

МГТУ им. Н. Э. Баумана, кафедра ИУ5  
курс “Технологии машинного обучения”

Рубежный контроль №2  
«Технологии использования и оценки моделей  
машинного обучения»

ВЫПОЛНИЛ:

Ерохин И.А.

Группа: ИУ5-61Б

Вариант: 8

ПРОВЕРИЛ:

Гапанюк Ю.Е.

Москва 2020

## Задача 2. Кластеризация данных:

Кластеризуйте данные с помощью двух алгоритмов кластеризации (варианты по группам приведены в таблице):

Группа	Алгоритм №1	Алгоритм №2
ИУ5-61Б, ИУ5Ц-81Б	K-Means	DBSCAN

Сравните качество кластеризации с помощью следующих метрик качества кластеризации (если это возможно для Вашего набора данных):

1. Adjusted Rand index Adjusted Mutual Information Homogeneity,
2. completeness,
3. V-measure,
4. Коэффициент силуэта.

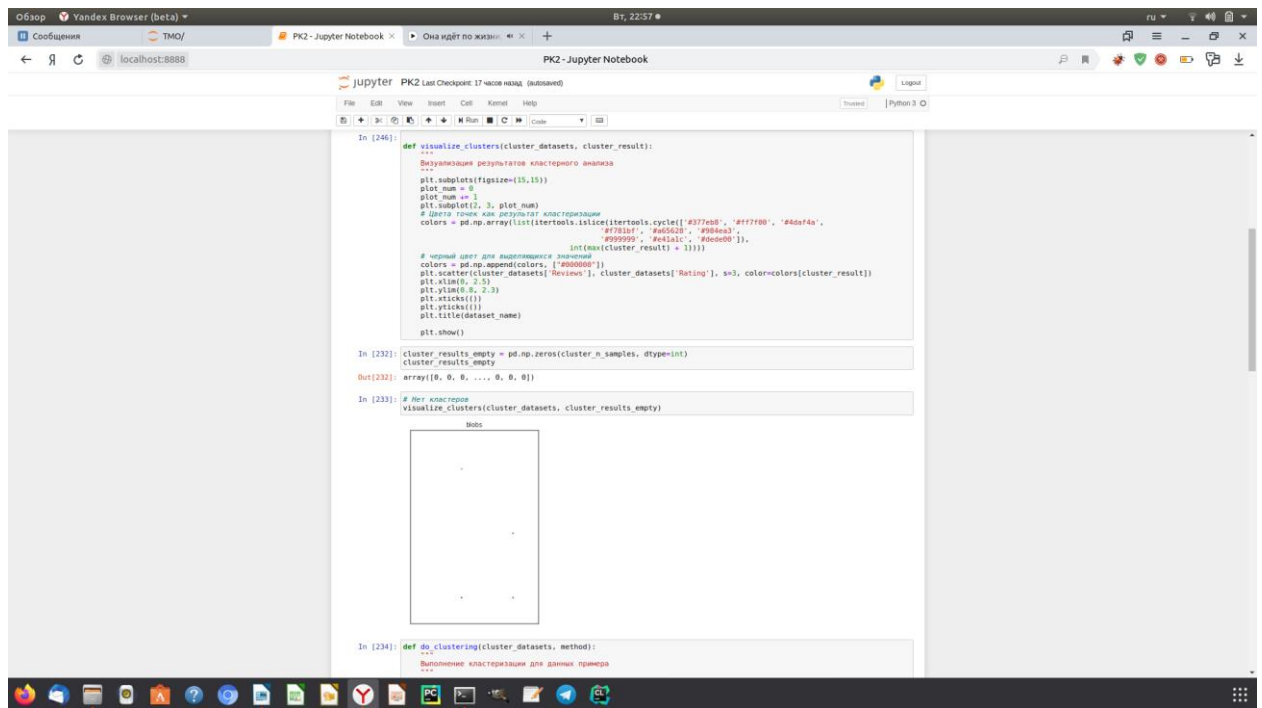
Сделайте выводы о том, какой алгоритм осуществляет более качественную кластеризацию на Вашем наборе данных.

## Набор данных (№8)

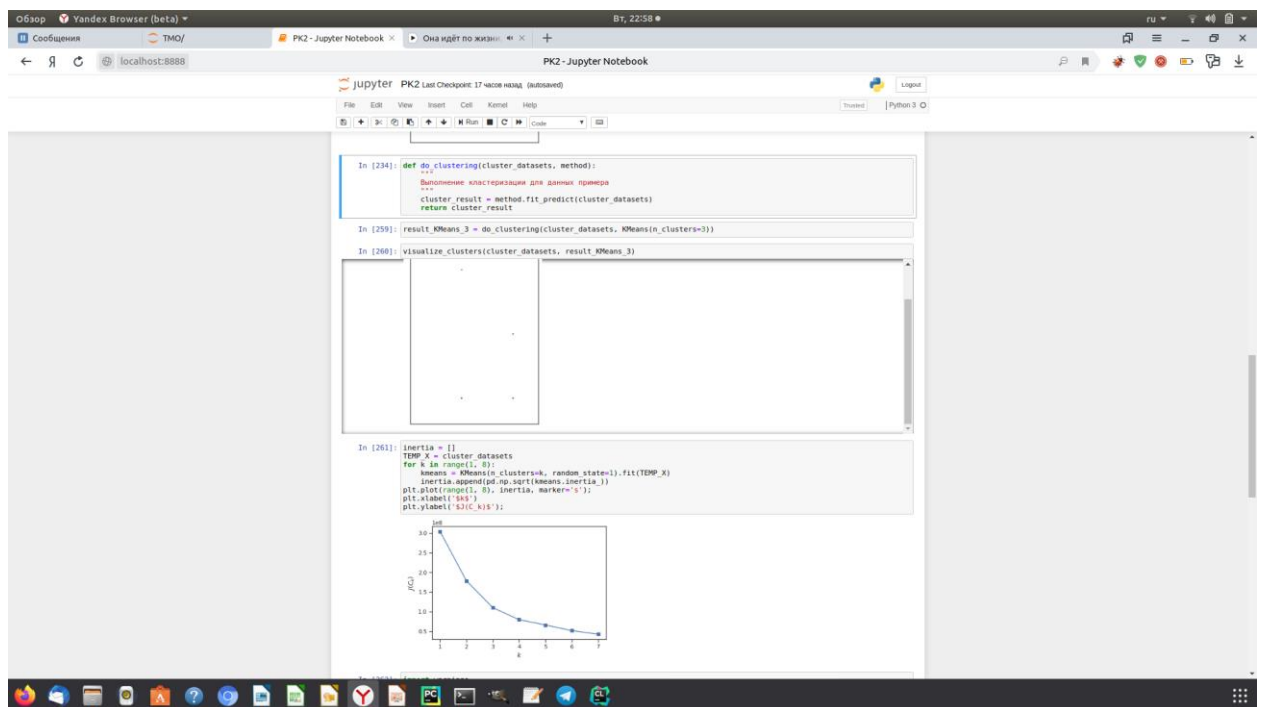
### Выполненная работа:

- ### 1. Первичный анализ датасета:

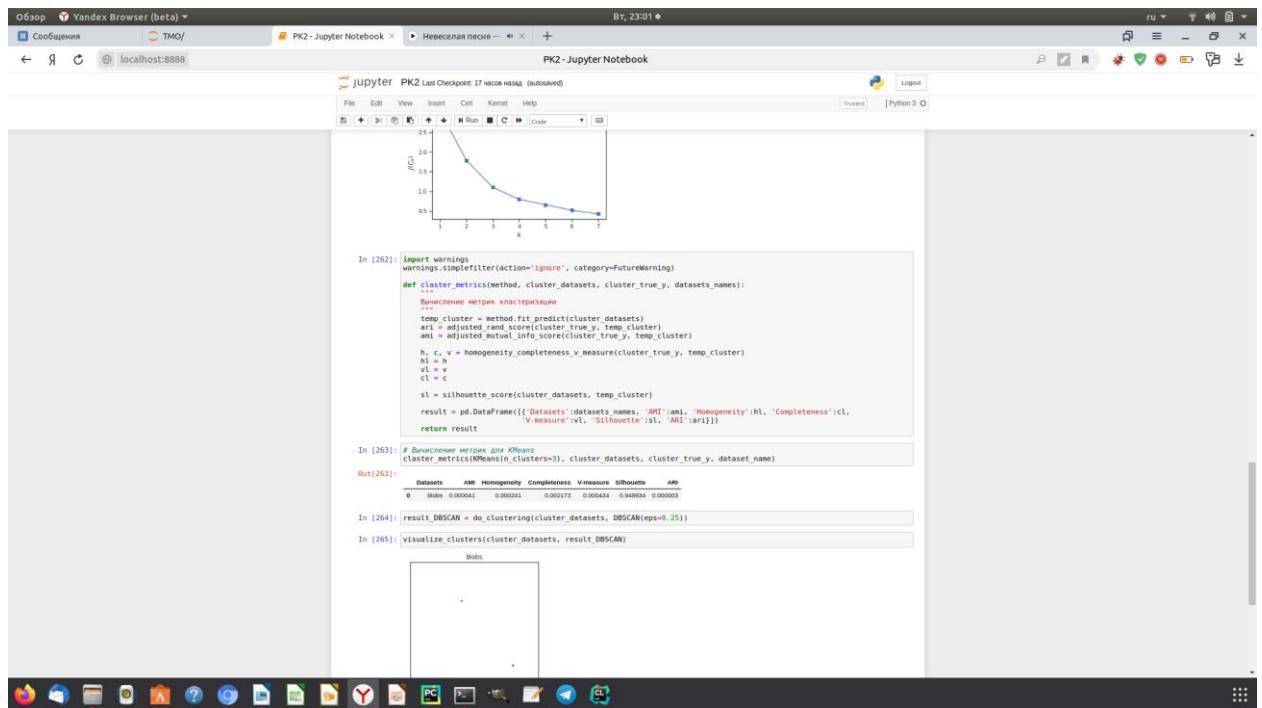
[illegible]



## 2. Кластеризация данных с помощью алгоритма K-Means:

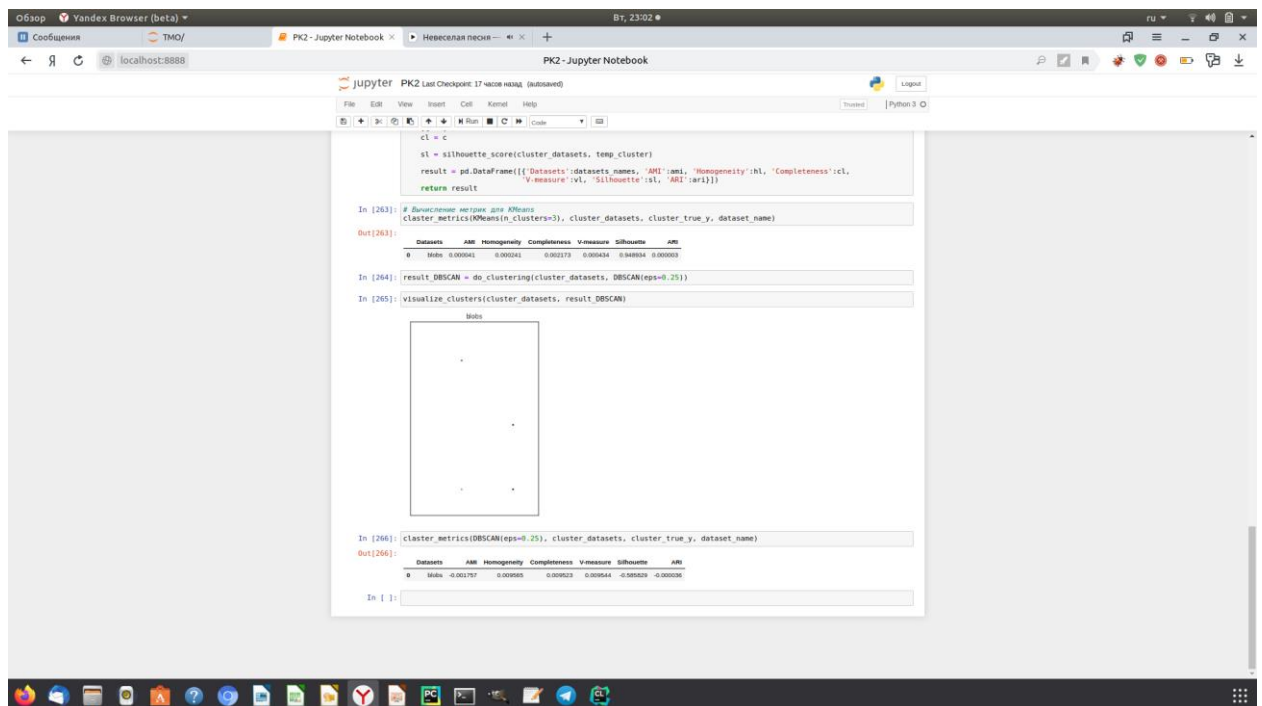


В качестве гиперпараметра К (количество кластеров) было выбрано число 4.



### 3. Кластеризация данных с помощью алгоритма DBSCAN:

Зададим параметр eps (максимальный радиус соседства) равным 0.25. В результате получим:



В данной работе были применены два алгоритма кластеризации данных: K-Means и DBSCAN. Анализируя результаты, нельзя сказать, какой из двух алгоритмов обеспечивает более качественную кластеризацию. Полученные метрики обоих алгоритмов, как видно, имеют одинаково плохие значения, что говорит о некачественном исходном наборе данных.