

## Листок 02. Задача кластеризации

Н.В. Артамонов

3 декабря 2024 г.

**Замечание** обязательно проводим предварительную обработку данных: удаление пропущенных значений, нормировку, преобразование категориальных признаков

**#1.** Для набора данных `countries` проведите разбиение на кластеры следующими методами:

Число кластеров	Метод
3	k-средних
4	k-средних
5	k-средних
3	иерархическая
4	иерархическая
5	иерархическая

Визуализируйте разбиение на кластеры на диаграмме рассеяния в переменных датасета

**#2.** Для набора данных `countries` найдите «оптимальное» число кластеров для метода

1. k-средних

2. иерархической кластеризации

относительно метрик: Silhouette, Calinski-Harabasz, Davies-Bouldin

**#3.** Из набора данных `sleep75` возьмите переменные `sleep`, `totwrk`, `age`, `educ` и проведите разбиение на кластеры следующими методами:

Число кластеров	Метод
3	k-средних
4	k-средних
5	k-средних
3	иерархическая
4	иерархическая
5	иерархическая

Визуализируйте разбиение на кластеры на диаграмме рассеяния в переменных датасета

#4. Из набора данных **sleep75** возьмите переменные **sleep**, **totwrk**, **age**, **educ** и найдите «оптимальное» число кластеров для метода

1. k-средних
2. иерархической кластеризации

относительно метрик: Silhouette, Calinski-Harabasz, Davies-Bouldin

#5. Для набора данных **Labour** проведите разбиение на кластеры следующими методами:

Число кластеров	Метод
3	k-средних
4	k-средних
5	k-средних
3	иерархическая
4	иерархическая
5	иерархическая

Визуализируйте разбиение на кластеры на диаграмме рассеяния в переменных датасета

#6. Для набора данных **Labour** найдите «оптимальное» число кластеров для метода

1. k-средних
2. иерархической кластеризации

относительно метрик: Silhouette, Calinski-Harabasz, Davies-Bouldin