

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)

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 : - основные классы научно-технических задач в области своей профессиональной деятельности, решаемые методами машинного обучения; основные методы - машинного обучения; основные метрики качества моделей машинного обучения;	Вопросы (№1 - №20)
	Обучающийся умеет : -использовать методы машинного обучения для решения прикладных задач; -оценивать качество модели обучения; -проводить машинный анализ данных в области информационных систем и технологий;	Задания (№21 - №25)
	Обучающийся владеет : -навыками оценки результативности применения информационных систем и с использованием технологий анализа данных и машинного обучения.	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:

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-6.2: Проводит расчеты основных показателей результативности создания и применения информационных систем и технологий	<p>Обучающийся знает:</p> <p>- основные классы научно-технических задач в области своей профессиональной деятельности, решаемые методами машинного обучения; основные методы - машинного обучения; основные метрики качества моделей машинного обучения;</p>
Задание 1	
Как называется метрика качества, которая оценивает долю правильно найденных положительных объектов среди всех объектов положительного класса?	
<ul style="list-style-type: none"> a) точностью (precision) b) полнота (recall) c) F1-мера d) Точность (score) 	
Задание 2	
Среди предложенных задач машинного обучения укажите задачи регрессии	
<ul style="list-style-type: none"> a) Поиск негативных отзывов о фильме 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 ...	
<ul style="list-style-type: none"> a) Tkinter b) Pandas c) NumPy d) Matplotlib 	
Задание 4	
Какие из перечисленных команд вернут данные для России и Зимбабве?	
<ul style="list-style-type: none"> 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	
<ul style="list-style-type: none"> 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?	
<ul style="list-style-type: none"> a) <code>min_samples_leaf</code> b) <code>max_depth</code> c) <code>min_samples_split</code> d) <code>min_impurity_decrease</code> 	
Задание 7	
Какой атрибут отвечает за минимальное число образцов в узле для разделения?	
<ul style="list-style-type: none"> a) <code>min_samples_leaf</code> b) <code>max_depth</code> c) <code>min_samples_split</code> d) <code>min_impurity_decrease</code> 	
Задание 8	
Отметьте верные утверждения о влиянии параметров решающего дерева на переобучение	
<ul style="list-style-type: none"> a) Чем меньше значение <code>min_samples_split</code>, тем меньше тенденция к переобучению b) Чем меньше значение <code>min_samples_leaf</code>, 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-6.2: 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>.ipynb (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 созданному 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, a 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 задаче 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.