

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Гаранин Максим Александрович
Должность: Ректор
Дата подписания: 02.10.2025 11:14:27
Уникальный программный ключ:
7708e3a47e66a8ee02711b298d7c78bd1e40bf88

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ПРИВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПУТЕЙ СООБЩЕНИЯ»

МОДУЛЬ "СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА"

Технологии машинного обучения рабочая программа дисциплины (модуля)

Направление подготовки 09.03.03 Прикладная информатика

Направленность (профиль) Управление цифровой инфраструктурой организации

Квалификация **бакалавр**

Форма обучения **очная**

Общая трудоемкость **10 ЗЕТ**

Виды контроля в семестрах:

экзамены 6

зачеты 5

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	5 (3.1)		6 (3.2)		Итого	
Неделя	16 3/6		16 3/6			
Вид занятий	уп	рп	уп	рп	уп	рп
Лекции	16	16	16	16	32	32
Лабораторные			48	48	48	48
Практические	16	16			16	16
Конт. ч. на аттест. в период ЭС	0,15	0,15	2,3	2,3	2,45	2,45
Итого ауд.	32	32	64	64	96	96
Контактная работа	32,15	32,15	66,3	66,3	98,45	98,45
Сам. работа	103	103	125	125	228	228
Часы на контроль	8,85	8,85	24,7	24,7	33,55	33,55
Итого	144	144	216	216	360	360

Программу составил(и):

к.п.н., доцент, Тюжина И.В.

Рабочая программа дисциплины

Технологии машинного обучения

разработана в соответствии с ФГОС ВО:

Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 09.03.03 Прикладная информатика (приказ Минобрнауки России от 19.09.2017 г. № 922)

составлена на основании учебного плана: 09.03.03-25-3-ПИБ.plm.plx

Направление подготовки 09.03.03 Прикладная информатика Направленность (профиль) Управление цифровой инфраструктурой организации

Рабочая программа одобрена на заседании кафедры

Цифровые технологии

Зав. кафедрой Горбатов С.В.

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.18.02
-------------------	------------

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

ОПК-6 Способен анализировать и разрабатывать организационно-технические и экономические процессы с применением методов системного анализа и математического моделирования

ОПК-6.2 Проводит расчеты основных показателей результативности создания и применения информационных систем и технологий

В результате освоения дисциплины (модуля) обучающийся должен

3.1	Знать:
3.1.1	функции и методы библиотеки Pandas: query, tail, sample, head, groupby, agg;
3.1.2	методы фильтрации и агрегации данных;
3.1.3	методы библиотек tensorflow, keras, seaborn, matplotlib;
3.1.4	основные методы машинного обучения;
3.1.5	ансамблевые методы
3.1.6	основные метрики качества моделей машинного обучения: accuracy, precision, recall, F1 score.
3.2	Уметь:
3.2.1	выполнять фильтрацию данных средствами Python;
3.2.2	агрегировать данные средствами Python;
3.2.3	визуализировать данные средствами библиотек pandas, seaborn, matplotlib;
3.2.4	выбирать тип классификатора в зависимости от поставленной задачи;
3.2.5	строить деревья решений и выполнять по ним предсказание;
3.2.6	строить модель "случайный лес" и выполнять по ней предсказание;
3.2.7	оценивать качество модели обучения;
3.2.8	проводить машинный анализ данных с применением информационных систем и технологий;
3.2.9	решать задачи классификации методами машинного обучения;
3.2.10	решать задачи регрессии методами машинного обучения.
3.3	Владеть:
3.3.1	навыками построения дерева решений;
3.3.2	реализации алгоритма случайного леса (Random forest);
3.3.3	написания однослойных и многослойных нейронных сетей;
3.3.4	настройки нейронных сетей;
3.3.5	обучения модели.
3.3.6	

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Примечание
	Раздел 1. Введение в обработку данных			
1.1	Основы обработки данных. Открытые данные. Машиночитаемые форматы. /Лек/	5	2	
1.2	Алгоритмы и структуры данных. Датафреймы. Вывод строк и столбцов. Методы loc и iloc. /Лек/	5	2	
1.3	Алгоритмы и структуры данных. Библиотека Pandas: структуры данных: Series и DataFrame. Работа с csv. /Пр/	5	2	
1.4	Датафреймы. Фильтрация данных. Метод query. /Лек/	5	2	
1.5	Датафреймы. Фильтрация данных. Метод query. /Пр/	5	2	

1.6	Датафреймы. Сортировка, группировка и агрегация данных. /Лек/	5	2	
1.7	Датафреймы. Сортировка, группировка и агрегация данных. /Пр/	5	2	
1.8	Визуализация данных. Визуализация данных в Phyton. Визуализация в Pandas. Библиотеки Seaborn и Matplotlib. /Лек/	5	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. /Пр/	5	2	
1.10	Фильтрация данных. Библиотека Pandas. Вывод строк и колонок. Методы query, tail, sample, head. Объединение методов. /Пр/	5	2	
1.11	Агрегация данных в Pandas: count, sum, max, min, mean, median. Методы groupby и aggregate. Объединение методов. /Пр/	5	2	
1.12	Отчистка данных. Работа с пропущенными значениями. /Лек/	5	2	
1.13	Предобработка данных. Преобразование данных. /Лек/	5	2	
1.14	Встроенные наборы данных. Работа с пропущенными значениями. Замена NaN. /Пр/	5	2	
1.15	Изменение датафрейма. Добавление строк и столбцов. Комбинирование методов. /Лек/	5	2	
1.16	Изменение датафрейма. Добавление строк и столбцов. Комбинирование методов. /Пр/	5	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/	6	2	
2.2	Применение ансамблевых методов в машинном обучении. Концепция случайного леса. Случайная выборка тренировочных образцов. Усреднение прогнозов. /Лек/	6	2	
2.3	Проблема переобучения. Определение переобучения, методы борьбы с ним. /Лек/	6	2	
2.4	Нейронные сети. Искусственные нейроны. Перцептрон. Обучение нейрона. /Лек/	6	2	
2.5	Градиентный спуск. Дробление шага при градиентном спуске. Стохастический градиентный спуск. Метод наискорейшего спуска. /Лек/	6	2	
2.6	Нейронная сеть. Однослойная модель. /Лек/	6	2	
2.7	Метод обратного распространения ошибки. /Лек/	6	2	
2.8	Метод опорных векторов. /Лек/	6	2	
2.9	Решающие деревья: обучение модели. Библиотеки sklearn, pandas, numpy. Выбор параметров модели. Кросс-валидация. Предварительная обработка данных. Тренировочный набор данных. Тестовый набор данных. Дерево классификации. /Лаб/	6	2	
2.10	Решающие деревья. Выбор оптимальных параметров дерева. Дерево решений в задачах регрессии. /Лаб/	6	2	
2.11	Библиотеки sklearn, pandas, numpy. Выбор параметров модели. Кросс-валидация. Предварительная обработка данных. Тренировочный набор данных. Тестовый набор данных. /Лаб/	6	2	
2.12	Применение ансамблевых методов в машинном обучении. Концепция случайного леса. Случайная выборка тренировочных образцов. Усреднение прогнозов. Проблема переобучения. /Лаб/	6	2	
2.13	Типы ошибок: true negative, false positive. Метрики качества: Precision, Recall, F1 score. Тестирование ROC AUC. /Лаб/	6	2	
2.14	Распознавание рукописных цифр. Алгоритмы случайного леса. Оптимальные параметры дерева и леса. /Лаб/	6	2	
2.15	Искусственный нейрон. Вход, функция активации (сигмоида, гиперболический тангенс, SoftMax). Подбор весов. /Лаб/	6	4	

2.16	Определение тональности текста с помощью нейронных сетей. Библиотеки tensorflow, Keras, набор данных IMDB. /Лаб/	6	2	
2.17	Библиотека Keras. Полносвязный и сглаживающий слой, слой Dropout. Решение проблем переобучения. /Лаб/	6	2	
2.18	Классификация изображений по нескольким классам (более двух). Набор данных Fashion Mnist. /Лаб/	6	4	
2.19	Внешние источники данных для анализа. Kuggle. Подготовка данных. Работа с изображениями разного формата. /Лаб/	6	2	
2.20	Распознавание образов. Классификация изображений кошек и собак. Библиотека keras, ImageDataGenerator. Достижение заданной точности модели на валидационной выборке. /Лаб/	6	4	
2.21	Написание собственных классов для описания нейронной сети. Нейрон. Слой. Сеть. Обучение модели. Выполнение предсказания. /Лаб/	6	4	
2.22	Предсказание пола по росту и весу с помощью нейронной сети. Выполнение предсказания. Минимизация значения функции потерь. Обучение: стохастический градиентный спуск. /Лаб/	6	2	
2.23	Метод опорных векторов. /Лаб/	6	4	
2.24	Метод обратного распространения ошибки /Лаб/	6	4	
2.25	Работа с неструктурированными данными. /Лаб/	6	4	
Раздел 3. Самостоятельная работа				
3.1	Библиотека NumPy. Работа с матрицами. /Ср/	5	6	
3.2	Подготовка к лекциям /Ср/	5	8	
3.3	Подготовка к практическим работам /Ср/	5	7	
3.4	Алгоритм обратного распространения ошибки. Функции активации. Оценка работы сети. /Ср/	6	18	
3.5	Задачи распознавания образов. Классификация объектов. Поиск изображения по образцу. /Ср/	6	18	
3.6	Компьютерное зрение. Библиотека OpenCV. Сегментация изображений. Детектирование объектов. /Ср/	6	18	
3.7	Компьютерное зрение. Библиотека OpenCV. Отслеживание движущихся объектов во времени. Распознавание лиц. /Ср/	6	20	
3.8	Подготовка к лекциям /Ср/	6	5	
3.9	Подготовка к лабораторным занятиям /Ср/	6	30	
3.10	Обработка естественного языка. Классификация длинных текстов. Векторная модель текста. /Ср/	5	10	
3.11	Обработка естественного языка. Дистрибутивная семантика. Векторные представления слов. /Ср/	5	10	
3.12	Математические основы искусственного интеллекта /Ср/	5	40	
3.13	Биологический нейрон. Принцип работы. /Ср/	5	10	
3.14	Лингвистический анализ. Извлечение признаков. /Ср/	5	12	
3.15	Языковые модели. Генерация текста. Рекуррентные нейронные сети. /Ср/	6	16	
Раздел 4. Контактные часы на аттестацию				
4.1	Зачёт /КЭ/	5	0,15	
4.2	Экзамен /КЭ/	6	2,3	

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) 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	http://e.lanbook.com/book/16

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/

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

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ
ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

(наименование дисциплины(модуля))

Направление подготовки / специальность

09.03.03 Прикладная информатика

(код и наименование)

Направленность (профиль)/специализация

Управление цифровой инфраструктурой организации

(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, 5 семестр

Перечень компетенций, формируемых в процессе освоения дисциплины

Код и наименование компетенции	Код индикатора достижения компетенции
ОПК-6: Способен анализировать и разрабатывать организационно-технические и экономические процессы с применением методов системного анализа и математического моделирования;	ОПК-6.1: Применяет методы математического моделирования для анализа информационных потоков, расчета экономической эффективности и надежности информационных систем и технологий

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

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Оценочные материалы
ОПК-6.1: Применяет методы математического моделирования для анализа информационных потоков, расчета экономической эффективности и надежности информационных систем и технологий	Обучающийся знает : основные понятия искусственного интеллекта: машинное обучение, глубокое обучение, обучение с учителем и без учителя; классификацию задач машинного обучения; метрики качества модели; признаки переобученности модели.	Вопросы (№1 - №20)
	Обучающийся умеет : настраивать вес нейронна; выбирать тип классификатора в зависимости от поставленной задачи; выполнять многоклассовую классификацию методами библиотеки sklearn;	Задания (№20 - №26)
	Обучающийся владеет : работы с генеративными нейронными моделями; написания нейронных сетей; обучения модели.	Задания (№27 - №30)

Промежуточная аттестация (зачет) проводится в одной из следующих форм:

- 1) собеседование;
- 2) выполнение заданий в ЭИОС университета.

2. Типовые¹ 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:

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.

Код и наименование индикатора достижения компетенции	Образовательный результат
ОПК-6.1: Применяет методы математического моделирования для анализа информационных потоков, расчета экономической эффективности и надежности информационных систем и технологий	Обучающийся знает : основные понятия искусственного интеллекта: машинное обучение, глубокое обучение, обучение с учителем и без учителя; классификацию задач машинного обучения; метрики качества модели; признаки переобученности модели.
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 основе данных, 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 Модель показывает на тренировочной выборке точность 0,95, а на тестовой 0,8. О чём это может свидетельствовать? a) Модель переобучена b) Модель недообучена c) Модель близка к идеальной d) Таких параметров быть не может 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 Какие из предложенных задач относятся к задачам обучения без учителя? a) Задача классификация b) Задача регрессии c) Задача кластеризации d) Задача поиска аномалий Задание 7 Какие из предложенных задач относятся к задачам обучения с учителем? a) Прогнозирование нагрузки на сервер b) Обнаружение вредоносной активности в сети c) Кластеризация пользователей по поведению d) Классификация типов запросов в службу поддержки: технические проблемы, вопросы о продукте, запросы на возврат Задание 8	

Отметьте верные утверждения о влиянии параметров решающего дерева 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 деревьев 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 датафрейма 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, versicolor, virginica) 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-6.1: Применяет методы математического моделирования для анализа информационных потоков, расчета экономической эффективности и надежности информационных систем и технологий	Обучающийся умеет : настраивать вес нейронна; выбирать тип классификатора в зависимости от поставленной задачи; выполнять многоклассовую классификацию методами библиотеки sklearn;

Задание 21

Необходимо используя исторические данные о нагрузке, время суток и другие факторы, предсказывать, сколько ресурсов (ЦП, память) потребуется в будущем. Какие методов из изученных нами можно использовать для построения модели?

- a) DecisionTreeClassifier
- b) DecisionTreeRegressor
- c) RandomForestClassifier
- d) RandomForestRegressor
- e) Нейронные сети

Задание 22

Что такое алгоритм коррекции ошибки. Расскажите, как он работает?

Ответ: Основная идея метода заключается в том, чтобы корректировать веса нейронов в случае возникновения ошибки предсказания. Основные шаги работы метода: 1. Инициализация весов. 2. Подача входных данных. 3. Вычисление выхода. 4. Оценка ошибки. 5. Коррекция весов: Коррекция весов производится по следующему правилу: • Для каждого веса w_i выполняется обновление веса $w_i = w_i + \Delta w_i$, где $\Delta w_i = \eta (d - y) x_i$, где η — скорость обучения (параметр, который контролирует размер шага обновления весов), d — желаемый выход (истинная метка), y — фактический выход, x_i — входное значение. 6. Шаги 2-5 повторяются для всех обучающих примеров и продолжаются до тех пор, пока ошибка не станет достаточно малой.

Задание 23

Как выбрать между DecisionTreeClassifier и RandomForestClassifier для задачи классификации?

Ответ: Выбор между DecisionTreeClassifier и RandomForestClassifier зависит от сложности задачи и объема данных. DecisionTreeClassifier может быть хорошим выбором для простых задач с небольшим объемом данных, но он подвержен переобучению. RandomForestClassifier, который состоит из множества деревьев решений, более устойчив к переобучению и обычно показывает лучшие результаты на более сложных задачах с большим количеством данных.

Задание 24.

Модель классифицирует обращения пользователей по категориям (технические проблемы, вопросы о продукте, запросы на возврат и т.д.) для более эффективного распределения ресурсов. С помощью каких метрик вы можете оценить производительность модель? Приведите не менее двух примеров.

Ответ: Может быть использована точность (accuracy) — доля правильных предсказаний, матрица ошибок (confusion matrix) — показывает, сколько объектов каждого класса было правильно или неправильно классифицировано. F1-score — гармоническое среднее точности и полноты.

Задание 25.

Загрузите встроенный набор iris.

Изучите остальные атрибуты классификатора. Создайте модель DecisionTreeClassifier с максимальной глубиной равной 5 уровням. Обучите полученную модель. Визуализируйте дерево. Выполните предсказание. Посчитайте точность (score) для дерева.

Задание 26.

Загрузите встроенный набор digits.

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-6.1: Применяет методы математического моделирования для анализа информационных потоков, расчета экономической эффективности и надежности информационных систем и технологий

Обучающийся владеет:

работы с генеративными нейронными моделями;
написания нейронных сетей;
обучения модели.

Задание 27.

Используя сервис для создания презентаций с использованием искусственного интеллекта, например, Gamma, создайте презентацию на тему «Проблемы переобучения моделей» (8 слайдов). Оцените сгенерированные ИИ факты. Проведите фактчекинг.

Задание 28.

Используя сервис, генерации изображений по текстовым запросам, например, Kandinsky, создайте три изображения на тему «Дерево решений». Оцените сгенерированные изображения. Перечислите проблемы, с которыми вы столкнулись.

Задание 29.

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 пяти слоев. Добейтесь точности 93 и выше.

Задание 30.

Импортируйте набор данных digits (данные библиотеки sklearn), содержащий изображения рукописных цифр, в разрешении 8x8 пикселей.

Разделите набор на тренировочный и тестовый в пропорции 70/30.

Визуализируйте первые десять объектов из тренировочного набора.

Постройте нейронную сеть, состоящую из трех и более слоев. Оцените точность полученной модели.

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 анализа данных.
7. Цели, этапы, методы и техники анализа данных.
8. Фильтрация группировка и агрегация данных.

9. Визуализация данных.
10. Визуализация данных в Python.
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 задаче 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.

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 объема заданных вопросов;
- оценка **«хорошо»** выставляется обучающемуся, если количество правильных ответов на вопросы – 89 – 76% от общего объема заданных вопросов;
- оценка **«удовлетворительно»** выставляется обучающемуся, если количество правильных ответов на тестовые вопросы – 75–60 % от общего объема заданных вопросов;
- оценка **«неудовлетворительно»** выставляется обучающемуся, если количество правильных ответов – менее 60% от общего объема 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.