Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Гаранин Максим Алексеевич
Должность: Ректор
Дата подписания: 12.11.2026 12:01:46
Уникальный программный ключ:
7708e3a47e66a8ee02711b298d7c78bd1e40bf88

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«ПРИВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПУТЕЙ СООБЩЕНИЯ»

УТВЕРЖДЕНА
Ученым советом университета
(протокол от 24.02.2026 №15)

МОДУЛЬ "СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА"

Системы искусственного интеллекта

рабочая программа дисциплины (модуля)

Направление подготовки 38.03.01 Экономика
Направленность (профиль) Экономика и финансы предприятий (организаций)
Квалификация **бакалавр**

Форма обучения **очно-заочная**

Общая трудоемкость **3 ЗЕТ**

Виды контроля в семестрах:
зачет 4

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	**4 (2.2)**		Итого	
Недель	17 2/6			
Вид занятий	УП	РП	УП	РП
Лекции	8	8	8	8
Лабораторные	8	8	8	8
Практические	8	8	8	8
Конт. ч. на аттест. в период ЭС	0,15	0,15	0,15	0,15
Итого ауд.	24	24	24	24
Контактная работа	24,15	24,15	24,15	24,15
Сам. работа	75	75	75	75
Часы на контроль	8,85	8,85	8,85	8,85
Итого	108	108	108	108

Программу составил(и):
к.п.н., доцент, Тюжина И.В.

Рабочая программа дисциплины
Системы искусственного интеллекта

разработана в соответствии с ФГОС ВО:
Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 38.03.01 Экономика (приказ Минобрнауки России от 12.08.2020 г. № 954)

составлена на основании учебного плана: 38.03.01-26-3-ЭФПб-оз.plz.plx
Направление подготовки 38.03.01 Экономика Направленность (профиль) Экономика и финансы предприятий (организаций)

Рабочая программа одобрена на заседании кафедры
Цифровые технологии

Зав. кафедрой Ефимова Т.Б.

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)	
1.1	овладение студентами основными методами методов машинного обучения, получение навыков программирования алгоритмов в области искусственного интеллекта и анализа полученных результатов.

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	
Цикл (раздел) ОП:	Б1.О.21.02

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

ОПК-5 Способен использовать современные информационные технологии и программные средства при решении профессиональных задач

ОПК-5.2 Использует методы искусственного интеллекта (машинного обучения) и анализа больших данных для решения прикладных задач

В результате освоения дисциплины (модуля) обучающийся должен

3.1	**Знать:**
3.1.1	основные методы машинного обучения; классификацию задач машинного обучения; метрики качества модели; функции и методы библиотеки Pandas: query, tail, sample, head; методы фильтрации и агрегации данных; признаки переобученности модели; методы библиотек sklearn, seaborn, matplotlib и numpy.
3.2	**Уметь:**
3.2.1	агрегировать данные средствами Python; выполнять фильтрацию данных средствами Python; настраивать вес нейронна; выбирать тип классификатора в зависимости от поставленной задачи; визуализировать данные средствами библиотек pandas, seaborn, matplotlib;
3.2.2	выполнять многоклассовую классификацию методами библиотеки sklearn;
3.2.3	строить деревья решений и выполнять по ним предсказание.
3.3	**Владеть:**
3.3.1	построения дерева решений;
3.3.2	реализации алгоритма случайного леса (Random forest);
3.3.3	написания нейронных сетей;
3.3.4	обучения модели.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Примечание
	Раздел 1. Введение искусственный интеллект			
1.1	Искусственный интеллект. Основные понятия ИИ. Машинное обучение. Глубокое обучение. Обучение с учителем и без учителя. /Лек/	4	2	
1.2	Данные, модель, обработка данных. Методологические принципы анализы данных. Цели, этапы, методы и техники. Фильтрация группировка и агрегация данных. /Лек/	4	2	
1.3	Визуализация данных. Визуализация данных в Phyton. Визуализация в Pandas. Библиотеки Seaborn и Matplotlib. /Ср/	4	2	
1.4	Алгоритмы и структуры данных. Библиотека Pandas: структуры данных: Series и DataFrame. Работа с csv. /Пр/	4	2	
1.5	Фильтрация данных. Библиотека Pandas. Вывод строк и колонок. Методы query, tail, sample, head. Объединение методов. /Пр/	4	2	
1.6	Фильтрация данных. Библиотека Pandas. Вывод строк и колонок. Методы query, tail, sample, head. Объединение методов. /Лаб/	4	2	
1.7	Агрегация данных в Pandas: count, sum, max, min, mean, median. Методы groupby и aggregate. Объединение методов. /Ср/	4	2	
1.8	Агрегация данных в Pandas: count, sum, max, min, mean, median. Методы groupby и aggregate. Объединение методов. /Лаб/	4	2	
1.9	Визуализация данных в Phyton. Методы визуализации в Pandas. Библиотеки Seaborn и Matplotlib. Графики, тепловые карты, диаграммы рассеивания. Анализ и интерпретация результатов визуализации. /Ср/	4	2	
	Раздел 2. Методы машинного обучения			
2.1	Решающие деревья: основные понятия. Классификация. Деревья решений. Основные параметры дерева. Дерево решений в задачи регрессии. Метод ближайших соседей. /Лек/	4	2	

2.2	Концепция случайного леса. Случайная выборка тренировочных образцов. Усреднение прогнозов. Проблема переобучения. Метрики качества модели. Тестирование ROC AUC. /Лек/	4	2	
2.3	Нейронные сети. Искусственные нейроны. Перцептрон. Обучение нейрона. /Ср/	4	2	
2.4	Градиентный спуск. Дробление шага при градиентном спуске. Стохастический градиентный спуск. Метод наискорейшего спуска. /Ср/	4	2	
2.5	Нейронная сеть. Однослойная модель. Сверточные нейронные сети. Свёрточный слой, слой подвыборки, полносвязный слой. Целевая функция. /Ср/	4	2	
2.6	Решающие деревья: обучение модели. Библиотеки sklearn, pandas, numpy. Выбор параметров модели. Кросс-валидация. Предварительная обработка данных. Тренировочный набор данных. Тестовый набор данных. /Пр/	4	2	
2.7	Библиотеки sklearn, pandas, numpy. Выбор параметров модели. Кросс-валидация. Предварительная обработка данных. Тренировочный набор данных. Тестовый набор данных. /Лаб/	4	2	
2.8	Концепция случайного леса. Случайная выборка тренировочных образцов. Усреднение прогнозов. Проблема переобучения. Типы ошибок: true negative, false positive. Метрики качества: Precision, Recall, F1 score. Тестирование ROC AUC. /Пр/	4	2	
2.9	Распознавание рукописных цифр. Алгоритмы случайного леса. Оптимальные параметры дерева и леса. /Лаб/	4	2	
2.10	Искусственный нейрон. Вход, функция активации (сигмоида, гиперболический тангенс, SoftMax). Подбор весов. /Ср/	4	2	
2.11	Определение тональности текста с помощью нейронных сетей. Библиотеки tensorflow, Keras, набор данных IMDb. /Ср/	4	2	
2.12	Библиотека Keras. Полносвязный и сглаживающий слои, слой Dropout. Решение проблем переобучения. /Ср/	4	2	
2.13	Классификация изображений по нескольким классам (более двух). Набор данных Fashion Mnist. /Ср/	4	2	
2.14	Внешние источники данных для анализа. Kuggle. Подготовка данных. Работа с изображениями разного формата. /Ср/	4	2	
2.15	Распознавание образов. Классификация изображений кошек и собак. Библиотека keras , ImageDataGenerator. Достижение заданной точности модели на валидационной выборке. /Ср/	4	2	
	Раздел 3. Самостоятельная работа			
3.1	Написание собственных классов для описания нейронной сети. Нейрон. Слой. Сеть. Обучение модели. Выполнение предсказания. /Ср/	4	4	
3.2	Предсказание пола по росту и весу с помощью нейронной сети. Выполнение предсказания. Минимизация значения функции потерь. Обучение: стохастический градиентный спуск. /Ср/	4	4	
3.3	Задачи распознавания образов. Классификация объектов. Поиск изображения по образцу. /Ср/	4	3	
3.4	Алгоритм обратного распространения ошибки. Функции активации. Оценка работы сети. /Ср/	4	4	
3.5	Компьютерное зрение. Библиотека OpenCV. Сегментация изображений. Детектирование объектов. /Ср/	4	4	
3.6	Компьютерное зрение. Библиотека OpenCV. Отслеживание движущихся объектов во времени. Распознавание лиц. /Ср/	4	4	
3.7	Обработка естественного языка. Основные понятия. Токенезация. Лемматизация. /Ср/	4	4	
3.8	Обработка естественного языка. Парсинг зависимостей. Распознавание именованных сущностей /Ср/	4	4	
3.9	Подготовка к лекциям /Ср/	4	4	
3.10	Подготовка к лабораторным занятиям /Ср/	4	16	
	Раздел 4. Контактные часы на аттестацию			
4.1	Зачёт /КЭ/	4	0,15	
5. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ				

Оценочные материалы для проведения промежуточной аттестации обучающихся приведены в приложении к рабочей программе дисциплины.

Формы и виды текущего контроля по дисциплине (модулю), виды заданий, критерии их оценивания, распределение баллов по видам текущего контроля разрабатываются преподавателем дисциплины с учетом ее специфики и доводятся до сведения обучающихся на первом учебном занятии.

Текущий контроль успеваемости осуществляется преподавателем дисциплины (модуля), как правило, с использованием ЭИОС или путем проверки письменных работ, предусмотренных рабочими программами дисциплин в рамках контактной работы и самостоятельной работы обучающихся. Для фиксирования результатов текущего контроля может использоваться ЭИОС.

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

6.1. Рекомендуемая литература

6.1.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год	Эл. адрес
Л1.1	Никольский С. Н.	Автоматизация информационного поведения и искусственный интеллект: Учебное пособие	Москва: МИРЭА, 2020	https://e.lanbook.com/bo

6.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год	Эл. адрес
Л2.1	Железнов М. М.	Методы и технологии обработки больших данных: учебно-методическое пособие	Москва: Московский государственный строительный университет, 2020	https://e.lanbook.com/bo

6.2 Информационные технологии, используемые при осуществлении образовательного процесса по дисциплине (модулю)

6.2.1 Перечень лицензионного и свободно распространяемого программного обеспечения

6.2.1.1	Python

6.2.2 Перечень профессиональных баз данных и информационных справочных систем

6.2.2.1	Информационная справочная система Техэксперт https://tech.company-dis.ru
6.2.2.2	Информационная справочная система "Гарант" http://www.garant.ru
6.2.2.3	База данных Государственных стандартов http://gostexpert.ru/
6.2.2.4	База данных «Железнодорожные перевозки» https://cargo-report.info/
6.2.2.5	

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

7.1	Учебные аудитории для проведения занятий лекционного типа, укомплектованные специализированной мебелью и техническими средствами обучения: мультимедийное оборудование для предоставления учебной информации большой аудитории и/или звукоусиливающее оборудование (стационарное или переносное).
7.2	Учебные аудитории для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные специализированной мебелью и техническими средствами обучения: мультимедийное оборудование и/или звукоусиливающее оборудование (стационарное или переносное).
7.3	Помещения для самостоятельной работы, оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду университета.
7.4	Помещения для хранения и профилактического обслуживания учебного оборудования.
7.5	Лаборатории, оснащенные специальным лабораторным оборудованием: компьютерной техникой с установленным ПО: Python, а также с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду университета

Приложение
к рабочей программе дисциплины

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

(наименование дисциплины(модуля)

Направление подготовки / специальность

38.03.01 Экономика

(код и наименование)

Направленность (профиль)/специализация

Экономика и финансы предприятий (организаций)

(наименование)

Содержание

1. Пояснительная записка.
2. Типовые контрольные задания или иные материалы для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих уровень сформированности компетенций.
3. Методические материалы, определяющие процедуру и критерии оценивания сформированности компетенций при проведении промежуточной аттестации.

1. Пояснительная записка

Цель промежуточной аттестации – оценивание промежуточных и окончательных результатов обучения по дисциплине, обеспечивающих достижение планируемых результатов освоения образовательной программы.

Формы промежуточной аттестации: Зачёт, 4 семестр (очная форма обучения)
зачет, 4 семестр (очно-заочная форма обучения)

Перечень компетенций, формируемых в процессе освоения дисциплины

Код и наименование компетенции	Код индикатора достижения компетенции
ОПК-5: Способен использовать современные информационные технологии и программные средства при решении профессиональных задач.	ОПК-5.2: Использует методы искусственного интеллекта (машинного обучения) и анализа больших данных для решения прикладных задач

Результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Оценочные материалы
ОПК-5.2: Использует методы искусственного интеллекта (машинного обучения) и анализа больших данных для решения прикладных задач	Обучающийся **знает**: основные методы машинного обучения; классификацию задач машинного обучения; метрики качества модели; функции и методы библиотеки Pandas: query, tail, sample, head; методы фильтрации и агрегации данных; признаки переобученности модели; методы библиотек sklearn, seaborn, matplotlib и numpy.	Вопросы (№1 - №20)
	Обучающийся **умеет**: агрегировать данные средствами Python; выполнять фильтрацию данных средствами Python; настраивать вес нейронна; выбирать тип классификатора в зависимости от поставленной задачи; визуализировать данные средствами библиотек pandas, seaborn, matplotlib; выполнять многоклассовую классификацию методами библиотеки sklearn; строить деревья решений и выполнять по ним предсказание.	Задания (№21 - №25)
	Обучающийся **владеет**: Навыками построения дерева решений; реализации алгоритма случайного леса (Random forest); написания нейронных сетей; обучения модели.	Задания (№ 26- №30)

Промежуточная аттестация (зачет) проводится в одной из следующих форм:
1) собеседование;
2) выполнение заданий в ЭИОС Университета.

2. Типовые[1] контрольные задания или иные материалы для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих уровень сформированности компетенций

2.1. Типовые вопросы (тестовые задания) для оценки знаниевого образовательного результата

Проверяемый образовательный результат:

Код и наименование индикатора достижения компетенции	Образовательный результат
ОПК-5.2: Использует методы искусственного интеллекта (машинного обучения) и анализа больших данных для решения прикладных задач	Обучающийся **знает**: основные методы машинного обучения; классификацию задач машинного обучения; метрики качества модели; функции и методы библиотеки Pandas: query, tail, sample, head; методы фильтрации и агрегации данных; признаки переобученности модели; методы библиотек sklearn, seaborn, matplotlib и numpy.

Задание 1
Машинное обучение —
a) специализированное программное решение (или набор решений), которое включает в себя все инструменты для извлечения закономерностей из сырых данных
b) эта группировка объектов (Наблюдений, событий) на основеданных, описывающих свойства объектов
c) набор данных, каждая запись которого представляет собой учебный пример, содержащий заданный входной набор данных, и соответствующий ему правильный выходной результат
d) подразделение искусственного интеллекта изучающий методы построения алгоритмов, способных обучаться на данных

Задание 2
Среди предложенных задач машинного обучения укажите задачи регрессии
a) Поиск негативных отзывов на фильм на сайте кинокомпании
b) Алгоритм фильтрации спама
c) Предсказание срока окупаемости проекта
d) Предсказание рыночной стоимости квартиры
e) Поиск мошеннических транзакций

Задание 3
К библиотекам анализа данных в Python относятся …
a) Tkinter
b) Pandas
c) NumPy
d) Matplotlib

Задание 4
Какие из перечисленных команд вернут данные для России и Зимбабве?
a) res.query('country == ["Russia","Zimbabwe"]')
b) res.query('country == "Russia" | "Zimbabwe"')
c) res.query('country == "Russia" | country == "Zimbabwe"')
d) res.query('country == "Russia" & country == "Zimbabwe"')
e) res[res.country == "Russia" | res.country == "Zimbabwe"]
f) res[(res.country == "Russia") | (res.country == "Zimbabwe")]

Задание 5
Отметьте верные утверждения об алгоритме случайного леса - Random Forest
a) Итоговым предсказанием модели является предсказание случайного дерева
b) Параметры для каждого дерева (глубина, минимальное число образцов в листе и т.д.) выбираются случайно
c) Каждое дерево в лесу получает случайный поднабор данных
d) Число деревьев в лесу выбирается случайным образом
e) Предсказание модели - усреднённые предсказания деревьев

Задание 6
Какой атрибут отвечает за минимальное число образцов в листе?
a) min_samples_leaf

[1] Приводятся типовые вопросы и задания. Оценочные средства, предназначенные для проведения аттестационного мероприятия, хранятся на кафедре в достаточном для проведения оценочных процедур количестве вариантов. Оценочные средства подлежат актуализации с учетом развития науки, образования, культуры, экономики, техники, технологий и социальной сферы. Ответственность за нераспространение содержания оценочных средств среди обучающихся университета несут заведующий кафедрой и преподаватель – разработчик оценочных средств.

b) max_depth
c) min_samples_split
d) min_impurity_decrease

Задание 7
Какой атрибут отвечает за минимальное число образцов в узле для разделения?
a) min_samples_leaf
b) max_depth
c) min_samples_split
d) min_impurity_decrease

Задание 8
Отметьте верные утверждения о влиянии параметров решающего деревья на переобучение
a) Чем меньше значение min_samples_split, тем меньше тенденция к переобучению
b) Чем меньше значение min_samples_leaf, тем меньше тенденция к переобучению
c) Чем меньше глубина дерева, тем меньше тенденция к переобучению
d) Глубина дерева никак не влияет на его переобученность
e) Чем меньше значение параметра max_leaf_nodes, тем меньше тенденция к переобучению

Задание 9
Какой классификатор необходимо использовать для предсказания размера одобренного кредита по заданным параметрам?
a) DecisionTreeRegressor
b) DecisionTreeClassifier
c) Может быть использован и DecisionTreeClassifier, и DecisionTreeRegressor

Задание 10
Отметьте верные утверждения о деревьях решений.
a) чем меньше свойство gini, тем однороднее примеры в листе
b) свойство gini определяется только для листьев, но не для узлов
c) свойство samples указывает на количество примеров в узле
d) параметр gini, во всех листьях должен иметь одинаковые значения

Задание 11
Отметьте верные утверждения о случайном лесе
a) чем больше деревьев в лесу, тем лучше качество предсказания
b) время настройки и работы случайного леса увеличивается пропорционально количеству деверев в лесу.
c) случайный лес применим для решения задач регрессии
d) случайный лес применим для решения задач классификации

Задание 12
Отметьте верные утверждения
a) RandomizedSearchCV находит лучшую комбинацию параметров для дерева решений
b) RandomizedSearchCV работает быстрее, чем GridSearchCV
c) GridSearchCV находит лучшую комбинацию параметров для дерева, из тех значений параметров, что были переданы
d) RandomizedSearchCV дает более точное предсказание, чем GridSearchCV

Задание 13
Выберите из списка задачи классификации:
a) разделить по фотографии животных на кошек и собак
b) предсказать стоимость квартиры по заданным параметрам
c) разделить грибы на съедобные и ядовитые
d) сгруппировать тексты по их эмоциональной окраске

Задание 14
Отметьте верные утверждения
a) В качестве переменных для обучения дерева могут быть использованы как количественные, так и категориальные признаки
b) Деревья решений могут решать задачу классификации с произвольным числом классов
c) Чем меньше значение min_samples_leaf тем точнее будут предсказания на тренировочной выборке
d) Чем меньше значение min_samples_leaf тем точнее будут предсказания на тестовой выборке

Задание 15
Обучающая выборка (X_text, y_test) это –
a) Выборка, по которой настраиваются оптимальные параметры дерева
b) Выборка, по которой оценивается качество полученного дерева решений
c) Выборка, по которой осуществляется выбор наилучшей модели из множества моделей
d) Множество целевых значений для данного набора данных

Задание 16
При обучении некоторой модели на тренировочной выборке и оценке её качества на тестовой выборке, получена большая разница между значениями метрик на тренировочной и тестовой выборках. О чем это может говорить?

a) Модель переобучена
b) Распределение данных в тестовой выборке сильно отличается от распределения в тренировочной
c) Модель недообучена
d) Модель идеально настроена

Задание 17
Data – pandas.DataFrame, размером 20 на 20. Какие из предложенных комбинаций не вызовут ошибку?

a) Data.iloc[[1,4,5],0:3]
b) Data.iloc[[0,15], [1,5,-1]]
c) Data.iloc[[1,4:15],0:3]
d) Data[1:5,[1,6]]

Задание 18
Какие строки позволяют отобрать 5 первых строк датафрэйма data (индексы последовательность чисел от 0 с шагом 1)

a) data.iloc[:5]
b) data.head(5)
c) data.loc[:5]
d) data.loc[:4]
e) data.tail(5)

Задание 19
В каких случаях может быть использована ступенчатая функция активации?

a) Определить к съедобным или несъедобным относится гриб?
b) Определить к какому из трех классов (setosa, vercicolor, verginica) относится ирис?
c) Разделить тексты на нейтральные, негативные и положительные.
d) Предсказать размер заработной платы специалиста по определенным параметрам.
e) Разделить изображения на содержащие и не содержащие открытый огонь.

Задание 20
Какое значение может иметь взвешенная сумма входов нейрона (до использования активационной функции)

a) [-1,1]
b) [0,1]
c) (-1,1)
d) $(-\infty,+\infty)$

2.2. Типовые задания для оценки навыкового образовательного результата

Проверяемый образовательный результат:

Код и наименование индикатора достижения компетенции	Образовательный результат
ОПК-5.2: использует методы искусственного интеллекта (машинного обучения) и анализа больших данных для решения прикладных задач	Обучающийся **умеет**: агрегировать данные средствами Python; выполнять фильтрацию данных средствами Python; настраивать вес нейронна; выбирать тип классификатора в зависимости от поставленной задачи; визуализировать данные средствами библиотек pandas, seaborn, matplotlib; выполнять многоклассовую классификацию методами библиотеки sklearn; строить деревья решений и выполнять по ним предсказание.

Задание 21
Откройте Jupter notebook, создайте файл lab2_<ваша фамилия>.ipyng (например, lab2_tiuzhna.ipynb), подключите библиотеки numpy и pandas и прочтите данные из файла StudentsPerformance.csv в датафрейм sp.
Получите распределение числовых величин датафрейма sp используя базовые методы визуализации библиотеки pandas.
Импортируйте библиотеку Seaborn. Используя интерфейс для подгонки регрессионных моделей к условным подмножествам набора данных выведите график успеваемости по математике и чтению с разделением по полу.

Задание 22
Загрузите модули библиотек numpay и matplotlib, а также модули библиотеки Scikit-learn, необходимые для работы с древом решений. Из sklearn.datasets рагрузите тестовый набор iris. Изучите структуру полученного набора данных. В наборе 150 экземпляров ("строк" в привычной нам терминологии). Каждому экземпляру соответствует четыре числовых атрибута (длина чашелистика в см; ширина чашелистика в см; длина лепестка в см; ширина лепестка в см) и класс ('setosa', 'versicolor', 'virginica'). Т.е. данный набор позволяет решить задачу классификации, отнесения растения "ирис" исходя из ширины лепестков, длины чашелистика и т.д. к одному из трех классов. Создайте демонстративный Decision Tree классификатор, используя одноимённый класс из библиотеки sklearn и сохраните его в переменную dt. Для разделения множества на

тестовое и тренировочное используйте метод train_test_split. Отобразите полученное дерево решений графически, используйте метод plot_tree

Задание 23
Загрузите данные из встроенного набора load_digits. Данные представляют собой набор из 1797 рукописных изображений цифр размерами 8 на 8. Каждая картинка представлена массивом из 64 значений цвета для каждой ячейки изображения. Обучите дерево решений и получите предсказание на тестовой выборке. Визуализируйте предсказание, укажите точность для полученного дерева.

Задание 24.
Загрузите встроенный набор boston.
Набор данных boston представляет собой словарь с ключами data, target, feature_names и DESCR. Ключу data соответствует массив списков:
{'data': array([[6.3200e-03, 1.8000e+01, 2.3100e+00, ..., 1.5300e+01, 3.9690e+02, 4.9800e+00], [2.7310e-02, 0.0000e+00, 7.0700e+00, ..., 1.7800e+01, 3.9690e+02, 9.1400e+00],
В каждой строчке 13 числовых значений. Ваша задача найти описание каждого из них. Так первое значение (6.3200e-03 или 0.00632) это уровень преступности на душу населения.
Постройте дерево решений, предсказывающее стоимость жилья по имеющимся данным. Обратите внимание, стоимость жилья - величина непрерывная, и дерево классификации DecisionTreeClassifier нам не подойдёт. Для предсказания значения целевой функции нужно регрессионное дерево DecisionTreeRegressor.
Переберите все деревья на данных boston по следующим параметрам:
максимальная глубина - от 2 до 15 уровней
минимальное число проб для разделения - от 2 до 10
минимальное число проб в листе - от 1 до 10
Отобразите лучшие параметры и лучшее дерево.

Задание 25.
Загрузите встроенный набор boston.
Изучите остальные атрибуты классификатора. Создайте модель RandomForestClassifier с 15 деревьями и максимальной глубиной равной 5 уровням для данных digits_load().
Обучите полученную модель. Выполните предсказание. Посчитайте точность (score) для леса.
Укажите три самых важных параметра для деревьев решений в созданном лесу, используя, например, атрибут feature_importances_.

ОПК-5.2: Использует методы искусственного интеллекта (машинного обучения) и анализа больших данных для решения прикладных задач	Обучающийся **владеет**: Навыками построения дерева решений. Навыками реализации алгоритма случайного леса (Random forest) Навыками написания нейронных сетей. Навыками обучения модели.

Задание 26
Загрузите данные из встроенного набора load_digits. Создайте классификатор, со следующими параметрами дерева: максимальная глубина - 5 уровней, минимальное число объектов в листе - 4.
Разделите полученные данные на тренировочный и тестовый наборы данных, 75% и 25% соответственно. Предскажите классы, к которым принадлежат данные из тестовой выборки, сохраните результат предсказаний в переменную predicted.

Задание 27
Загрузите встроенный набор boston.
Набор данных boston представляет собой словарь с ключами data, target, feature_names и DESCR. Ключу data соответствует массив списков:
Используя алгоритм случайный лес (random forest) Создайте модель RandomForestClassifier с 20 деревьями и максимальной глубиной равной 5 уровням для данных. Обучите полученную модель. Выполните предсказание. Посчитайте точность (score) для леса

Задание 28
Импортируйте набор данных Fashion MNIST (данные библиотеки keras), содержащий 70 000 изображений элементов одежды в градациях серого, в разрешении 28x28 пикселей.и прочтите данные/
Разделите набор на тренировочный и тестовый в пропорции 85/15.
Визуализируйте первые десять объектов из тренировочного набора.
Постройте нейронную сеть, состоящую из трех слоев: входного, скрытого и выходного.

Задание 29
Скачайте данные с сайта kugle (https://www.kaggle.com/c/dogs-vs-cats/data). Для скачивания необходимо пройти процедуру регистрации. Распакуйте архивы train.zip и test.zip с данными в каталоге со своим проектом.
В каталоге train создайте две папки cat и dog. Разнесите изображения из каталога train по вновь созданным папкам. Вы может сделать это вручную (0 баллов) или программно (1 балл) используя методы языка Python. Вам понадобится библиотека os и методы строк.
Сгенерируйте датасета из изображений используя метод библиотеки keras ImageDataGenerator.

С помощью метода flow_from_directory создайте генератор данных на основе изображений из каталога.
Исследуйте полученный генератор. Выведите на экран, изучите форму.
Проведите обратную операцию. Визуализируйте первые пять объектов из полученного набора данных.
Расщепите имеющиеся данные на три набора (train,test,validation), либо используйте данные из архива test, подготовив их соответствующим образом.
Самостоятельно создайте и обучите нейронную сеть для различения кошек и собак (по примеру лабораторной 5, часть 2).
Добейтесь точности модели на валидационной выборке не ниже 80%

Задание 30
Реализуйте класс Neuron осуществляющий выполнение функции =x1*w1+x2*w2 +... +xnwn+b (1), где x – значения, поданные на вход, w - веса входов, а b - смещение. В качестве функции активации используйте сигмоиду. Убедитесь, что созданный вами нейрон работает. Для этого самостоятельно задайте веса и смещения, и значения входов и проверьте ответ.
Соберите из нейронов простейшую искусственную сеть, состоящую из трех слоев: входного, скрытого и выходного.
Скрытый слой – это любой слой между входным (первым) слоем сети и выходным (последним). Скрытых слоев может быть несколько!

2.3. Перечень вопросов для подготовки обучающихся к промежуточной аттестации

1. Искусственный интеллект. Основные понятия ИИ.
2. Машинное обучение.
3. Глубокое обучение.
4. Обучение с учителем и без учителя.
5. Основные понятия анализа данных. Данные, модель, обработка данных.
6. Методологические принципы анализы данных.
7. Цели, этапы, методы и техники анализа данных.
8. Фильтрация группировка и агрегация данных.
9. Визуализация данных.
10. Визуализация данных в Phyton.
11. Визуализация в Pandas.
12. Библиотеки Seaborn и Matplotlib
13. Решающие деревья: основные понятия.
14. Задача классификации.
15. Деревья решений. Основные параметры дерева.
16. Дерево решений в задачи регрессии.
17. Метод ближайших соседей.
18. Концепция случайного леса.
19. Случайная выборка тренировочных образцов.
20. Усреднение прогнозов.
21. Проблема переобучения.
22. Типы ошибок: true negative, false positive.
23. Метрики качества модели: Precision, Recall, F1 score.
24. Тестирование ROC AUC.
25. Нейронные сети.
26. Искусственные нейроны.
27. Функции активации нейрона.
28. Перцептрон.
29. Обучение нейрона.
30. Градиентный спуск.
31. Дробление шага при градиентном спуске.
32. Стохастический градиентный спуск.
33. Метод наискорейшего спуска.
34. Нейронная сеть.
35. Однослойная модель.
36. Сверточные нейронные сети.
37. Свёрточный слой, слой подвыборки, полносвязный слой.
38. Целевая функция.
39. Задачи распознавания образов.
40. Классификация объектов.
41. Поиск изображения по образцу.

3. Методические материалы, определяющие процедуру и критерии оценивания сформированности компетенций при проведении промежуточной аттестации

Критерии формирования оценок по ответам на вопросы, выполнению тестовых заданий

- оценка **«отлично»** выставляется обучающемуся, если количество правильных ответов на вопросы составляет 100 – 90% от общего объёма заданных вопросов;

- оценка **«хорошо»** выставляется обучающемуся, если количество правильных ответов на вопросы – 89 – 76% от общего объёма заданных вопросов;

- оценка **«удовлетворительно»** выставляется обучающемуся, если количество правильных ответов на тестовые вопросы –75–60 % от общего объёма заданных вопросов;

- оценка **«неудовлетворительно»** выставляется обучающемуся, если количество правильных ответов – менее 60% от общего объёма заданных вопросов.

Критерии формирования оценок по результатам выполнения заданий

«Отлично» – ставится за работу, выполненную полностью без ошибок и недочетов.

«Хорошо» – ставится за работу, выполненную полностью, но при наличии в ней не более одной негрубой ошибки и одного недочета, не более трех недочетов.

«Удовлетворительно» – ставится за работу, если обучающийся правильно выполнил не менее 2/3 всей работы или допустил не более одной грубой ошибки и двух недочетов, не более одной грубой и одной негрубой ошибки, не более трех негрубых ошибок, одной негрубой ошибки и двух недочетов.

«Неудовлетворительно» – ставится за работу, если число ошибок и недочетов превысило норму для оценки «удовлетворительно» или правильно выполнено менее 2/3 всей работы.

Виды ошибок:

- *грубые ошибки: незнание основных понятий, правил, норм; незнание приемов решения задач; ошибки, показывающие неправильное понимание условия предложенного задания.*

- *негрубые ошибки: неточности формулировок, определений; нерациональный выбор хода решения.*

- *недочеты: нерациональные приемы выполнения задания; отдельные погрешности в формулировке выводов; небрежное выполнение задания.*

Критерии формирования оценок по зачету

«Зачтено» – студент демонстрирует знание основных разделов программы изучаемого курса: его базовых понятий и фундаментальных проблем; приобрел необходимые умения и навыки, освоил вопросы практического применения полученных знаний, не допустил фактических ошибок при ответе, достаточно последовательно и логично излагает теоретический материал, допуская лишь незначительные нарушения последовательности изложения и некоторые неточности.

«Незачтено» – выставляется в том случае, когда студент демонстрирует фрагментарные знания основных разделов программы изучаемого курса: его базовых понятий и фундаментальных проблем. У экзаменуемого слабо выражена способность к самостоятельному аналитическому мышлению, имеются затруднения в изложении материала, отсутствуют необходимые умения и навыки, допущены грубые ошибки и незнание терминологии, отказ отвечать на дополнительные вопросы, знание которых необходимо для получения положительной оценки.