Национальный исследовательский Томский государственный университет

Отчёт по лабораторной работе

«Кластеризация»

Вариант 9

Выполнил студент гр. 932001

Андрюшина М. А.

Преподаватель

Аксёнов С. В.

2023 г.

**Работа с данными**

Для работы с бинарными классификаторами использовался набор данных Bike Sharing Dataset (Аренда велосипедов).

Набор данных не содержал пропущенных значений. Для анализа и обучения были использованы признаки hum и atemp. Данные были преобразованы через MinMaxScaler.

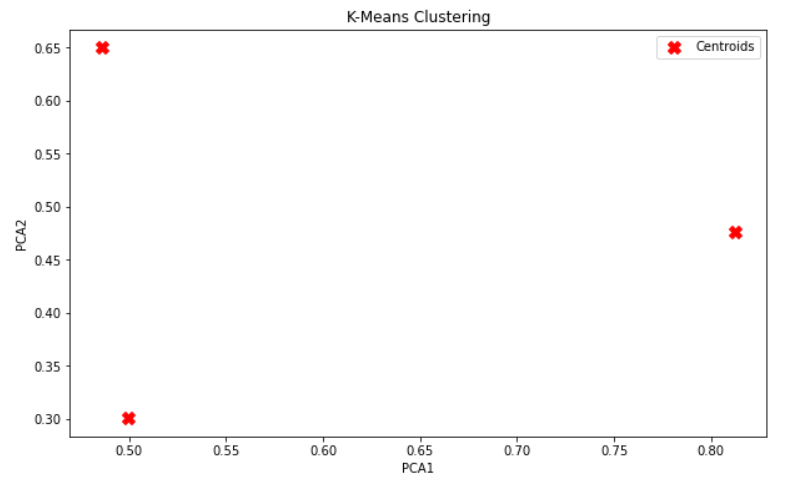
**K-means**

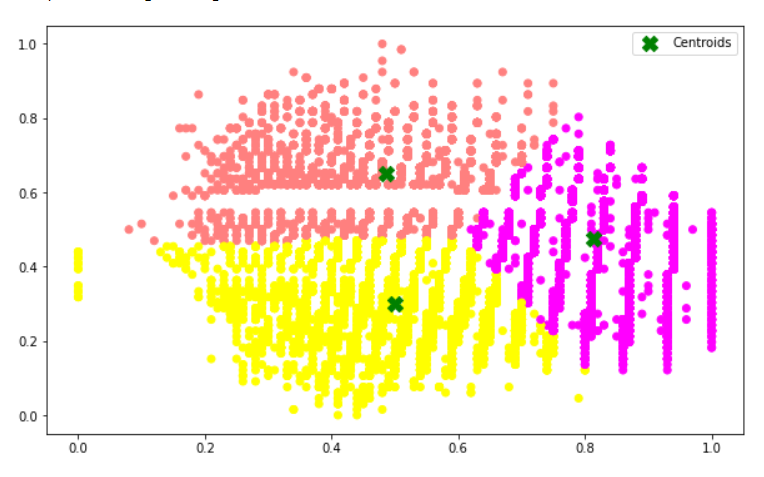
При создании модели Kmeans использовались следующие гиперпараметры: n\_clusters = 3, n\_init = 10, random\_state = 77.

**Метрики качества:**

Коэффициент силуэта: 0.4145711290093497

**Центроиды кластеров:**

**Кластеры:**



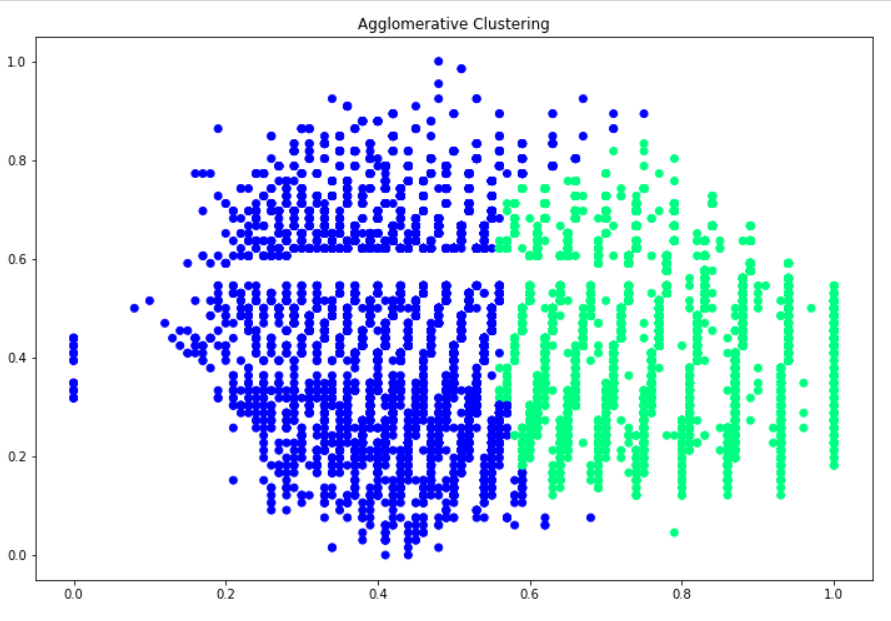
**Аггломеративная кластеризация**

При создании модели AgglomerativeClustering использовались следующие гиперпараметры: n\_clusters = 2.

**Метрики качества:**

Коэффициент силуэта: 0.36511934599116525

**Кластеры:**

****

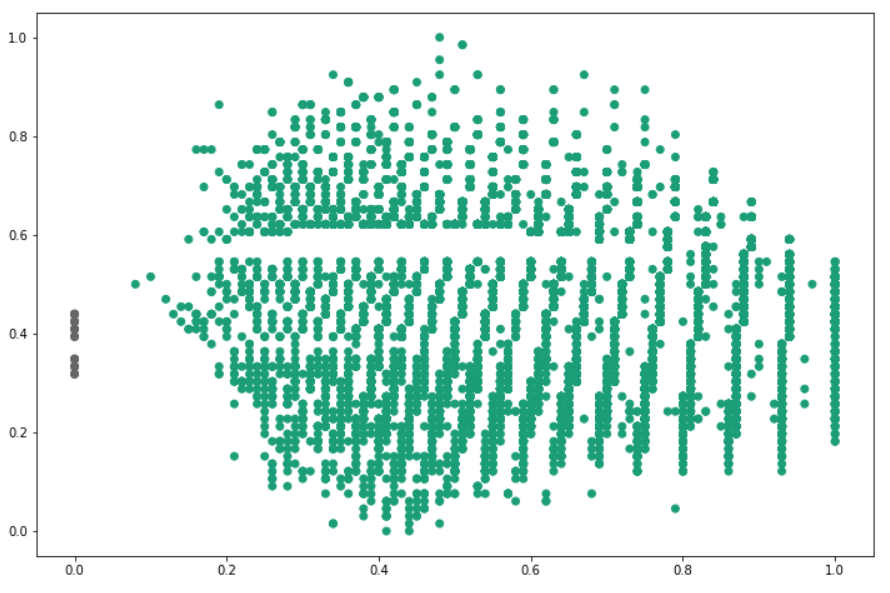
**DBSCAN**

При создании модели DBSCAN использовались следующие гиперпараметры: eps = 0.1, min\_samples = 1. На графике выделено 2 кластера.

**Метрики качества:**

Коэффициент силуэта: 0.45729474310063295

**Кластеры:**

 **Вывод**

Лучше всех показала себя модель DBSCAN с коэффициентом силуэта, равным 0.457. При этом число элементов в первом кластере сильно отличается от числа элементов во втором.

Модели K-means и AgglomerativeClustering выделяли кластеры более равномерно. В каждом кластере получилось примерно одинаковое количество элементов.

**Программный код**

Программный код доступен по ссылке:

<https://github.com/AnMari24/MachineLearning/tree/main/Lab4>