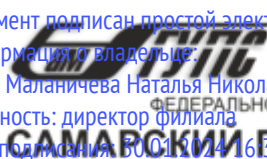


Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Маланичева Наталья Николаевна
Должность: директор филиала
Дата подписания: 00.01.2024 16:34:44
Уникальный программный ключ:
94732c3d953a82d495dcc3155d5c573883fedd18

 **МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ**
ФЕДЕРАЛЬНОЕ АГЕНТСТВО ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
САМАРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПУТЕЙ СООБЩЕНИЯ

Приложение
к рабочей программе дисциплины

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)

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 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 (№20 - №30)
	Обучающийся владеет: Навыками построения дерева решений; реализации алгоритма случайного леса (Random forest); написания нейронных сетей; обучения модели.	Задания (№31 - №35)

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-6.1: Применяет методы математического моделирования для анализа информационных потоков, расчета экономической эффективности и надежности информационных систем и технологий	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) <code>res.query('country == ["Russia", "Zimbabwe"]')</code>	
b) <code>res.query('country == "Russia" "Zimbabwe"')</code>	
c) <code>res.query('country == "Russia" country == "Zimbabwe"')</code>	
d) <code>res.query('country == "Russia" & country == "Zimbabwe"')</code>	
e) <code>res[res.country == "Russia" res.country == "Zimbabwe"]</code>	
f) <code>res[(res.country == "Russia") (res.country == "Zimbabwe")]</code>	
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	

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 6

Какой атрибут отвечает за минимальное число образцов в листе?

- a) min_samples_leaf
- 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 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 отобразить 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) (-∞,+∞)

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: 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

Импортируйте набор данных brain_networks. Для загрузки используйте библиотеку seaborn и команду seaborn.load_dataset('brain_networks'). Ответьте, сколько столбцов в этом наборе. В поле ответа введите число без пробелов.

Ответ: 63

Задание 22

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 34, 57? Скопируйте ответ в поле ниже.

Ответ: -0.06777933984994887

Задание 23

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

Ответ: 12082.

Задание 24.

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду seaborn.load_dataset('diamonds'). Каков средний вес бриллиантов органики Very Good в этом наборе? За огранку отвечает столбец 'cut', за вес столбец 'carat'. Скопируйте ответ в поле ниже.

Ответ: 0.806381

Задание 25.

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду seaborn.load_dataset('diamonds'). Отберите только бриллианты с огранкой Premium. Сколько стоит последний бриллиант в полученном наборе? За огранку отвечает столбец 'cut', за цену столбец 'price'. Скопируйте ответ в поле ниже.

Ответ: 2757

Задание 26.

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду seaborn.load_dataset('diamonds'). Отберите только бриллианты цвета «E». Какая глубина у третьего бриллианта в полученном наборе? За цвет отвечает столбец 'color', за глубину столбец 'depth'. Скопируйте ответ в поле ниже.

Ответ: 56.9

Задание 27.

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду seaborn.load_dataset('diamonds'). Отберите только бриллианты цвета «D» и огранки «Very Good». Сколько строк в полученном наборе?

Ответ: 1513

Задание 28.

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду seaborn.load_dataset('diamonds'). Отсортируйте набор по возрастанию цены. Выберите в полученном наборе строку с индексом 78. Чему равно значение столбца price в этой строке?

Ответ: 368

Задание 29.

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду seaborn.load_dataset('diamonds'). Отберите только бриллианты, весящие больше 1, но меньше 3 карат (не включительно). Какова их средняя стоимость? За вес отвечает столбец 'carat', за стоимость 'price'.

Ответ: 8385.952583

Задание 30.

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.

Ответ: 13791

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

Обучающийся владеет:

Навыками построения дерева решений.
Навыками реализации алгоритма случайного леса (Random forest)
Навыками написания нейронных сетей.
Навыками обучения модели.

Задание 31

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 32

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 33

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 34

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 35

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. 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 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.