАШ КОД. Обучите модель
coef, intercept = clf.coef_, clf.intercept_
coef, intercept

```

Задание № 3. Предскажите значения целевой переменной, используя метод `predict`. Вычислите accuracy.

```

from sklearn.metrics import accuracy_score
predictions = #ВАШ КОД
score = #ВАШ КОД
print(f'Model accuracy = {score}')

```

Задание № 4. Прodelайте процедуру из заданий 2, 3 для константы регуляризации $C=2$ и вычислите accuracy, а также проанализируйте, как отличаются веса моделей и

разделяющие поверхности. Постройте графики сходимости функции потерь и попробуйте объяснить полученный эффект.

Задание № 5. Исследование регуляризации в sklearn.

Для каждого значения C вычислите `accuracy_score` на обучающем и тестовом множествах. Запишите результаты в массив `train_accuracies` и `test_accuracies`.

```
train_accuracies = []
```

```
test_accuracies = []
```

```
for C in C_grid:
```

```
    # ВАШ КОД. Заполните массивы train_accuracies и test_accuracies.
```

```
    # ИСПОЛЬЗУЙТЕ LogisticRegression из sklearn, а не вашу имплементацию!
```

Раздел 3. Кластеризация

Лабораторная работа № 3. Решающие деревья.

Вам предлагается решить задачу бинарной классификации, а именно построить алгоритм, определяющий превысит ли средний заработок человека порог \$50k. Каждый объект выборки — человек, для которого известны следующие признаки:

- age
 - workclass
 - fnlwgt
 - education
 - education-num
 - marital-status
 - occupation
 - relationship
 - race
 - sex
 - capital-gain
 - capital-loss
 - hours-per-week
- Первым делом вы произведете загрузку и обработку данных. В частности, вам необходимо будет закодировать категориальные признаки с помощью One-hot encoding.
 - Сначала мы построим для нашей задачи обычный случайный лес и измерим его качество. Мы подберем оптимальный гиперпараметр "глубина дерева" для случайного леса.
 - Далее мы обучим алгоритм градиентного бустинга с помощью библиотеки Catboost. Catboost --- это библиотека для градиентного бустинга, которая автоматически обрабатывает категориальные признаки. Поэтому для этого пункта вам нужно будет использовать не One-hot признаки, а изначальные категориальные признаки.

Раздел 4. Искусственные нейронные сети

Лабораторная работа № 4. Полносвязные и свёрточные нейронные сети.

В этом занятии вам предстоит потренироваться построению нейронных сетей с помощью библиотеки Pytorch.

Создайте тензоры с обучающими и тестовыми данными.

Вам необходимо написать модуль на PyTorch реализующий $\text{logits} = XW + b$, где W и b — параметры (`nn.Parameter`) модели. Иначе говоря, здесь мы реализуем своими руками

модуль nn.Linear (в этом пункте его использование запрещено). Инициализируйте веса нормальным распределением (torch.randn).

Реализация цикла обучения.

Реализуйте predict и посчитайте accuracy на test.

Создайте полносвязную нейронную сеть с помощью класса Sequential. Сеть состоит из:

- Уплощения матрицы в вектор (nn.Flatten);
- Двух скрытых слоёв из 128 нейронов с активацией nn.ELU;
- Выходного слоя с 10 нейронами.

Протестируйте разные функции активации.

- Построим график accuracy/epoch для каждой функции активации.
- Реализуйте LeNet
- Обучите CNN

Раздел 5. Методы и подходы продвинутого машинного обучения

1. Лабораторная работа

Лабораторная работа № 5. Классификация изображений.

Сегодня вам предстоит помочь телекомпании FOX в обработке их контента. Как вы знаете, сериал "Симпсоны" идет на телеэкранах более 25 лет, и за это время скопилось очень много видеоматериала. Персонажи менялись вместе с изменяющимися графическими технологиями, и Гомер Симпсон-2018 не очень похож на Гомера Симпсона-1989. В этом задании вам необходимо классифицировать персонажей, проживающих в Спрингфилде.

В нашем тесте будет 990 картинок, для которых вам будет необходимо предсказать класс. мы используем wrapper над датасетом для удобной работы. Вам стоит понимать, что происходит с LabelEncoder и с torch.Transformation.

ToTensor конвертирует PIL Image с параметрами в диапазоне [0, 255] (как все пиксели) в FloatTensor размера (C x H x W) [0,1], затем производится масштабирование:

$$input = \frac{input - \mu}{standard\ deviation},$$
 константы - средние и дисперсии по каналам на основе ImageNet

Стоит также отметить, что мы переопределяем метод __getitem__ для удобства работы с данной структурой данных. Также используется LabelEncoder для преобразования строковых меток классов в id и обратно. В описании датасета указано, что картинки разного размера, так как брались напрямую с видео, поэтому следуем привести их к одному размеру (это делает метод _prepare_sample)

```
class SimpsonsDataset(Dataset):
```

```
    """
    Датасет с картинками, который параллельно подгружает их из папок производит
    скалирование и превращение в торчевые тензоры
    """
```

```
    def __init__(self, files, mode):
        super().__init__()
        # список файлов для загрузки
        self.files = sorted(files)
        # режим работы
        self.mode = mode
        if self.mode not in DATA_MODES:
            print(f"{self.mode} is not correct; correct modes: {DATA_MODES}")
            raise NameError
```

```

self.len_ = len(self.files)
self.label_encoder = LabelEncoder()
if self.mode != 'test':
    self.labels = [path.parent.name for path in self.files]
    self.label_encoder.fit(self.labels)
    with open('label_encoder.pkl', 'wb') as le_dump_file:
        pickle.dump(self.label_encoder, le_dump_file)
def __len__(self):
    return self.len_
def load_sample(self, file):
    image = Image.open(file)
    image.load()
    return image
def __getitem__(self, index):
    # для преобразования изображений в тензоры PyTorch и нормализации входа
    transform = transforms.Compose([
        transforms.ToTensor(),
        transforms.Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225])
    ])
    x = self.load_sample(self.files[index])
    x = self._prepare_sample(x)
    x = np.array(x / 255, dtype='float32')
    x = transform(x)
    if self.mode == 'test':
        return x
    else:
        label = self.labels[index]
        label_id = self.label_encoder.transform([label])
        y = label_id.item()
        return x, y

def _prepare_sample(self, image):
    image = image.resize((RESCALE_SIZE, RESCALE_SIZE))
    return np.array(image)

```

Построение нейросети:

Данная архитектура будет очень простой и нужна для того, чтобы установить базовое понимание и получить простой сабмит на Kaggle.

Описание слоев:

1. размерность входа: 3x224x224 2.размерности после слоя: 8x111x111
2. 16x54x54
3. 32x26x26
4. 64x12x12
5. ВЫХОД: 96x5x5.

2. Итоговое тестирование

Пример итогового теста

Какой метод используется для предсказания категориальных переменных?

- 1) Линейная регрессия
- 2) Логистическая регрессия
- 3) Кластеризация

4) ARIMA

Что такое нормализация данных?

- 1) Процесс удаления выбросов
- 2) Процесс приведения данных к единой шкале
- 3) Процесс кластеризации данных
- 4) Процесс классификации данных

Какой метод используется для предсказания числовых значений?

- 1) Линейная регрессия
- 2) Логистическая регрессия
- 3) Кластеризация
- 4) ARIMA

Выберите верное утверждение:

- 1) Чем ближе к 1 индекс корреляции, тем выше качество модели множественной регрессии.
- 2) Чем ближе к 0 коэффициент детерминации, тем выше качество модели множественной регрессии.
- 3) Независимость остатков проверяется с помощью критерия Дарбина – Уотсона.
- 4) Качество регрессора характеризуется фактом обоснованной зависимости остатков от целевой переменной.

Поставьте в соответствие библиотеку и ее описание

1. Scikit-learn
 2. NumPy
 3. Pandas
 4. Seaborn
- a) библиотека для работы с многомерными массивами числовых данных и со сложными математическими операциями;
- b) библиотека визуализации, основанная на Matplotlib;
- c) опенсорсная библиотека машинного обучения Python, с широким спектром алгоритмов кластеризации, регрессии и классификации;
- d) это библиотека с открытым исходным кодом, которая предлагает широкий спектр инструментов для обработки и анализа данных. С ее помощью можно читать данные из широкого спектра источников, таких как CSV, базы данных SQL, файлы JSON и Excel.

Вопросы к дифзачету

1. Машинное обучение: задачи, этапы, обучение с учителем и без учителя
2. Модель и процесс машинного обучения
3. Сбор данных для обучения
4. Подготовка данных для моделирования
5. Визуализация данных
6. Линейная регрессия
7. Логистическая регрессия
8. Задачи классификации
9. Задачи кластеризации
10. Оценка модели
11. Оценка классификаторов
12. Оценка регрессоров
13. Оптимизация модели

14. Проектирование признаков
15. Деревья решений, случайный лес
16. Наивный байесовский классификатор
17. Бустинг
18. Метод k-ближайших соседей
19. Модель искусственного нейрона
20. Нейронная сеть. Типы нейронных сетей. Архитектуры нейронных сетей
21. Методы и подходы продвинутого машинного обучения. Практическое применение
22. Системы компьютерного зрения
23. Системы распознавания речи
24. Системы интеллектуального анализа текста

**Перечень вопросов и заданий,
выносимых на дифференцированный зачёт**

Таблица 9 – Примеры оценочных средств с ключами правильных ответов

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
ОПК-3. Способен использовать необходимые математические методы для решения задач профессиональной деятельности				
1.	Задание закрытого типа	<p>Укажите соответствие между мерами схожести и формулами по которым они вычисляются</p> <ol style="list-style-type: none"> 1) Евклидово расстояние 2) Квадрат евклидова расстояния 3) Расстояние городских кварталов (манхэттенское расстояние) 4) Расстояние Чебышева <p>a) $\rho(x, x') = \max(x_i - x'_i)$</p> <p>b) $\rho(x, x') = \sum_i^n (x_i - x'_i)^2$</p> <p>c) $\rho(x, x') = \sqrt{\sum_i^n (x_i - x'_i)^2}$</p> <p>d) $\rho(x, x') = \sum_i^n x_i - x'_i$</p>	<ol style="list-style-type: none"> 1) – c 2) – b 3) – d 4) – a 	2
2.		<p>В основе какого алгоритма классификации лежат положения из теории вероятности:</p> <ol style="list-style-type: none"> 1) Градиентный бустинг 2) Наивный байесовский 3) J48 4) Случайный лес 	2	2
3.		<p>Машинное обучение — это</p> <ol style="list-style-type: none"> 1) Подраздел технологий искусственного интеллекта, изучающий построения алгоритмов, способных 	1	2

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
		<p>обучаться</p> <p>2) Подраздел технологий искусственного интеллекта, изучающий принципы разработки робототехнических систем</p> <p>3) Подраздел технологий искусственного интеллекта, изучающий построения алгоритмов для управления машинами</p> <p>4) Подраздел технологий искусственного интеллекта, изучающий человеко-машинное взаимодействие</p>		
4.		<p>Спортсмен Зайцев пробежал некоторую дистанцию за 10,5 сек. и занял 1-е место. Здесь числовой признак:</p> <p>1) 10.5 (10,5 с.)</p> <p>2) 1 (1-е место)</p> <p>3) Оба</p>	1	2
5.		<p>Линейная регрессия относится к задачам:</p> <p>1) Обучения с подкреплением</p> <p>2) Обучения без учителя</p> <p>3) Обучения с учителем</p>	3	2
6.	Задание открытого типа	<p>Дайте определение понятию «ложноотрицательные результаты»</p>	<p>Результаты распознавания называются ложноположительными, если классификатор неверно утверждает, что объект не принадлежит к рассматриваемому классу.</p>	5
7.		<p>Дайте определение понятию «ложноположительные результаты»</p>	<p>Результаты распознавания называются ложноположительными, если классификатор неверно отнёс объект к рассматриваемому классу.</p>	5
8.		<p>Дайте определение понятию «регрессия».</p>	<p>Регрессия – задача предсказания фактического значения на основе заранее собранных</p>	8

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
			данных. Относится к типу «обучение с учителем».	
9.		Дайте определение понятию «классификация».	Классификация – это задача отнесения объекта к какому-либо из заранее определенных классов. Относится к типу «обучение с учителем».	8
10.		Дайте определение понятию «понижение размерности».	Понижение размерности данных - задача преобразования данных в форму, наиболее удобную для анализа и интерпретации. Относится к типу «обучение без учителя».	8
ОПК-7. Способен использовать языки программирования и технологии разработки программных средств для решения задач профессиональной деятельности				
11.	Задание закрытого типа	Что такое нормализация данных? 1. Усреднение данных 2. Преобразование категориальных признаков в численные 3. Преобразование численных признаков в категориальные 4. Подгонка под единую шкалу	3	3
12.		Укажите соответствие между типами входных/целевых признаков и диаграммой, которую целесообразно использовать для визуализации 1. Входной признак- категориальный, целевая переменная- категориальная 2. Входной признак- категориальный, Целевая переменная- числовая 3. Входной признак- числовой, Целевая переменная- категориальная 4. Входной признак- числовой, Целевая переменная- числовая а. Диаграмма рассеяния б. Диаграмма размаха в. График плотности г. Мозаичная диаграмма	1-d 2-с 3-b 4-a	3
13.		Если вам необходимо, рассортировать содержимое корзины с фруктами, то какую задачу вы будете решать? 1. Понижения размерности 2. Регрессии 3. Классификации	4	3

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
		4. Кластеризации		
14.		Для оценки эффективности регрессора применяют: 1. Точность 2. Верность 3. Долю истинно положительных результатов 4. Квадратный корень из среднеквадратичной ошибки 5. Частотой ошибки	3	3
15.		Какой алгоритм основан на гипотезе «Набор слабых обучающих алгоритмов способен создать сильный обучающий алгоритм»? 1. Бустинг 2. Случайный лес 3. Нейронные сети 4. Наивный Байес	1	3
16.	Задание открытого типа	Что делать в случае, если в обучающем множестве отсутствуют какие-либо данные	Существует несколько стратегий: <ul style="list-style-type: none"> • создать новую категорию для отсутствующих данных • удалить экземпляры с отсутствующими данными • подставить значение предшествующего экземпляра • заместить отсутствующее значение средним значением столбца • заместить с помощью модели МО 	5
17.		Что такое точность классификации?	Точность-это доля правильно распознанных экземпляров.	5
18.		Какие признаки называются категориальными?	Признаки называются категориальными, если их можно отнести к какой-либо группе, но при этом не важен порядок	5
19.		В чем выражается проблема переобучения?	Модель эффективно работает только с теми	5

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
			данными, на которых была обучена	
20.		В проверочном множестве 800 объектов: 300 объектов класса <i>a</i> , 500 объектов класса <i>b</i> . Правильно были распознаны 275 объектов класса <i>a</i> и 480 объектов класса <i>b</i> . Рассчитайте верность (accuracy) классификатора.	0,94	5

7.4. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине (модулю)

Отчет по лабораторной работе

Отчет по лабораторной работе представляется в электронном виде. Защита отчета проходит в форме доклада студента по выполненной работе и ответов на вопросы преподавателя. В случае, если оформление отчета и поведение студента во время защиты соответствуют указанным требованиям, студент получает максимальное количество баллов.

Основаниями для снижения количества баллов в диапазоне от max до min являются:

- отсутствие списка использованной литературы,
- небрежное выполнение,
- отсутствие выводов.

Отчет не может быть принят и подлежит доработке в случае:

- отсутствия необходимых разделов,
- отсутствия необходимого графического материала,
- неверных результатов расчета.

В отчете по выполненной лабораторной работе должны быть указаны:

- тема лабораторной работы,
- пакет документов в соответствии с темой лабораторной работы,
- использованная литература.

Критерии оценки лабораторной работы:

– оценка «отлично» выставляется обучающемуся, если студент продемонстрировал глубокие знания теоретического материала и умение их применять, обоснованно изложил свои мысли, сделал необходимые выводы;

– оценка «хорошо» выставляется обучающемуся, если студент продемонстрировал глубокие знания теоретического материала и умение их применять, обоснованно изложил свои мысли, сделал необходимые выводы, допущены некоторые неточности, имеется одна негрубая ошибка;

– оценка «удовлетворительно» выставляется обучающемуся, если студент ответил на вопросы преимущественно верно, имеются затруднения в формулировке выводов, имеются одна или две негрубые ошибки;

– оценка «неудовлетворительно» выставляется обучающемуся, если студент не дал ответы на поставленные вопросы, обоснования неверные, либо дан верный ответ без его обоснования, сделаны грубые ошибки, отсутствуют знания по основам математики.

Контрольные работы

Критерии оценки теста:

- оценка «отлично» выставляется студенту, если он умеет безошибочно самостоятельно обрабатывать и интерпретировать данные при решении задач, как в стандартной, так и в нестандартной формулировке;
- оценка «хорошо» выставляется студенту, если он умеет безошибочно самостоятельно обрабатывать и интерпретировать данные при решении задач в стандартной ситуации или за верное решение 75% - 89% заданий теста;
- оценка «удовлетворительно» выставляется студенту, если он умеет при решении задач обрабатывать данные с опорой на справочные материалы и помощь преподавателя, верно выполняя при этом 60% - 74% работы.
- оценка «неудовлетворительно» выставляется студенту, если он не умеет правильно обрабатывать данные, выполнил менее 60% заданий теста.
- оценка «зачтено» выставляется студенту, если тест студента оценен не ниже чем «удовлетворительно»;
- оценка «не зачтено», если тест оценен ниже чем «удовлетворительно».

Оценивание студентов на **зачете** осуществляется в соответствии с требованиями и критериями 100-балльной шкалы. Учитываются как результаты текущего контроля, так и знания, навыки и умения, непосредственно показанные студентами в ходе зачета.

Текущий контроль осуществляется в ходе учебного процесса и консультирования студентов, по результатам выполнения самостоятельных и тематических контрольных работ. Он предусматривает проверку готовности студентов к плановым занятиям, оценку качества и самостоятельности выполнения заданий на практических занятиях, проверку правильности решения задач, выданных на самостоятельную проработку.

Зачет

Оценивание студентов на **зачете** осуществляется в соответствии с требованиями и критериями 100-балльной шкалы. Учитываются как результаты текущего контроля, так и знания, навыки и умения, непосредственно показанные студентами в ходе зачета.

Текущий контроль осуществляется в ходе учебного процесса и консультирования студентов, по результатам выполнения самостоятельных и тематических контрольных работ. Он предусматривает проверку готовности студентов к плановым занятиям, оценку качества и самостоятельности выполнения заданий на практических занятиях, проверку правильности решения задач, выданных на самостоятельную проработку.

На зачете осуществляется комплексная проверка знаний, навыков и умений студентов по всему теоретическому материалу дисциплины и с проверкой практических навыков и умений по разработке документов различных видов. Теоретические знания оцениваются путем компьютерного тестирования или на основании письменных ответов студентов по нескольким теоретическим вопросам.

Таблица 10 – Технологическая карта рейтинговых баллов по дисциплине (модулю)

№ п/п	Контролируемые мероприятия	Количество мероприятий / баллы	Максимальное количество баллов	Срок представления
Основной блок				
1.	<i>Выполнение лабораторной работы</i>	5/15	75	По расписанию
2.	<i>Тест</i>	1/15	15	
Всего			90	-
Блок бонусов				
3.	<i>Посещение занятий без пропусков</i>	1	3	
4.	<i>Своевременное выполнение всех заданий</i>	1	3	

№ п/п	Контролируемые мероприятия	Количество мероприятий / баллы	Максимальное количество баллов	Срок представления
5.	<i>Активность студента на занятии</i>	1	4	
Всего			10	-
ИТОГО			100	-

Таблица 11 – Система штрафов (для одного занятия)

Показатель	Балл
<i>Опоздание на занятие</i>	- 1
<i>Нарушение учебной дисциплины</i>	- 1
<i>Неготовность к занятию</i>	- 2
<i>Пропуск занятия без уважительной причины</i>	- 2

Таблица 12 – Шкала перевода рейтинговых баллов в итоговую оценку за семестр по дисциплине (модулю)

Сумма баллов	Оценка по 4-балльной шкале	
90–100	5 (отлично)	зачтено
85–89	4 (хорошо)	
75–84		
70–74		
65–69	3 (удовлетворительно)	
60–64	2 (неудовлетворительно)	незачтено
Ниже 60		

При реализации дисциплины (модуля) в зависимости от уровня подготовленности обучающихся могут быть использованы иные формы, методы контроля и оценочные средства, исходя из конкретной ситуации.

8. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

8.1. Основная литература

1. Кольер, Р. Машинное обучение в Elastic Stack / Р. Кольер, К. Монтонен, Б. Азарми; пер. с англ. В. С. Яценкова. - Москва : ДМК Пресс, 2021. - 380 с. - ISBN 978-5-93700-107-8. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785937001078.html> (дата обращения: 21.12.2022). - Режим доступа : по подписке. Андреева Г. М. Социальная психология: учебник. М.: Аспект Пресс, 2002. 364 с. (23 экз.).

2. Машинное обучение с использованием библиотеки H2O / Д. Кук - Москва : ДМК Пресс, 2018. - ISBN 978-5-97060-508-0. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970605080.html> (дата обращения: 21.12.2022). - Режим доступа : по подписке.

3. Машинное обучение с использованием библиотеки H2O / Д. Кук - Москва : ДМК Пресс, 2018. - ISBN 978-5-97060-508-0. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970605080.html> (дата обращения: 21.12.2022). - Режим доступа : по подписке.

4. Python и машинное обучение / С. Рашка - Москва : ДМК Пресс, 2017. - ISBN 978-5-97060-409-0. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970604090.html> (дата обращения: 21.12.2022). - Режим доступа : по подписке.

5. Python и машинное обучение / С. Рашка - Москва : ДМК Пресс, 2017. - ISBN 978-5-97060-409-0. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970604090.html> (дата обращения: 21.12.2022). - Режим доступа : по подписке.

6. Горбаченко, В. И. Машинное обучение : учебное пособие / В. И. Горбаченко, К. Е. Савенков, М. А. Малахов. — Москва : Ай Пи Ар Медиа, 2023. — 217 с. — ISBN 978-5-4497-1860-0. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/125886.html> (дата обращения: 28.11.2022). — Режим доступа: для авторизир. пользователей. - DOI: <https://doi.org/10.23682/125886>

7. Павлова, А. И. Искусственные нейронные сети : учебное пособие / А. И. Павлова. — Москва : Ай Пи Ар Медиа, 2021. — 190 с. — ISBN 978-5-4497-1165-6. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/108228.html> (дата обращения: 21.12.2022). — Режим доступа: для авторизир. пользователей. - DOI: <https://doi.org/10.23682/108228>

8. Барский, А. Б. Введение в нейронные сети : учебное пособие / А. Б. Барский. — 3-е изд. — Москва, Саратов : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 357 с. — ISBN 978-5-4497-0309-5. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/89426.html> (дата обращения: 21.12.2022). — Режим доступа: для авторизир. пользователей

8.2. Дополнительная литература

1. Паттерсон, Дж. , Гибсон А. Глубокое обучение с точки зрения практика / Паттерсон Дж. , Гибсон А. , пер. с англ. А. А. Слинкина. - Москва : ДМК Пресс, 2018. - 418 с. - ISBN 978-5-97060-481-6. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970604816.html> (дата обращения: 21.12.2022). - Режим доступа : по подписке.

2. (Манро), Р. Машинное обучение с участием человека / Монарх Р. (Манро) ; перевод В. И. Бахур. — Москва : ДМК Пресс, 2022. — 498 с. — ISBN 978-5-97060-934-7. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/125122.html> (дата обращения: 20.10.2022). — Режим доступа: для авторизир. Пользователей

3. Яхьяева, Г. Э. Нечеткие множества и нейронные сети : учебное пособие / Г. Э. Яхьяева. — 4-е изд. — Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 315 с. — ISBN 978-5-4497-0665-2. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/97552.html> (дата обращения: 21.12.2022). — Режим доступа: для авторизир. Пользователей

4. Барский, А. Б. Искусственный интеллект и логические нейронные сети : учебное пособие / А. Б. Барский. — Санкт-Петербург : Интермедия, 2019. — 360 с. — ISBN 978-5-4383-0155-4. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/95270.html> (дата обращения: 21.12.2022). — Режим доступа: для авторизир. пользователей

8.3. Интернет-ресурсы, необходимые для освоения дисциплины (модуля)

1. **Электронно-библиотечная система (ЭБС) ООО «Политехресурс» «Консультант студента».** Многопрофильный образовательный ресурс «Консультант студента» является электронной библиотечной системой, предоставляющей доступ через сеть Интернет к учебной литературе и дополнительным материалам, приобретенным на основании прямых договоров с правообладателями. Каталог в настоящее время содержит около 15000 наименований. www.studentlibrary.ru.

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

(МОДУЛЯ)

Для проведения лекционных занятий необходима мультимедийная аудитория, оснащенная компьютерной презентационной техникой.

Для проведения лабораторных занятий необходима аудитория, оснащенная компьютерами.

Рабочая программа дисциплины (модуля) при необходимости может быть адаптирована для обучения (в том числе с применением дистанционных образовательных технологий) лиц с ограниченными возможностями здоровья, инвалидов. Для этого требуется заявление обучающихся, являющихся лицами с ограниченными возможностями здоровья, инвалидами, или их законных представителей и рекомендации психолого-медико-педагогической комиссии. Для инвалидов содержание рабочей программы дисциплины (модуля) может определяться также в соответствии с индивидуальной программой реабилитации инвалида (при наличии).