

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

ТЕХНОЛОГИИ МАШИННОГО ОБУЧЕНИЯ

(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; экзамен, 6 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-6.2: Проводит расчеты основных показателей результативности создания и применения 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-6.2: Проводит расчеты основных показателей результативности создания и применения информационных систем и технологий	Обучающийся знает : функции и методы библиотеки Pandas: query, tail, sample, head, groupby, agg; методы фильтрации и агрегации данных; методы библиотек tensorflow, keras, seaborn, matplotlib; основные методы машинного обучения; ансамблевые методы основные метрики качества моделей машинного обучения: accuracy, precision, recall, F1 score.	Вопросы (№1 - №20)
	Обучающийся умеет : выполнять фильтрацию данных средствами Python; агрегировать данные средствами Python; визуализировать данные средствами библиотек pandas, seaborn, matplotlib; выбирать тип классификатора в зависимости от поставленной задачи; строить деревья решений и выполнять по ним предсказание; строить модель "случайный лес" и выполнять по ней предсказание; оценивать качество модели обучения; проводить машинный анализ данных с применением информационных систем и технологий; решать задачи классификации методами машинного обучения; решать задачи регрессии методами машинного обучения.	Задания (№21 - №25)
	Обучающийся владеет : навыками построения дерева решений; реализации алгоритма случайного леса (Random forest); написания однослойных и многослойных нейронных сетей; настройки нейронных сетей; обучения модели.	Задания (№26 - №30)

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

- 1) ответ на билет, состоящий из теоретических вопросов и практических заданий;
- 2) выполнение заданий в ЭИОС СамГУПС.

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

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

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:

Код и наименование индикатора достижения компетенции	Образовательный результат
ОПК-6.2: Проводит расчеты основных показателей результативности создания и применения информационных систем и технологий	Обучающийся знает : - основные классы научно-технических задач в области своей профессиональной деятельности, решаемые методами машинного обучения; основные методы - машинного обучения; основные метрики качества моделей машинного обучения;
Задание №1	
Импортируйте набор данных fmri. Для загрузки используйте библиотеку seaborn и команду <code>seaborn.load_dataset('fmri')</code> . Ответьте, сколько строк в этом наборе. В поле ответа введите число без пробелов.	
Запишите число:	
1)	Ответ:
Задание №2	
Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду <code>seaborn.load_dataset('diamonds')</code> . Сколько бриллиантов органики Very Good в этом наборе? За огранку отвечает столбец 'cut'. Скопируйте ответ в поле ниже.	
Запишите число:	
1)	Ответ:
Задание №3	
Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду <code>seaborn.load_dataset('diamonds')</code> . Отсортируйте набор по возрастанию цены. Выберите в полученном наборе строку с индексом 78. Чему равно значение столбца price в этой строке?	
Запишите число:	
1)	Ответ:
Задание №4	
Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду <code>seaborn.load_dataset('diamonds')</code> . Отберите только бриллианты с огранкой Ideal. Сколько стоит третий бриллиант в полученном наборе? За огранку отвечает столбец 'cut', за цену столбец 'price'. Скопируйте ответ в поле ниже.	
Запишите число:	
1)	Ответ:

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.

Задание №5

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду `seaborn.load_dataset('diamonds')`. Сколько бриллиантов органики Premium в этом наборе? За огранку отвечает столбец 'cut'. Скопируйте ответ в поле ниже.

Запишите число:

1)	Ответ:	
----	--------	--

Задание №6

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду `seaborn.load_dataset('diamonds')`. Отберите только бриллианты цвета «J» и огранки «Premium». Какова стоимость третьего бриллианта в наборе? За цвет отвечает столбец 'color', за огранку 'cut'.

Запишите число:

1)	Ответ:	
----	--------	--

Задание №7

Импортируйте набор данных brain_networks. Для загрузки используйте библиотеку seaborn и команду `seaborn.load_dataset('brain_networks')`. Ответьте какое значение находится в ячейке датасета с индексами 134, 7? Скопируйте ответ в поле ниже.

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №8

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду `seaborn.load_dataset('diamonds')`. Отберите только бриллианты, весящие меньше 1-х карата, огранки «Ideal». Какова максимальная стоимость бриллианта из этого набора? За вес отвечает столбец 'carat', за огранку 'cut', за стоимость 'price'.

Запишите ответ:

1)	Ответ:	
----	--------	--

Задание №9

Какая из команд выведет столбцы 'gender' и 'math score' датафрейма sp?

Выберите один из 4 вариантов ответа:

1)	<code>sp[['gender','math score']]</code>
2)	<code>sp.loc(['gender','math score'])</code>
3)	<code>sp.iloc['gender','math score']</code>
4)	<code>sp['gender','math score']</code>

Задание №10

Как называется негативное явление, возникающее, когда алгоритм обучения вырабатывает предсказания, которые слишком близко или точно соответствуют конкретному набору данных и не подходят для применения алгоритма к дополнительным данным или будущим наблюдениям?

Выберите один из 4 вариантов ответа:

1)	переобучение
----	--------------

2)		недообучение
3)		кроссвалидация
4)		решающее правило

Задание №11

Для чего используется подвыборочный слой (иначе слой пулинга, подвыборки, субдискретизации)?

Выберите несколько из 4 вариантов ответа:

1)		уменьшение изображения,
2)		ускорение вычислений
3)		распознавание мелких деталей изображения
4)		разделение изображения на фрагменты

Задание №12

Какая функция применяется для того, чтобы количественно оценить величину расхождения между полученными и ожидаемыми решениями нейронной сети?

Выберите один из 4 вариантов ответа:

1)		функция потерь
2)		функция активации
3)		сумматорная функция
4)		синусоида

Задание №13

Какую метрику целесообразно использовать для оценки качества модели определения опухоли на рентгеновском снимке?

Выберите один из 4 вариантов ответа:

1)		recall
2)		precision
3)		Accuracy
4)		F1

Задание №14

Как называется слой, в котором каждый нейрон соединен со всеми нейронами на предыдущем уровне, причем каждая связь имеет свой весовой коэффициент?

Выберите один из 4 вариантов ответа:

1)		полносвязный
2)		Dropout
3)		сумматорный
4)		активационный

Задание №15

Какую метрику целесообразно использовать для оценки качества модели определения съедобных ягод по фотографии (из набора съедобных и ядовитых)?

Выберите один из 4 вариантов ответа:

1)		recall
2)		precision
3)		Accuracy
4)		F1

Задание №16

Отметьте верные утверждения о влиянии параметров решающего дерева на переобучение.

Выберите несколько из 5 вариантов ответа:

1)		Чем меньше значение min_samples_split, тем меньше тенденция к переобучению
2)		Чем меньше значение min_samples_leaf, тем меньше тенденция к переобучению
3)		Чем меньше глубина дерева, тем меньше тенденция к переобучению
4)		Глубина дерева никак не влияет на его переобученность
5)		Чем меньше значение параметра max_leaf_nodes, тем меньше тенденция к переобучению

Задание №17

Отметьте верные утверждения о метрике качества accuracy...

Выберите несколько из 4 вариантов ответа:

1)		accuracy показывает отношение верно классифицированных наблюдений к общему количеству наблюдений
2)		accuracy особенно полезна в задачах с несбалансированными классами - когда одних значений (ядовитых грибов, например) значительно меньше, чем других (съедобных грибов).
3)		accuracy используется преимущественно в задачах классификации, но не регрессии
4)		accuracy показывает долю объектов, названных классификатором положительными и при этом действительно являющихся положительными

Задание №18

Отметьте верные утверждения о схеме построения отдельного дерева в модели «Случайный лес»?

Выберите один из 4 вариантов ответа:

1)		для обучения используется подвыборка обучающей выборки
2)		для обучения используется вся обучающая выборка
3)		для обучения всегда используется половина обучающей выборки
4)		для обучения каждого дерева используется строго независимая (без возвращения) подвыборка обучающей выборки

Задание №19

Отметьте верные утверждения о случайном лесе

Выберите несколько из 5 вариантов ответа:		
1)		чем больше деревьев в лесу, тем лучше качество предсказания
2)		время настройки и работы случайного леса увеличивается пропорционально количеству деревьев в лесу.
3)		случайный лес применим для решения задач регрессии
4)		случайный лес в общем случае менее точен, чем дерево решений
5)		количество рассматриваемых признаков всегда равно половине известных признаков

Задание №20

Дерево решений —

Выберите один из 4 вариантов ответа:

1)		средство поддержки принятия решений, использующееся в машинном обучении, анализе данных и статистике, представляет собой иерархическую структуру правил вида «если - то»
2)		свойство интеллектуальных систем выполнять отдельные функции интеллекта человека, в том числе творческие;
3)		разнообразные данные, которые поступают с постоянно растущей скоростью и объем которых постоянно растет;
4)		метод решения задач, точный алгоритм решения которых заранее известен.

#1 (1 б.)	Ответ = 1064
#2 (1 б.)	Ответ = 12082
#3 (1 б.)	Ответ = 368
#4 (1 б.)	Ответ = 344
#5 (1 б.)	Ответ = 13791
#6 (1 б.)	Ответ = 2807
#7 (1 б.)	Ответ = 48.09652709960938
#8 (1 б.)	Ответ = 9636.00
#9 (1 б.)	1
#10 (1 б.)	1
#11 (1 б.)	1, 2
#12 (1 б.)	1
#13 (1 б.)	1
#14 (1 б.)	1
#15 (1 б.)	2
#16 (1 б.)	3, 5
#17 (1 б.)	1, 3
#18 (1 б.)	1
#19 (1 б.)	1, 2, 3
#20 (1 б.)	1

2.2. Типовые задания для оценки навыкового образовательного результата

Проверяемый образовательный результат:

Код и наименование индикатора достижения компетенции	Образовательный результат
ОПК-6.2: Проводит расчеты основных показателей результативности создания и применения информационных систем и технологий	Обучающийся умеет: -использовать методы машинного обучения для решения прикладных задач; -оценивать качество модели обучения; -проводить машинный анализ данных о применения информационных систем и технологий;
	Задание 21 Откройте Jupyter 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 numpy и matplotlib, а также модули библиотеки Scikit-learn, необходимые для работы с древом решений. Из sklearn.datasets загрузите 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 одноименный 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 подойдет. 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-6.2: Проводит расчеты основных показателей результативности создания и применения информационных систем и технологий	Обучающийся владеет: -навыками анализа данных об информационной системе в том числе и с использованием технологий машинного обучения.
	Задание 26 Загрузите log-файл активности пользователей в ЭИОС вуза. Проведите первичный анализ. Сгруппируйте данные по пользователю. Выведите гистограмму по времени проводимому в системе. Постройте круговую диаграмму, отображающую затраты времени по разделам. Ответьте письменно, сколько времени в системе проводят чаще всего, какой раздел посещают чаще всего.

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 можете сделать это вручную (0 баллов) или программно (1 балл) используя методы языка `Python`. Вам понадобится библиотека `os` и методы строк.

Сгенерируйте датасет 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 $=x_1*w_1+x_2*w_2+...+x_n*w_n+b$ (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. Цели, этапы, методы и техники анализа данных.
7. Фильтрация группировка и агрегация данных.
8. Визуализация данных.
9. Визуализация данных в `Python`.
10. Визуализация в `Pandas`.
11. Библиотеки `Seaborn` и `Matplotlib`
12. Решающие деревья: основные понятия.
13. Задача классификации.
14. Деревья решений. Основные параметры дерева.
15. Дерево решений в задачи регрессии.
16. Метод ближайших соседей.
17. Концепция случайного леса.
18. Случайная выборка тренировочных образцов.
19. Усреднение прогнозов.
20. Проблема переобучения.
21. Типы ошибок: `true negative`, `false positive`.
22. Метрики качества модели: `Precision`, `Recall`, `F1 score`.
23. Тестирование `ROC AUC`.
24. Нейронные сети.
25. Искусственные нейроны.
26. Функции активации нейрона.
27. Перцептрон.

28. Обучение нейрона.
29. Градиентный спуск.
30. Дробление шага при градиентном спуске.
31. Стохастический градиентный спуск.
32. Метод наискорейшего спуска.
33. Нейронная сеть.
34. Однослойная модель.
35. Сверточные нейронные сети.
36. Свёрточный слой, слой подвыборки, полносвязный слой.
37. Целевая функция.
38. Задачи распознавания образов.
39. 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 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.

Критерии формирования оценок по экзамену

«Отлично» – студент приобрел необходимые умения и навыки, продемонстрировал навык практического применения полученных знаний, не допустил логических и фактических ошибок.

«Хорошо» – студент приобрел необходимые умения и навыки, продемонстрировал навык практического применения полученных знаний; допустил незначительные ошибки и неточности.

«Удовлетворительно» – студент допустил существенные ошибки.

«Неудовлетворительно» – студент демонстрирует фрагментарные знания изучаемого курса; отсутствуют необходимые умения и навыки, допущены грубые ошибки.