

ДПО МЛ 4 - Кластеризация

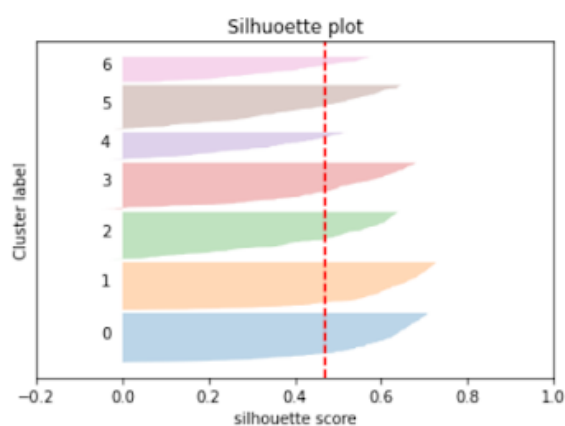
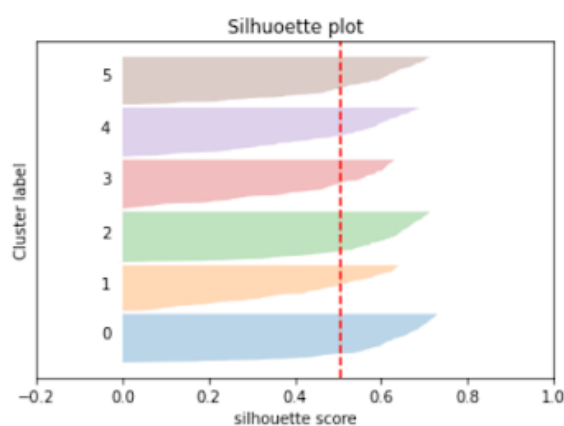
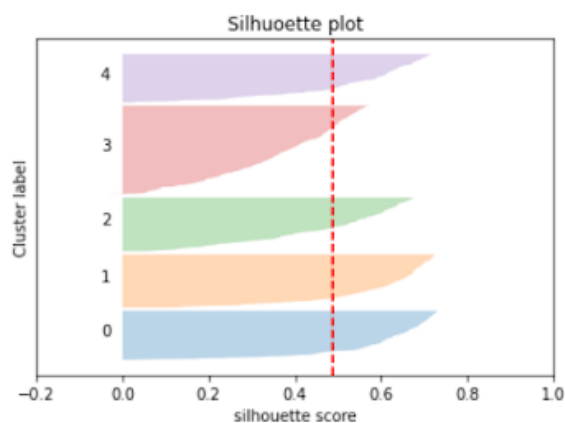
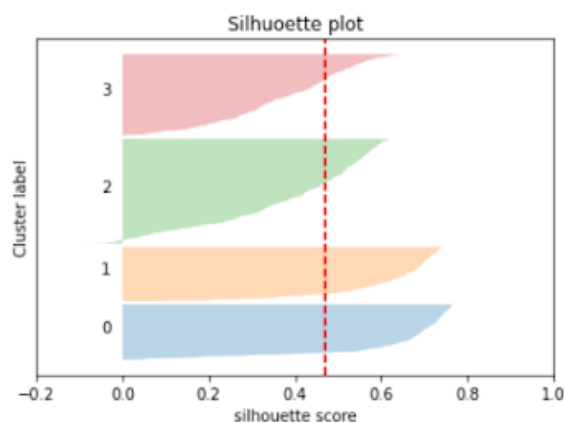
*Обязательный вопрос

1. Электронная почта *

2. Ваше имя и фамилия *

3. Определите оптимальное количество кластеров в данном случае, ориентируясь на силуэты:

* 1 балл



Отметьте только один овал.

- ☐ 4
- ☐ 5
- ☐ 6
- ☐ 7

4. К вам попала чья-то оцифрованная домашняя библиотека. Никакой информации о книгах, кроме собственно текстов, не сохранилось. Вы хотите кластеризовать полученные тексты. Какие метрики качества кластеризации вы используете?

* 1 балл

Отметьте все подходящие варианты.

- ☐ Коэффициент силуэта
- ☐ Гомогенность и полнота
- ☐ Adjusted Rand Index
- ☐ V-мера

5. У какого из этих алгоритмов есть центральные точки? Выберите все правильные ответы. ★ 1 балл

Отметьте все подходящие варианты.

- ☐ KMeans
☐ Mini-batch KMeans
☐ DBSCAN
☐ Agglomerative Clustering

6. Посмотрите на предложенный фрагмент кода. Какую ошибку вы заметили? ★ 1 балл

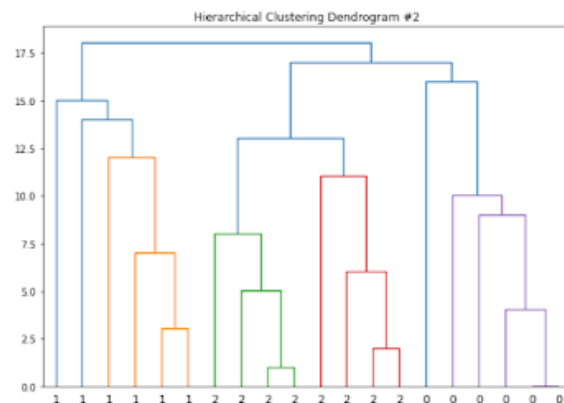
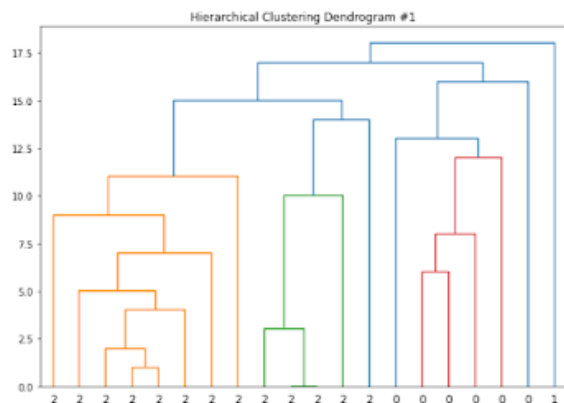
```
1 from sklearn.datasets import make_circles
2 from sklearn.cluster import DBSCAN
3 from sklearn.metrics import adjusted_rand_score
4
5
6 X, y = make_circles()
7 dbscan = DBSCAN()
8 dbscan.fit(X)
9 print(adjusted_rand_score(y, dbscan.predict(X)))
```

Отметьте только один овал.

- ☐ Здесь нельзя применять ARI
☐ У DBSCAN нет метода .predict()
☐ Не задано количество кластеров
☐ Предсказания происходят на всей выборке

7. Эти дендрограммы созданы из одних и тех же данных с применением разных методик измерения расстояния между кластерами. Сопоставьте дендрограмму и ее метрики гомогенности (H), полноты (C) и v-меры (V). Напоминание: все эти метрики имеют шкалу от 0 до 1, где 1 - идеальная кластеризация.

* 2 балла



Отметьте только один овал в каждом ряду.

	Dendrogram #1	Dendrogram #2
H: 0.739,	<input type="radio"/>	<input type="radio"/>
C: 0.744,	<input type="radio"/>	<input type="radio"/>
V: 0.742	<input type="radio"/>	<input type="radio"/>
H: 0.467,	<input type="radio"/>	<input type="radio"/>
C: 0.648,	<input type="radio"/>	<input type="radio"/>
V: 0.543	<input type="radio"/>	<input type="radio"/>

Компания Google не имеет никакого отношения к этому контенту.

Google Формы

