

КЛАСТЕРИЗАЦИЯ ПОЛЬЗОВАТЕЛЕЙ НА ОСНОВЕ ИХ ОТЗЫВОВ

Афанасьева Александра
Осипенкова Юлия

ЦЕЛЬ РАБОТЫ

2

Научиться определять людей со
схожими вкусами и интересами,
схожим времяпрепровождением

ИСХОДНЫЕ ДАННЫЕ

3

Датасет из ~ 5,5 тысяч строк

24 типа мест

В каждой строке - информация о пользователе(возраст, пол) и средний рейтинг по определенным посещенным им местам(по церквям, паркам, ресторанам и так далее)



МЕТРИКА – DAVIES_BOULDIN_SCORE

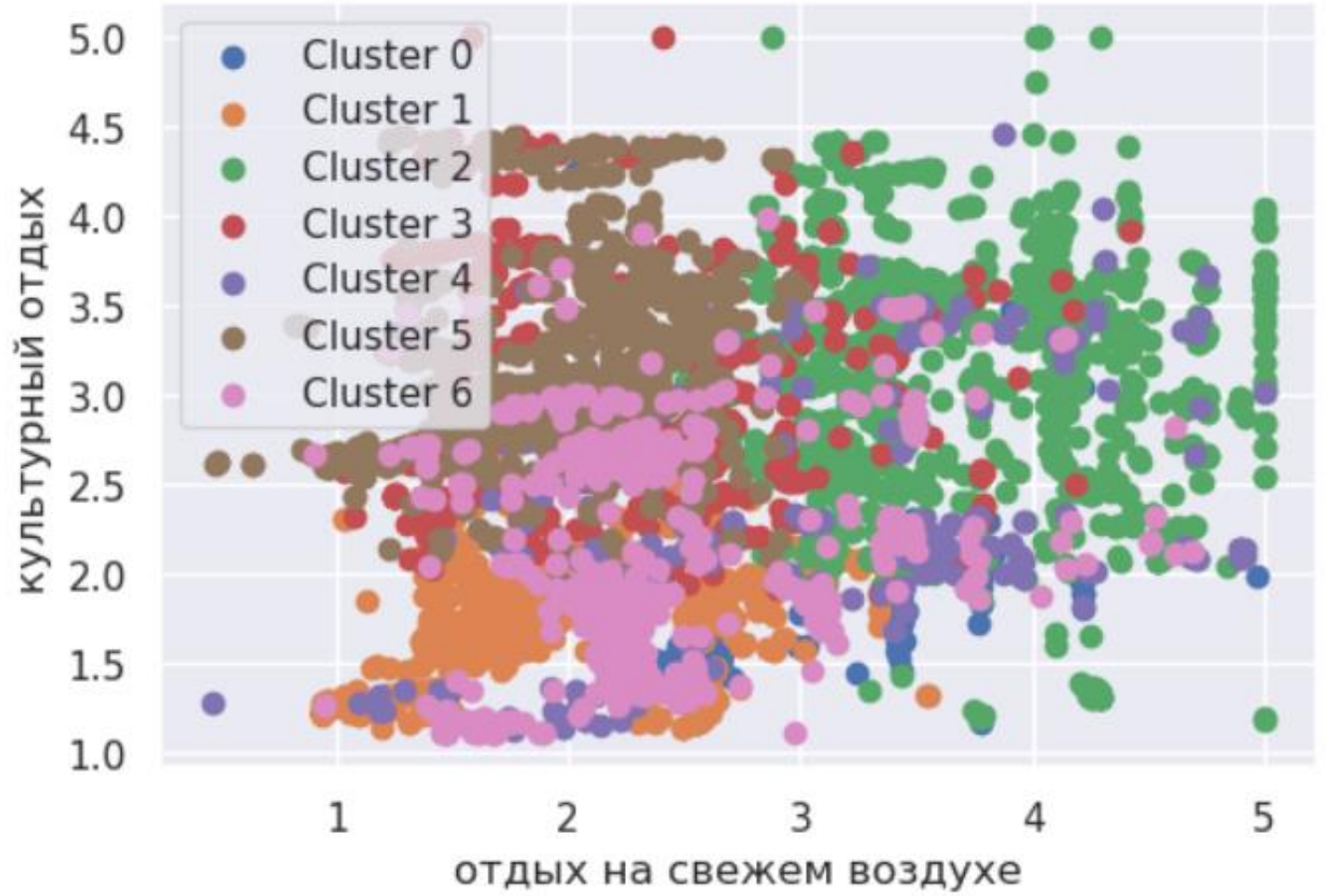
Оценивает, какой алгоритм делит на более
«кучные» кластеры

Лучше всего делит на кластеры модель DBSCAN
с параметрами $\text{eps}=0.2$, $\text{min_samples}=20$

Модель	Дополнительно	DBS
KMeans	N_CLUSTERS = 3	1.46
KMeans	N_CLUSTERS = 7	1.23
DBSCAN	$\text{eps}=0.2$, $\text{min_samples}=20$	0.43
DBSCAN	$\text{eps}=0.6$, $\text{min_samples}=70$	0.63

РЕЗУЛЬТАТЫ

6



РЕЗУЛЬТАТЫ

7

