**Лабораторная работа 2**

1. Решите задачу 1 (Кластеризация.ipynb).
2. Задача 2

Умная камера фотосепаратора (аппарата для сортировки сыпучих продуктов) настроена на выявление признаков подсолнечника, ячменя, пшеницы и гороха.

В файле «Train.csv» представлены признаки 4-х видов сыпучих продуктов:

* **Area** – площадь объекта;
* **MajorAxisLenght** – длина главной оси инерции;
* **MinorAxisLenght** – длина побочной оси инерции;
* **Eccentricity** – экстрецесинтет объекта;
* **Orientation** – угол ориентации объекта;
* **ConvexArea** – площадь выпуклой оболочки, построенной над объектом;
* **FilledArea** – заполненность;
* **EulerNumber** – число Эйлера;
* **EquivDiameter** – эквивалентый диметр;
* **Solidity** – твердость;
* **Extent**;
* **Perimeter** – периметр объекта;

**Целевая переменная Class** – класс объекта, принимает 4 значения.

Файл «Train.csv» содержит признаки объектов и их маркировку.

Ваша задача состоит в том, чтобы по имеющимся признакам разделить продукты по видам: ячмень, подсолнечник, пшеница и горох.

1. Загрузить данные из файла Train.csv в DataFrame.
2. Удалить столбец Class
3. Загрузить данные из файла Test.csv в DataFrame.
4. Объединить оба DataFrame в один.
5. Вывести статистическую информацию о наборе данных.
6. Вывести названия столбцов.
7. Заменить или удалить пропущенные значения.
8. Заменить категориальные данные количественными.
9. Выбрать важные признаки для кластеризации.
10. Решить задачу кластеризации, используя методы К-средних и DBSAN. Оптимальное количество кластеров определить методом локтя.

from sklearn.cluster import DBSCAN, KMeans

kmeans = KMeans(n\_clusters = 4, init = 'k-means++', max\_iter = 2000, n\_init = 10, random\_state = 42)

model\_dbscan\_s = DBSCAN(eps=0.01, min\_samples=100)

1. Проведите оценку полученных моделей методом силуэта.

from sklearn.metrics import silhouette\_score

print(silhouette\_score(X=x, labels=y\_pred)) # силуэт

1. Сделайте выводы