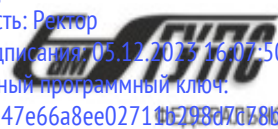


Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Гаранин Максим Алексеевич
Должность: Ректор
Дата подписания: 05.12.2023 16:07:50
Уникальный программный ключ:
7708e3a47e66a8ee027116298d7d7861e40b188



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

Приложение к рабочей программе
дисциплины

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОМЕЖУТОЧНОЙ
АТТЕСТАЦИИ
ПО ДИСЦИПЛИНЕ**

Работа с большими данными и машинное обучение
(наименование дисциплины)

Направление

09.04.02 Информационные системы и технологии
(код и наименование)

Направленность (профиль)

Цифровые технологии в образовании
(наименование)

Содержание

- Пояснительная записка.
- Типовые контрольные задания или иные материалы для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих уровень сформированности компетенций.
- 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.

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

?о умолчанию для РсБДиМО

УК-2: Способен управлять проектом на всех этапах его жизненного цикла

УК-2.2: Разрабатывает проект с учетом анализа альтернативных вариантов его реализации, определяет основные направления работ, управляет проектом на всех этапах его жизненного цикла

Знать: теоретические основы методов анализа больших данных и машинного обучения

(Множественный выбор / Только один ответ)

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

a. (0%)

```
sp.iloc['gender','math score']
```

b. (0%)

```
sp.loc(['gender','math score'])
```

c. (0%)

```
sp['gender','math score']
```

d. (100%)

```
sp[['gender','math score']]
```

(Множественный выбор)

Какие строки позволяют отобразить 5 последних строк датафрейма data?

a. (-100%)

```
data.iloc[-5:-1]
```

b. (-100%)

```
data.loc[-5:]
```

c. (50%)

```
data.iloc[-5:]
```

d. (-100%)

```
data.head(5)
```

e. (50%)

```
data.tail(5)
```

(Числовой ответ)

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

(Числовой ответ)

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

Ответ: 923

(Числовой ответ)

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

Ответ: 53940

(Числовой ответ)

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

Ответ: 6

(Числовой ответ)

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

Ответ: 2

(Числовой ответ)

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

Ответ: 649

(Числовой ответ)

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

Ответ: 144

(Числовой ответ)

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

Ответ: 3

(Числовой ответ)

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

Ответ: 1064

(Числовой ответ)

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

Ответ: 5

Уметь: ставить задачи и разрабатывать программу исследования с использованием анализа больших данных и машинного обучения

(Короткий ответ)

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

(Короткий ответ)

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.

Ответ: 48.09652709960938

(Короткий ответ)

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 13, 17? Скопируйте ответ в поле ниже.

Ответ: 1.8887639045715328

(Короткий ответ)

Импортируйте набор данных `brain_networks`. Для загрузки используйте библиотеку `seaborn` и команду `seaborn.load_dataset('brain_networks')`. Ответьте какое значение находится в столбце датасета с

названием 16.5 и строке с индексом 922? Скопируйте ответ в поле ниже.

Ответ: 102.0863037109375

(Короткий ответ)

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 313, 27? Скопируйте ответ в поле ниже.

Ответ: 1.5263108015060425

(Короткий ответ)

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 113, 34? Скопируйте ответ в поле ниже.

Ответ: -45.07174682617188

(Короткий ответ)

Импортируйте набор данных `diamonds`. Для загрузки используйте библиотеку `seaborn` и команду `seaborn.load_dataset('diamonds')`. Ответьте какое значение находится в ячейке датасета с индексами 45, 3? Скопируйте ответ в поле ниже.

Ответ: SI1

(Короткий ответ)

Импортируйте набор данных `diamonds`. Для загрузки используйте библиотеку `seaborn` и команду `seaborn.load_dataset('diamonds')`. Ответьте какое значение находится в ячейке датасета с индексами 145, 3? Скопируйте ответ в поле ниже.

Ответ: VVS2

(Короткий ответ)

Импортируйте набор данных `diamonds`. Для загрузки используйте библиотеку `seaborn` и команду `seaborn.load_dataset('diamonds')`. Ответьте какое значение находится в ячейке датасета с индексами 157, 5? Скопируйте ответ в поле ниже.

Ответ: 58.0

(Короткий ответ)

Импортируйте набор данных `diamonds`. Для загрузки используйте библиотеку `seaborn` и команду `seaborn.load_dataset('diamonds')`. Ответьте какое значение находится в ячейке датасета с индексом строки 107, и меткой столбца `'depth'`? Скопируйте ответ в поле ниже.

Ответ: 62.3

(Короткий ответ)

Импортируйте набор данных `diamonds`. Для загрузки используйте библиотеку `seaborn` и команду `seaborn.load_dataset('diamonds')`. Ответьте какое значение находится в ячейке датасета с индексом строки 1409, и меткой столбца `'table'`? Скопируйте ответ в поле ниже.

Ответ: 57.0

(Числовой ответ)

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

(Числовой ответ)

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

(Числовой ответ)

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

Ответ: 1610

(Числовой ответ)

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду `seaborn.load_dataset('diamonds')`. Сколько бриллиантов чистоты VVS1 в этом наборе? За чистоту отвечает столбец 'clarity'. Скопируйте ответ в поле ниже.

Ответ: 3655

(Числовой ответ)

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду `seaborn.load_dataset('diamonds')`. Сколько бриллиантов чистоты VVS2 в этом наборе? За чистоту отвечает столбец 'clarity'. Скопируйте ответ в поле ниже.

Ответ: 5066

(Числовой ответ)

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду `seaborn.load_dataset('diamonds')`. Сколько бриллиантов чистоты SI1 в этом наборе? За чистоту отвечает

столбец 'clarity'. Скопируйте ответ в поле ниже.

Ответ: 13065

(Числовой ответ)

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду seaborn.load_dataset('diamonds'). Сколько бриллиантов чистоты VS2 в этом наборе? За чистоту отвечает столбец 'clarity'. Скопируйте ответ в поле ниже.

Ответ: 12258

(Короткий ответ)

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

Ответ: 1.046137

(Короткий ответ)

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

(Короткий ответ)

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

Ответ: 0.891955

ОПК-2: Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач ОПК-2.2: Разрабатывает программные средства с использованием современных интеллектуальных технологий для решения профессиональных задач
Знать: типы задач, решаемые с использованием методов анализа больших данных и машинного обучения

(Множественный выбор)

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

а. (-100%)

распознавание мелких деталей изображения

b. (-100%)

разделение изображения на фрагменты

c. (50%)

ускорение вычислений

d. (50%)

уменьшение изображения,

(Множественный выбор / Только один ответ)

Имеется нейронная сеть на 10.000 входов. Первый слой – Dense. Сколько нейронов должно быть в этом слое?

a. (0%)

10.000, так как Dense полносвязный слой

b. (0%)

50.000

c. (0%)

20.000

d. (0%)

любое количество нейронов

(Множественный выбор / Только один ответ)

Имеется однослойная нейронная сеть на 6 нейронах. Сколько нейронов содержится в выходном слое?

a. (0%)

12

b. (0%)

1

c. (0%)

любое число меньше либо равно 6

d. (100%)

6

e. (0%)

(Множественный выбор / Только один ответ)

Как называется вектор, своим направлением указывающий направление возрастания некоторой скалярной величины ϕ , (значение которой меняется от одной точки пространства к другой, образуя скалярное поле), а по величине (модулю) равный скорости роста этой величины в этом направлении?

a. (0%)

производная

b. (0%)

перцептрон

c. (100%)

градиент

d. (0%)

среднеквадратичное отклонение

(Множественный выбор / Только один ответ)

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

a. (100%)

полносвязный

b. (0%)

Dropout

c. (0%)

сумматорный

d. (0%)

активационный

(Множественный выбор / Только один ответ)

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

a. (0%)

функция активации

b. (100%)

функция потерь

c. (0%)

сумматорная функция

d. (0%)

синусоида

(Множественный выбор / Только один ответ)

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

a. (0%)

$(-1,1)$

b. (0%)

$[0,1]$

c. (0%)

$[-1,1]$

d. (100%)

$(-\infty, +\infty)$

(Множественный выбор)

Отметьте верные утверждения

a. (50%)

В качестве функции активации нейрона может выступать сигмоидная функция

b. (-100%)

Результат функции активации всегда число от 0 до 1.

c. (50%)

Функция активации получает на вход результат выполнения сумматорной функции

d. (-100%)

В качестве функции активации может выступать только функция ReLU

(МНОЖЕСТВЕННЫЙ ВЫБОР)

Отметьте верные утверждения

a. (-100%)

В искусственном нейроне не менее 5 входов

b. (50%)

аналогом входов искусственного нейрона в естественном нейроне являются дендриты

c. (-100%)

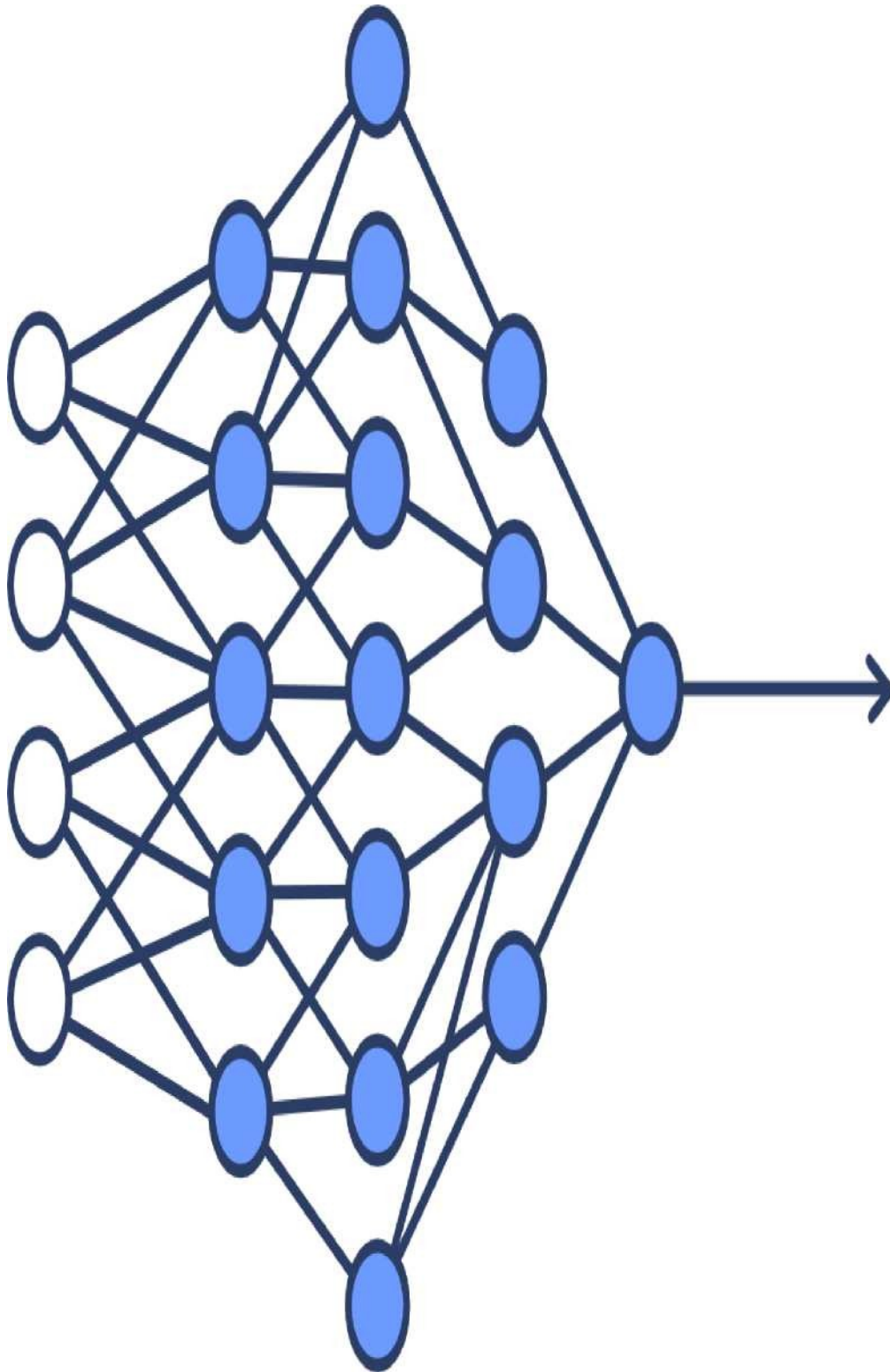
В искусственном нейроне неограниченное количество выходов

d. (50%)

Количество входов и весов в нейроне должно совпадать

(МНОЖЕСТВЕННЫЙ ВЫБОР)

Отметьте верные утверждения, о нейронной сети, которую видите на экране



a. (-100%)

сеть - графическое отображение Перцептрона

c. (50%)

в сети 4 входа

d. (50%)

сеть многослойная

d. (-100%)

все слои сети полносвязные

(Множественный выбор / Только один ответ)

Отметьте верные утверждения:

a. **(100%)**

В выходном слое нейронной сети неограниченное число нейронов

b. **(0%)**

В выходном слое нейронной сети не может быть больше 10 нейронов

c. **(0%)**

В выходном слое нейронной сети всегда находится только один нейрон

d. **(0%)**

Выходной слой нейронной сети не содержит нейронов

(Множественный выбор / Только один ответ)

Что такое Перцептрон?

a. **(100%)**

одна из первых моделей искусственного нейрона

b. **(0%)**

линейная функция активации

c. **(0%)**

название слоя субдескрипции в библиотеке keras

d. **(0%)**

фамилия автора концепции сверточных нейронных сетей, и общее название таких сетей

Уметь: выбирать соответствующие методы решения экспериментальных и теоретических задач

(Множественный выбор / Только один ответ)

В каких задачах для оценки качества модели используется средняя абсолютная ошибка (mae)?

a. (0%)

в задачах кластеризации

b. (0%)

в задачах классификации

c. (100%)

в задачах регрессии

d. (0%)

как в задачах регрессии, так и в задачах классификации

(МНОЖЕСТВЕННЫЙ ВЫБОР)

В каких задачах для оценки качества модели может быть использована средняя квадратичная ошибка (mse)?

a. (-100%)

как в задачах определения вида ириса по размеру лепестка и чашелистника

b. (-100%)

в задаче классификации кошек и собак по фотографии

c. (50%)

в задаче предсказания стоимости драгоценного камня, по его чистоте и весу

d. (50%)

в задаче предсказания размера одобренного кредита по заработной плате, кредитной истории и т.д.

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

Есть набор изображений зеленого горошка, вам необходимо выделить все изображения заболевших растений (важно, чтобы все больные растения были в выборке, попадание туда небольшого количества здоровых растений – не критично). Какую метрику качества модели целесообразнее использовать?

a. (0%)

precision

F1

c. (0%)

b. (0%)

Accuracy

d. (100%)

recall

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

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

a. (0%)

Accuracy

b. (0%)

F1

c. (0%)

precision

d. (100%)

recall

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

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

a. (100%)

precision

b. (0%)

F1

c. (0%)

recall

d. (0%)

Accuracy

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

Модель машинного обучения классифицирует кошек и собак по фото. Модели предъявляется

фотография кошки, и модель классифицирует её как фотографию кошки. Как называется такой исход?

a. (100%)

Либо TP, либо TN, в зависимости от того, какой класс был признан целевым (положительным)

b. (0%)

FN. FalseNegative

c. (0%)

FP. FalsePositive

d. (0%)

Либо FP, либо FN, в зависимости от того, какой класс был признан целевым (положительным)

(МНОЖЕСТВЕННЫЙ ВЫБОР)

Отметьте верные утверждения о метрике качества ассигасу...

a. (-100%)

ассигасу показывает долю объектов, названных классификатором положительными и при этом действительно являющихся положительными

b. (50%)

ассигасу показывает отношение верно классифицированных наблюдений к общему количеству наблюдений

c. (-100%)

ассигасу особенно полезна в задачах с несбалансированными классами - когда одних значений (ядовитых грибов, например) значимо меньше, чем других (съедобных грибов).

d. (50%)

ассигасу используется преимущественно в задачах классификации, но не регрессии

(Числовой ответ)

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

(Числовой ответ)

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

Ответ: 19479

(Числовой ответ)

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду `seaborn.load_dataset('diamonds')`. Отберите только бриллианты цвета «J» или огранки «Premium». Сколько строк в полученном наборе.

Ответ: 15791

(Числовой ответ)

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

Ответ: 808

Владеть: навыком критически анализировать современные проблемы инноватики с использованием анализа больших данных и машинного обучения

(Числовой ответ)

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду `seaborn.load_dataset('diamonds')`. 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.

Ответ: 344

(Числовой ответ)

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

Скопируйте ответ в поле ниже.

Ответ: 345

(Числовой ответ)

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

(Короткий ответ)

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

(Короткий ответ)

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

Ответ: 61.2

(Короткий ответ)

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

Ответ: 58.4

ОПК-7: Способен разрабатывать и применять математические модели процессов и объектов при решении задач анализа и синтеза распределенных информационных систем и систем поддержки принятия решений ОПК-7.2: Применяет математические модели процессов и объектов при решении прикладных задач Знать: современные технологии анализа больших данных и машинного обучения

(Множественный выбор / Только один ответ)

За что отвечает параметр `n_estimator`?

a. (0%)

количество примеров в листе

b. (100%)

количество деревьев в модели «Случайный лес»

c. (0%)

максимальная глубина дерева в модели «Случайный лес»

d. (0%)

количество признаков, которые рассматривает каждое отдельное дерево в модели «Случайный лес»

(Множественный выбор / Только один ответ)

Как называется алгоритм, который заключается в использовании большого ансамбля решающих деревьев?

a. (0%)

дерево классификации (Classifier Tree)

b. (0%)

рандомный бор (Random Pine Forest)

c. (100%)

случайный лес (Random Forest)

d. (0%)

дерево регрессии (Regressor Tree)

(Множественный выбор / Только один ответ)

Какой из предложенных методов используется для предсказания модели?

a. (100%)

predict

b. (0%)

fit

c. (0%)

score

d. (0%)

decision_path

(МНОЖЕСТВЕННЫЙ ВЫБОР)

Отметьте верные утверждения о количестве признаков, которые рассматривает каждое отдельное дерево в модели «Случайный лес»?

a. (-100%)

в задачах классификации это значение по умолчанию равно половине всех имеющихся признаков

b. (-100%)

это значение всегда должно быть больше глубины леса

c. (50%)

чем больше это значение, тем дольше настраивается модель

d. (-100%)

чем больше это значение, тем разнообразнее деревья

e. (50%)

в задачах регрессии это значение по умолчанию равно трети всех имеющихся признаков

(МНОЖЕСТВЕННЫЙ ВЫБОР)

Отметьте верные утверждения о случайном лесе

a. (33.33333%)

время настройки и работы случайного леса увеличивается пропорционально количеству деревьев в лесу.

b. (-100%)

количество рассматриваемых признаков всегда равно половине известных признаков

c. (-100%)

случайный лес в общем случае менее точен, чем дерево решений

d. (33.33333%)

случайный лес применим для решения задач регрессии

e. (33.33333%)

чем больше деревьев в лесу, тем лучше качество предсказания

(Множественный выбор / Только один ответ)

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

a. (0%)

для обучения используется вся обучающая выборка

b. (0%)

для обучения каждого дерева используется строго независимая (без возвращения) подвыборка обучающей выборки

c. (100%)

для обучения используется подвыборка обучающей выборки

d. (0%)

для обучения всегда используется половина обучающей выборки

(Множественный выбор)

Отметьте верные утверждения об алгоритме случайного леса - Random Forest.

a. (-100%)

Число деревьев в лесу всегда равно удвоенному количеству предсказываемых классов

b. (-100%)

Итоговым предсказанием модели является предсказание случайного дерева

c. (-100%)

Параметры для каждого дерева (глубина, минимальное число образцов в листе и т.д.) выбираются случайно

d. (50%)

Предсказание модели в задачах регрессии - усреднённые предсказания деревьев

e. (50%)

Каждое дерево в лесу получает случайный поднабор данных

(Числовой ответ)

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

Ответ: 4406695

(КОРОТКИЙ ОТВЕТ)

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

- o 3538.9144201968334
- o 3538.914420

(КОРОТКИЙ ОТВЕТ)

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду `seaborn.load_dataset('diamonds')`. Отберите только бриллианты, весящие больше 2-х карат и огранки «Premium». Какова их средняя стоимость? За вес отвечает столбец 'carat', за огранку 'cut'.

Ответ: 15589.128964

(КОРОТКИЙ ОТВЕТ)

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'.

- o 9636.00
- O 9636

(КОРОТКИЙ ОТВЕТ)

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

Уметь: применять методы и подходы анализа больших данных и машинного обучения для решения экспериментальных и теоретических задач

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

За что отвечает параметр `max_features`?

a. (0%)

количество деревьев в модели «Случайный лес»

b. (0%)

максимальная глубина дерева в модели «Случайный лес»

c. (100%)

количество признаков, которые рассматривает каждое отдельное дерево в модели «Случайный лес»

d. (0%)

количество примеров в листе

(Числовой ответ)

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

(Числовой ответ)

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду `seaborn.load_dataset('diamonds')`. Отсортируйте набор по убыванию цены. Сколько стоит третий бриллиант в наборе?

Ответ: 18806

(Короткий ответ)

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 67. Чему равно значение столбца carat в этой строке?

Ответ: 0.2

(Короткий ответ)

Импортируйте набор данных diamonds. Для загрузки используйте библиотеку seaborn и команду `seaborn.load_dataset('diamonds')`. Отсортируйте набор по убыванию цены. Сколько весит (столбец 'carat') пятый бриллиант в наборе?

2.00

2.0

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

Дерево решений —

a. (0%)

свойство интеллектуальных систем выполнять отдельные функции интеллекта человека, в том числе творческие;

b. (0%)

метод решения задач, точный алгоритм решения которых заранее известен.

c. (100%)

средство поддержки принятия решений, использующееся в машинном обучении, анализе данных и статистике, представляет собой иерархическую структуру правил вида «если - то»

d. (0%)

разнообразные данные, которые поступают с постоянно растущей скоростью и объем которых постоянно растет;

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

Как в деревьях решений называется удаление листов и узлов, потеря которых минимально скажется на точности модели и увеличении ошибки?

a. (0%)

ранняя остановка

b. (0%)

ограничение глубины дерева

c. (100%)

отсечение ветвей

d. (0%)

обучение модели

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

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?

a. (100%)

переобучение

b. (0%)

кроссвалидация

c. (0%)

недообучение

d. (0%)

решающее правило

(Множественный выбор / Только один ответ)

Какой атрибут отвечает за минимальное число образцов в листе?

a. (0%)

min_impurity_decrease

b. (0%)

max_depth

c. (0%)

min_samples_split

d. (100%)

min_samples_leaf

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

Какой атрибут отвечает за минимальное число образцов в узле для разделения?

a. (0%)

max_depth

b. (0%)

min_samples_leaf

c. (100%)

min_samples_split

d. (0%)

min_impurity_decrease

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

Какой из предложенных ниже методов отвечает за обучение модели?

a. (100%)

fit

b. (0%)

get_depth

predict

d. (0%)

score

c. (0%)

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

Какой классификатор необходимо использовать для предсказания размера одобренного кредита по заданным параметрам?

a. (0%)

Query

b. (100%)

DecisionTreeRegressor

c. (0%)

Может быть использован и DecisionTreeClassifier, и DecisionTreeRegressor

d. (0%)

DecisionTreeClassifier

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

На представленном ниже дереве решений закрыто одно значение. Какое?

a. (-100%)

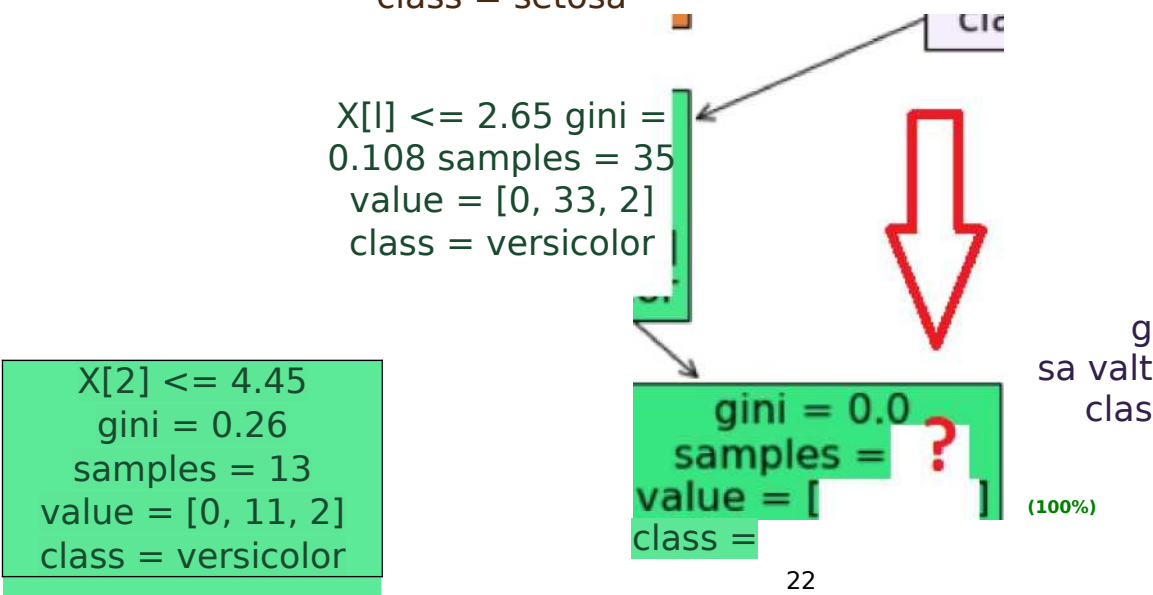
```

X[2] <= 2.45
gini = 0.665
samples = 112
value = [41. 34. 37]
class = setosa
gini = 0.499

gini = 0.0
samples = 41
value = [41, 0. 0]
class = setosa

samples = 71
value = [0, 34. 37]
= virginica

```



- b. (0%)
- 14
- c. (0%)
- 41
- d. (0%)
- 39

(Множественный выбор)

Отметьте верные утверждения о влиянии параметров решающего дерева на переобучение.

- Чем меньше значение min_samples_split, тем меньше тенденция к переобучению
- b. (-100%)
- Чем меньше значение min_samples_leaf, тем меньше тенденция к переобучению

c. (50%)

Чем меньше значение параметра `max_leaf_nodes`, тем меньше тенденция к переобучению

d. (-100%)

Глубина дерева никак не влияет на его переобученность

e. (50%)

Чем меньше глубина дерева, тем меньше тенденция к переобучению

(МНОЖЕСТВЕННЫЙ ВЫБОР)

Отметьте верные утверждения о деревьях решений.

a. (50%)

свойство `samples` указывает на количество примеров в узле

b. (-100%)

свойство `gini` определяется только для листьев, но не для узлов

c. (50%)

чем меньше свойство `gini`, тем однороднее примеры в листе

d. (-100%)

параметр `gini`, во всех листьях должен иметь одинаковые значения

(МНОЖЕСТВЕННЫЙ ВЫБОР)

Отметьте верные утверждения о подборе оптимальных параметров модели

a. (-100%)

`RandomizedSearchCV` дает более точное предсказание, чем `GridSearchCV`

b. (50%)

`GridSearchCV` находит лучшую комбинацию параметров для дерева, из тех значений параметров, что были переданы

c. (-100%)

`RandomizedSearchCV` находит лучшую комбинацию параметров для дерева решений

d. (50%)

a. (-100%)

RandomizedSearchCV работает быстрее, чем GridSearchCV

(МНОЖЕСТВЕННЫЙ ВЫБОР)

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. (-100%)

Модель идеально настроена

b. (-100%)

Модель недообучена

c. (50%)

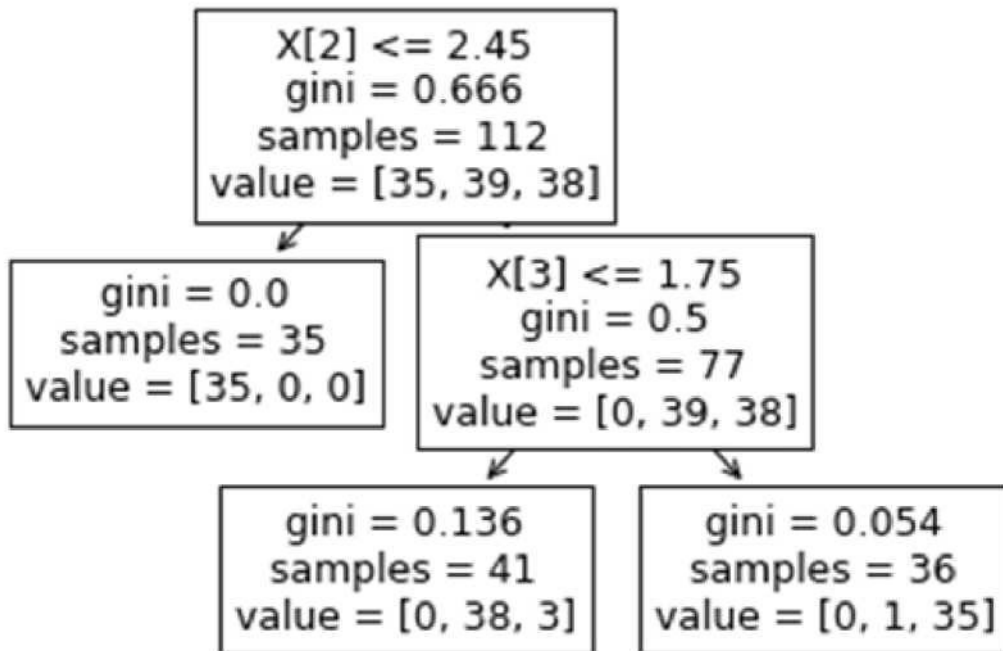
Модель переобучена

d. (50%)

Распределение данных в тестовой выборке сильно отличается от распределения в тренировочной

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

Сколько листьев в указанном ниже дереве?



a. (0%)

4

b. (0%)

5

c. (0%)

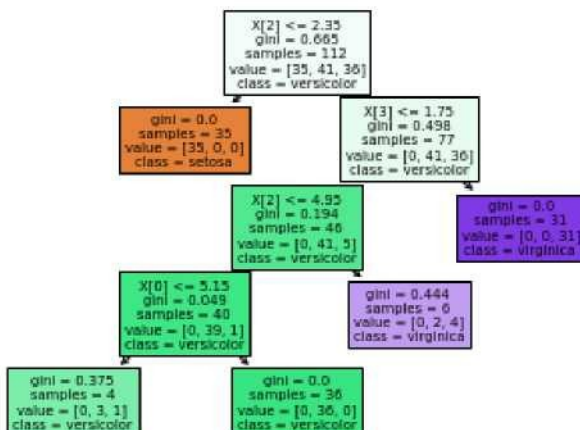
2

d. (100%)

3

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

Сколько листьев в указанном ниже дереве?



a. (0%)

4

b. (0%)

3

c. (0%)

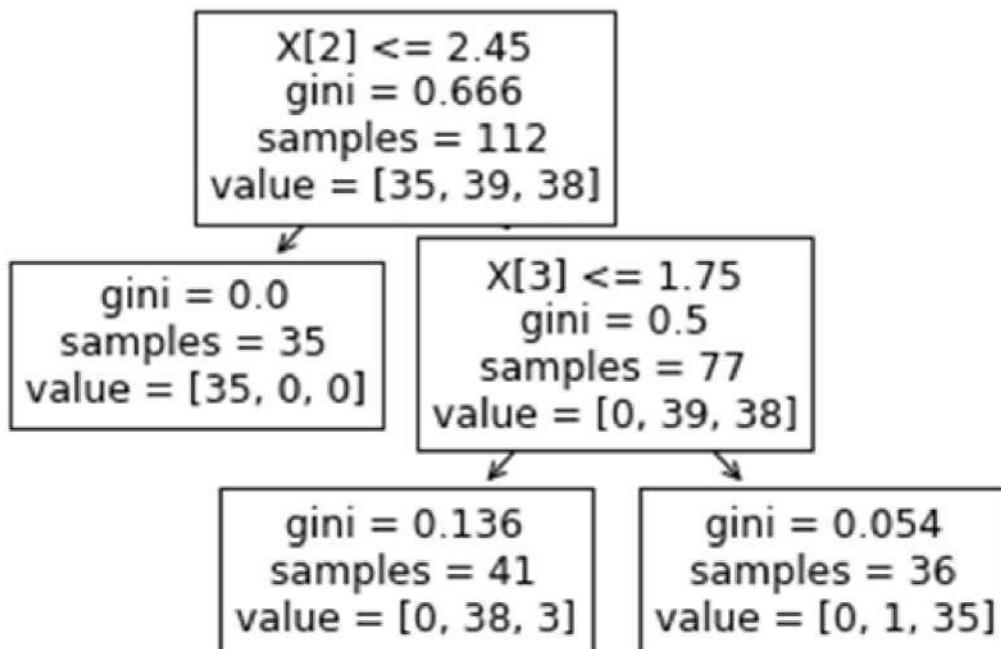
9

d. (100%)

5

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

Сколько узлов в указанном ниже дереве?



a. (0%)

4

b. (100%)

2

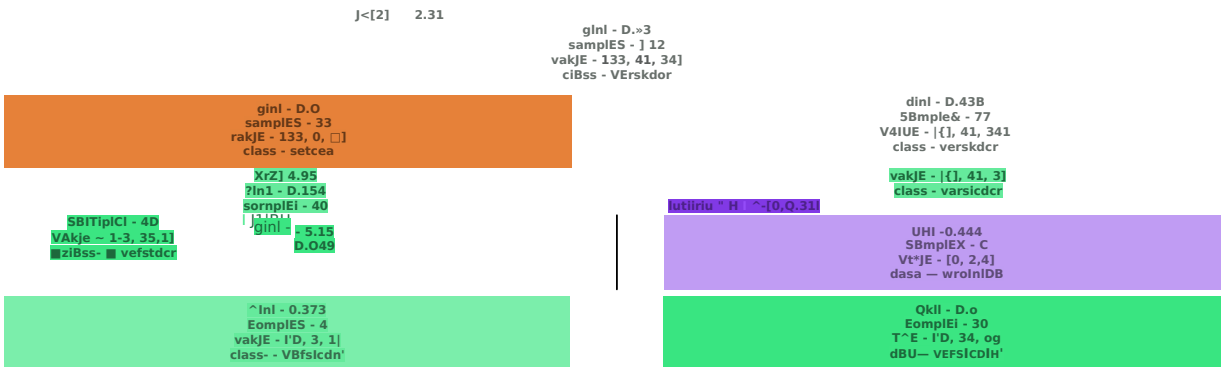
c. (0%)

3

d. (0%)

(МНОЖЕСТВЕННЫЙ ВЫБОР / ТОЛЬКО ОДИН ОТВЕТ)

Сколько узлов в указанном ниже дереве?



C. (-100%)

a. (0%)

5

b. (0%)

9

c. (0%)

2

d. (100%)

(МНОЖЕСТВЕННЫЙ ВЫБОР)

Среди предложенных задач машинного обучения укажите задачи регрессии.

a. (-100%)

Поиск негативных отзывов о фильме на сайте кинокомпании

b. (-100%)

Поиск мошеннических транзакций

Алгоритм фильтрации спама

d. (50%)

Предсказание срока окупаемости проекта

e. (50%)

Предсказание рыночной стоимости квартиры

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/зачтено» – ставится за работу, выполненную полностью без ошибок и недочетов.

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

«Удовлетворительно/зачтено» – 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/не зачтено» – 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» - обучающийся 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.

«Не зачтено» - выставляется в том случае, когда обучающийся 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