

Итоговый отчет по результатам классификация БД сети магазинов Рябко Д.А.

Используемые алгоритмы:

Principal component analysis - для понижения размерности данных

Mean shift - для кластеризации

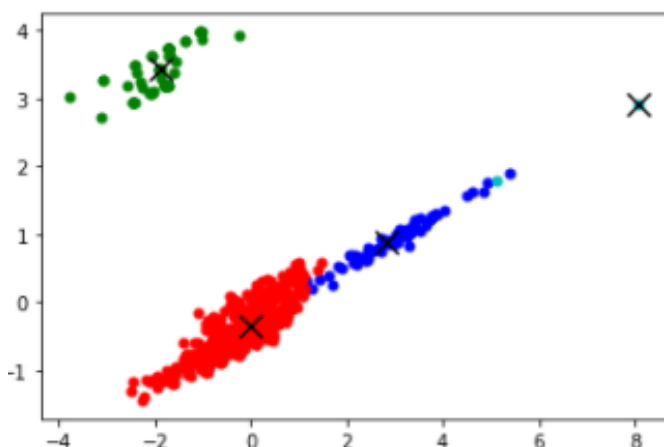
Необходимость понижения данных обусловлена большим количеством критериев и сложностью кластеризации. Как итог- кластеризация основана более чем на 80% информации, полученных из БД.

Mean shift (Сдвиг среднего значения) — это непараметрическая техника анализа пространства признаков для определения местоположения максимума плотности вероятности, так называемый алгоритм поиска моды.

Результаты:

Распределение

0	721
1	63
2	59
3	2



Итог:

Полученные результаты неидеальны и требуют доработок, так как присутствует выброс, при наличие большего количества информации можно было бы определить- выброс это или неописанная группа магазина