

## Лабораторная работа №2 - Кластеризация

Наборы данных: lab2\_blobs.csv, lab2\_luckyset.csv, lab2\_circles.csv

**В пунктах 1-4 исследуются наборы данных, отмеченные цифрами вашего вариант.**

### 1. Подготовка наборов данных:

- 1.1. Загрузите наборы.
- 1.2. Проверьте корректность загрузки.
- 1.3. Постройте диаграмму рассеяния набора данных. Опишите форму данных.
- 1.4. Подготовьте наборы данных проведя стандартизация или нормировку данных. Обоснуйте выбор операции.

### 2. K-Means:

- 2.1. Проведите исследование оптимального количества кластеров методом локтя. Сделайте выводы, о наиболее подходящем количестве кластеров.
- 2.2. Проведите исследование оптимального количества кластеров методом силуэта. Сделайте выводы, о наиболее подходящем количестве кластеров.
- 2.3. Проведите кластеризацию алгоритмом K-means, с выбранным оптимальным количеством кластеров.
- 2.4. Постройте диаграмму рассеяния результатов кластеризации с выделением разным цветом разных кластеров.
- 2.5. Постройте диаграмму Вороного для результатов кластеризации. На диаграмме отметьте центроиды полученных кластеров.
- 2.6. Постройте для каждого признака диаграмму “box-plot” или “violin-plot”, с разделением по кластерам. Сделайте выводы о разделении кластеров и успешности применения кластеризации K-means к набору данных.
- 2.7. Рассчитайте для каждого кластера кол-во точек, среднее, СКО, минимум и максимум. Сопоставьте результаты с построенными графиками.

### 3. DBSCAN:

- 3.1. Подберите параметры алгоритма DBSCAN, которые на ваш взгляд дают наилучшие результаты. Опишите процесс (почему и как изменяли параметры) подбора параметров.
- 3.2. Постройте диаграмму рассеяния результатов кластеризации с выделением разным цветом разных кластеров.
- 3.3. Сделайте выводы об успешности кластеризации.

### 4. Иерархическая кластеризация:

- 4.1. Проведите иерархическую кластеризацию при всех возможных параметрах linkage, используя количество кластеров полученных в п.2 или п.3. Для каждого из результатов постройте дендрограмму. Сделайте выводы, о разделении кластеров и необходимости изменить количество кластеров (если считаете, что необходимо изменить количество кластеров, то повторите кластеризацию с другим количеством кластеров).
- 4.2. Постройте диаграмму рассеяния результатов кластеризации с выделением разным цветом разных кластеров. Используйте лучшие результаты, полученные для определенного параметра linkage.
- 4.3. Сравните результаты кластеризации с результатами полученными в п.2 и п.3. Сделайте выводы о том, какой метод кластеризации подходит под каждый из наборов данных.