

Машинное обучение: квиз 0

*Обязательный вопрос

1. Электронная почта *

2. Какой из этих этапов предобработки избыточен для **глубокого** обучения? * 1 балл

Отметьте только один овал.

- ☐ Токенизация
- ☐ Очистка текста от шума
- ☐ Конструирование новых признаков
- ☐ Разбиение на тренировочные и тестовые данные

3. Сопоставьте задачу и метод машинного обучения: * 2 балла

Задача 1: имея корпус новостей, разделенных по тематикам, обучить модель определять тему новости;

Задача 2: имея неаннотированный корпус новостных текстов, обучить модель разделять их на тематические кластеры.

Отметьте только один овал в каждом ряду.

	Обучение без учителя	Обучение с учителем
Задача 1	<input type="radio"/>	<input type="radio"/>
Задача 2	<input type="radio"/>	<input type="radio"/>

4. Сопоставьте конкретную задачу и класс задач:

★ 4 балла

Задача 1: предскажите среднее время чтения текста;

Задача 2: предскажите пол автора текста;

Задача 3: визуализируйте эмбединги ключевых слов вашего текста (вы уже нашли ключевые слова);

Задача 4: объедините способности нескольких моделей для решения одной и той же проблемы.

Отметьте только один овал в каждом ряду.

	Классификация	Ансамблевое обучение	Снижение размерности	Регрессия
Задача 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Задача 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Задача 3	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Задача 4	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5. Рассмотрите псевдокод на картинке. Считайте, что X - корпус текстов. * 1 балл
Зависимая переменная включает только те классы, которые вы видите на экране. Какую ошибку вы заметили?

(train_test_split - это функция из библиотеки sklearn, которая делит выборку на тренировочную и тестовую части.)

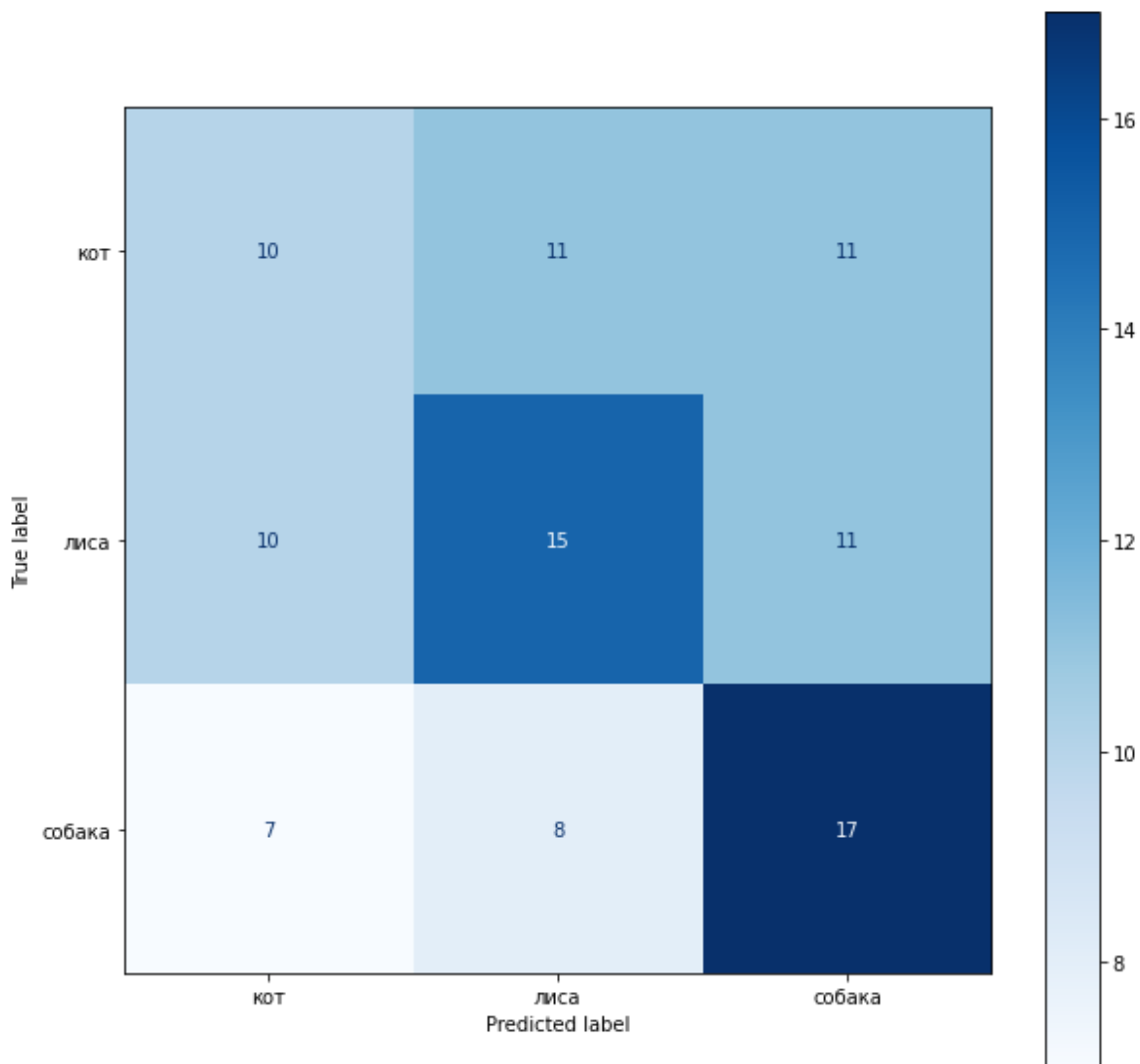
```
1 X = [.....]
2 y = [0, 1, 0, 0.....]
3
4
5 tf_idf = TfidfVectorizer()
6 X_transformed = tf_idf.fit_transform(X)
7
8 X_train, y_train, X_test, y_test = train_test_split(X_transformed, y)
9
10 lr = LogisticRegression()
11 lr.fit(X_train)
```

Отметьте только один овал.

- ☐ Нужно было закодировать лейблы
- ☐ Строка 6: следовало применять метод fit векторайзера только на тренировочной выборке
- ☐ Строка 10: нужно выбрать другую модель

6. Посмотрите на приведенную матрицу путаницы и выберите ВСЕ ВЕРНЫЕ утверждения. Не забудьте: истинные лейблы классов находятся на оси Y, предсказанные моделью - на оси X.

* 3 балла



Отметьте все подходящие варианты.

- ☐ Модель лучше всего определяет, что на картинке собака
- ☐ Модель чаще принимает лис за собак, чем за котов
- ☐ Модель чаще принимает лис за собак, чем собак за лис
- ☐ Модель реже принимает котов за лис, чем за собак

Компания Google не имеет никакого отношения к этому контенту.

Google Формы

