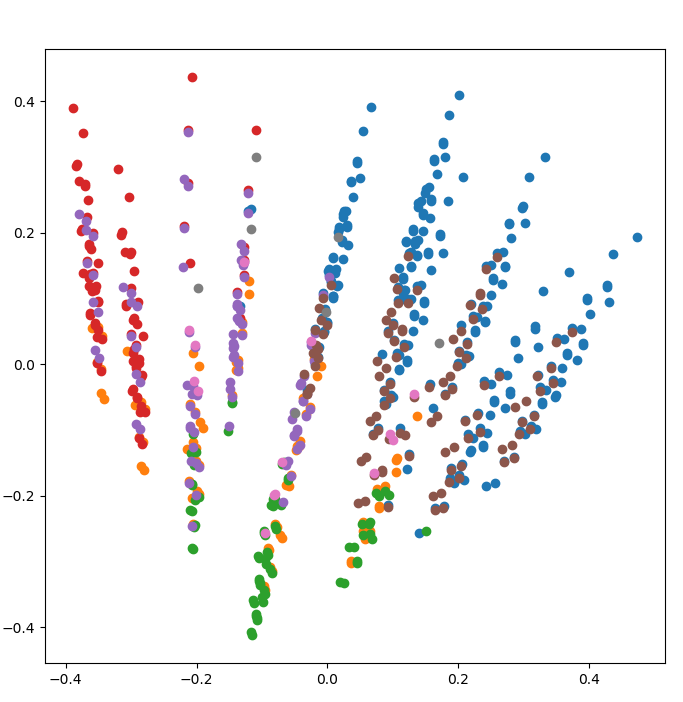
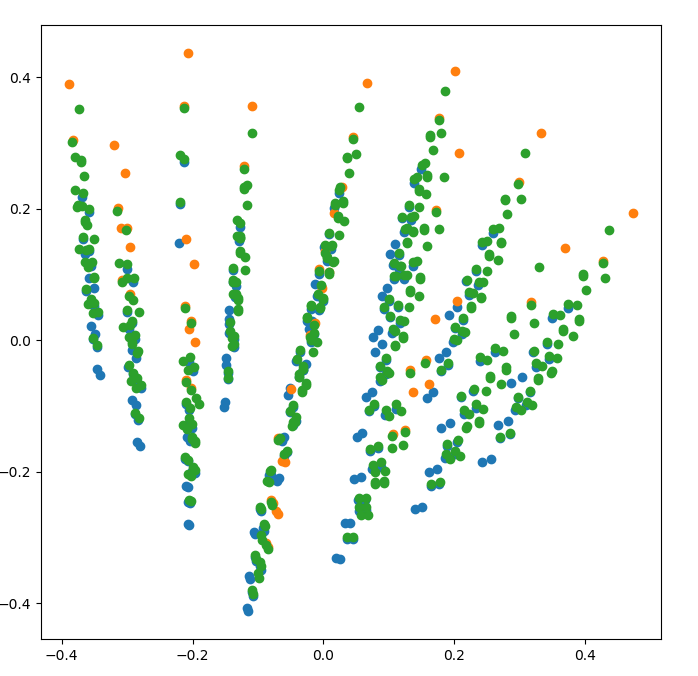
Кластеризация

Взяла <https://www.openml.org/d/1568>

Кластеры (8 штук):



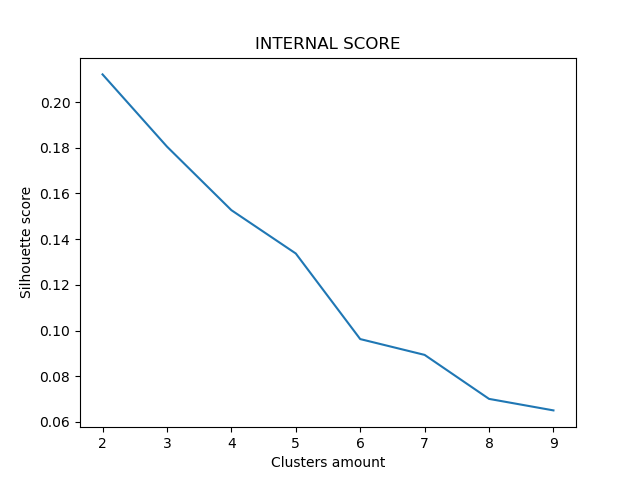
Реальное распределение по классам (4 класса):



Лучшие показатели при 8 кластерах:

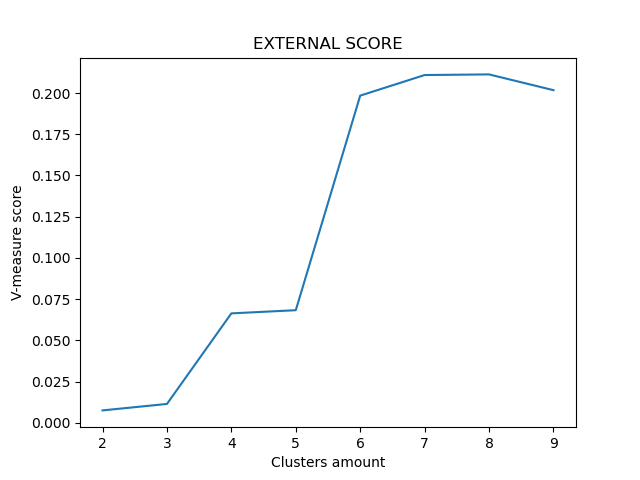
EXTERNAL: 0.2112921183980282

INTERNAL: 0.0700985729011146



Внутренняя метрика - silhouette\_score, рассчитывается как (b - a) / max(a, b). Где b – расстояние между точкой и ближайшим кластером, в котором точка не лежит, а - mean intra-cluster distance.

Принимает значения от -1 до 1, оптимально – единица – кластеры четко разделены между собой. Тут получилось не так ввиду несовпадающих с кластерами меток.



Оптимизация проводилась по внешней метрике - v\_measure\_score, видно, что оптимальное значение у метрики при 8 кластерах

