Министерство науки и высшего образования Российской Федерации **Федеральное государственное автономное образовательное учреждение  
высшего образования**

**«ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
(национальный исследовательский университет)**

Высшая школа электроники и компьютерных наук

Кафедра системного программирования

**ОТЧЁТ ПО ЗАДАНИЮ №8**

по дисциплине «Интеллектуальный анализ больших данных»

Тема: Разделительная кластеризация

Выполнил

студент группы КЭ-120

Глизница Максим Николаевич

E-mail: letadllo@mail.ru

Челябинск 2021

# Задание

Выполните кластеризацию набора 2-х или 3-мерных данных с помощью алгоритма k-Means (предполагается, что полученные кластеры будут выпуклыми), используя различные значения параметра 𝑘 (из интервала 3..9). Выполните визуализацию полученных результатов в виде точечных графиков, на которых цвет точки отражает принадлежность кластеру.

Внесите шум в набор данных (случайным образом изменить определенную долю объектов набора: 1%, 3%, 5%, 10%; изменение может заключаться в добавлении/вычитании к/из одной/нескольких координат объекта случайного числа). Выполните кластеризацию зашумленного набора данных с помощью алгоритмов k-Means и k-Medoids (или PAM), используя различные значения параметра 𝑘 (из интервала 3..9). Выполните визуализацию полученных результатов в виде точечных графиков, на которых цвет точки отражает принадлежность кластеру.

Выполните кластеризацию набора данных из задания 9 (с невыпуклыми кластерами) с помощью алгоритмов k-Means и k-Medoids (или PAM), используя различные значения параметра 𝑘 (из интервала 3..9). Выполните визуализацию полученных результатов в виде точечных графиков, на которых цвет точки отражает принадлежность кластеру.

# Краткие сведения о наборах данных

Использованный набор данных:

# Краткие сведения о средствах реализации

Для реализации методов были использованы библиотеки scikit-learn и scikit-learn-extra, включающие в себя множество алгоритмов для анализа данных.

Репозиторий по дисциплине: <https://github.com/Airpllane/DAAlgorithms>. Каталог для задания: 8. SepCluster

# Визуализация показателей качества