

МИНОБРНАУКИ РОССИИ
Санкт-Петербургский государственный
электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
Кафедра Вычислительной техники

отчет

по лабораторной работе № 3

по дисциплине «Основы машинного обучения»

**Тема: «Применение методов машинного обучения и их оценка для
задач кластеризации»**

Студенты гр. 3311

Аршин А. Д
Баймухамедов Р. Р.
Пасечный Л. В.

Преподаватель

Петруша П. Г.

Санкт-Петербург

Цель работы

Получение и закрепление навыков в применении методов машинного обучения и их оценки для задач кластеризации.

Задание работы

На самостоятельно выбранном в первой лабораторной работе наборе данных найти скрытые структуры - методами решения задач кластеризации. Такими методами как:

- K-means (K-средних)
- Hierarchical clustering (Иерархическая кластеризация)
- DBSCAN

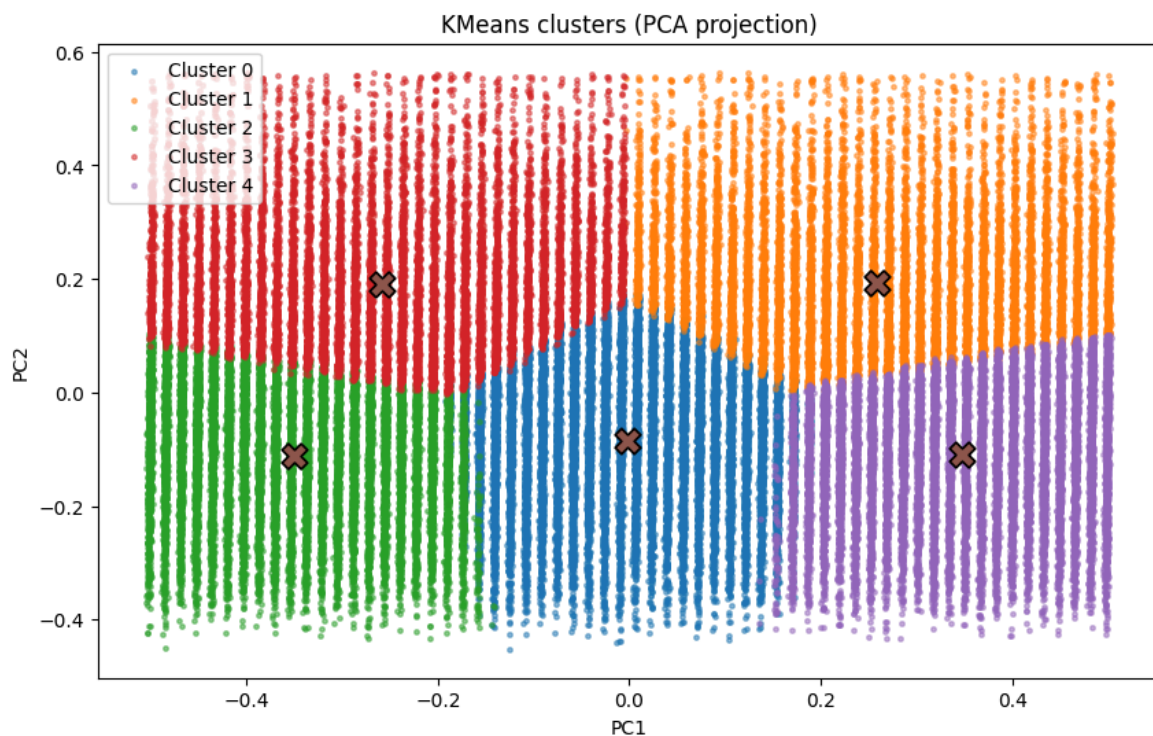
Основные шаги для нахождения кластеров разными методами описаны в Google Colab. Здесь же приведем результат обучения

Лабораторная работа в среде Google Colab -

<https://colab.research.google.com/drive/1gp-gtVqpvHakC2VSQaK8uc2dxl02JxrO?usp=sharing>

Альтернативный вариант - GitHub

https://github.com/brick1ng5654/course-3/tree/RafaelB/boml/lab_02



Распределение объектов по кластерам методом K-Means

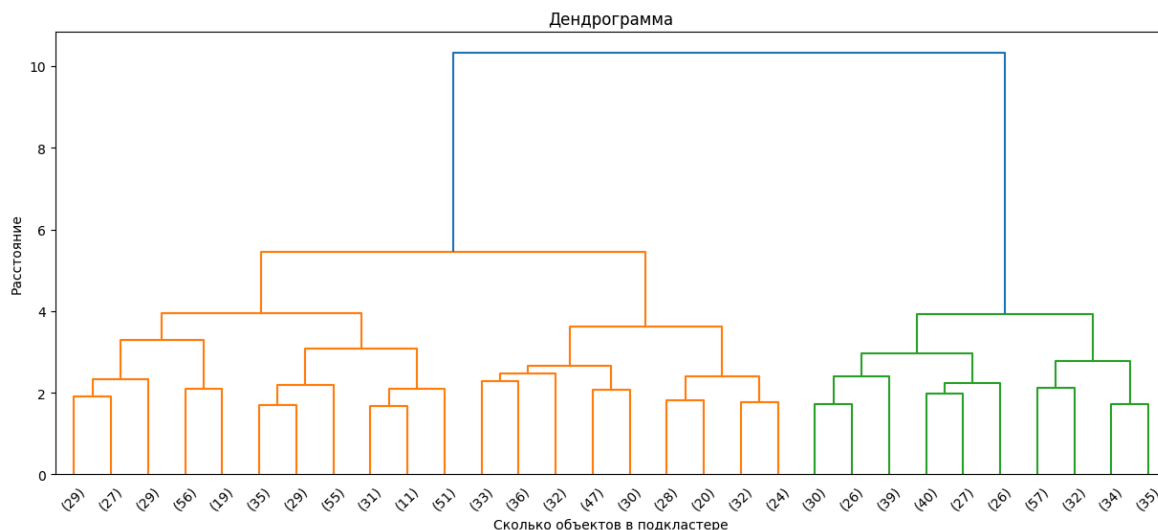
Кластер №0 - средний возраст и умеренно низкий BMI

Кластер №1 - взрослый возраст и высокий BMI

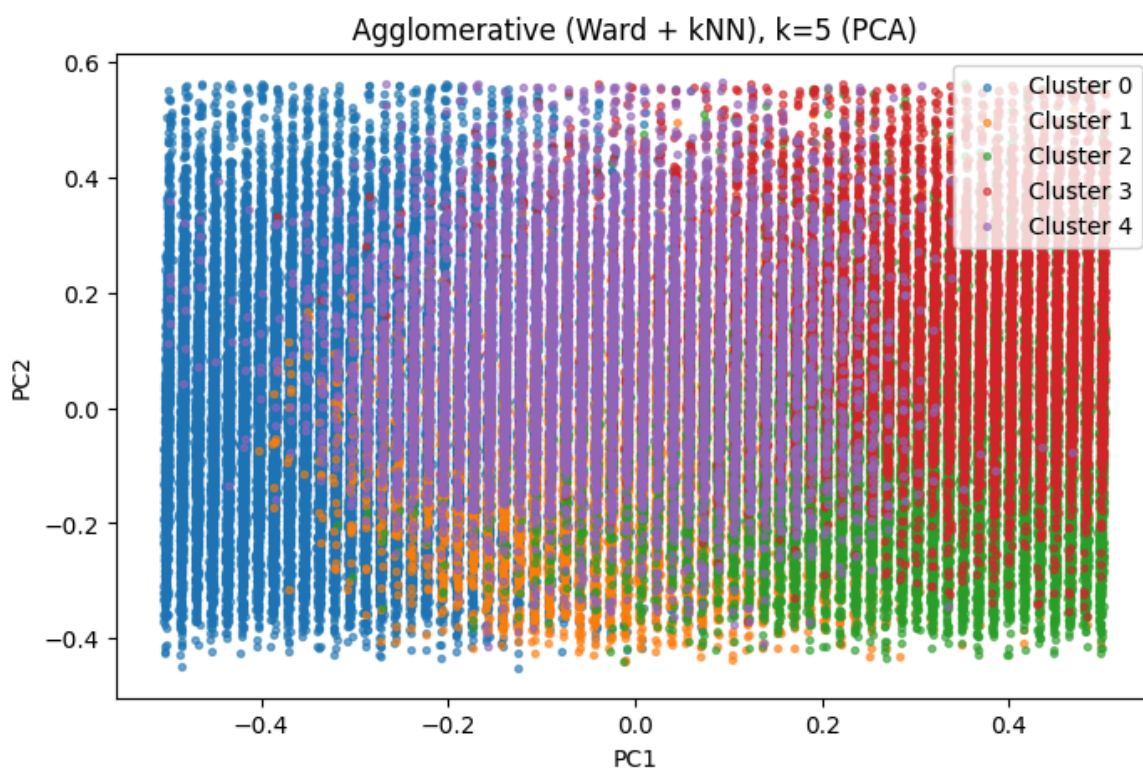
Кластер №2 - самые молодые и низкий BMI

Кластер №3 - молодые с высоким BMI

Кластер №4 - Самые старшие с низким BMI



Дендрограмма метода иерархической кластеризации



Распределения объектов по кластерам методом иерархической кластеризации

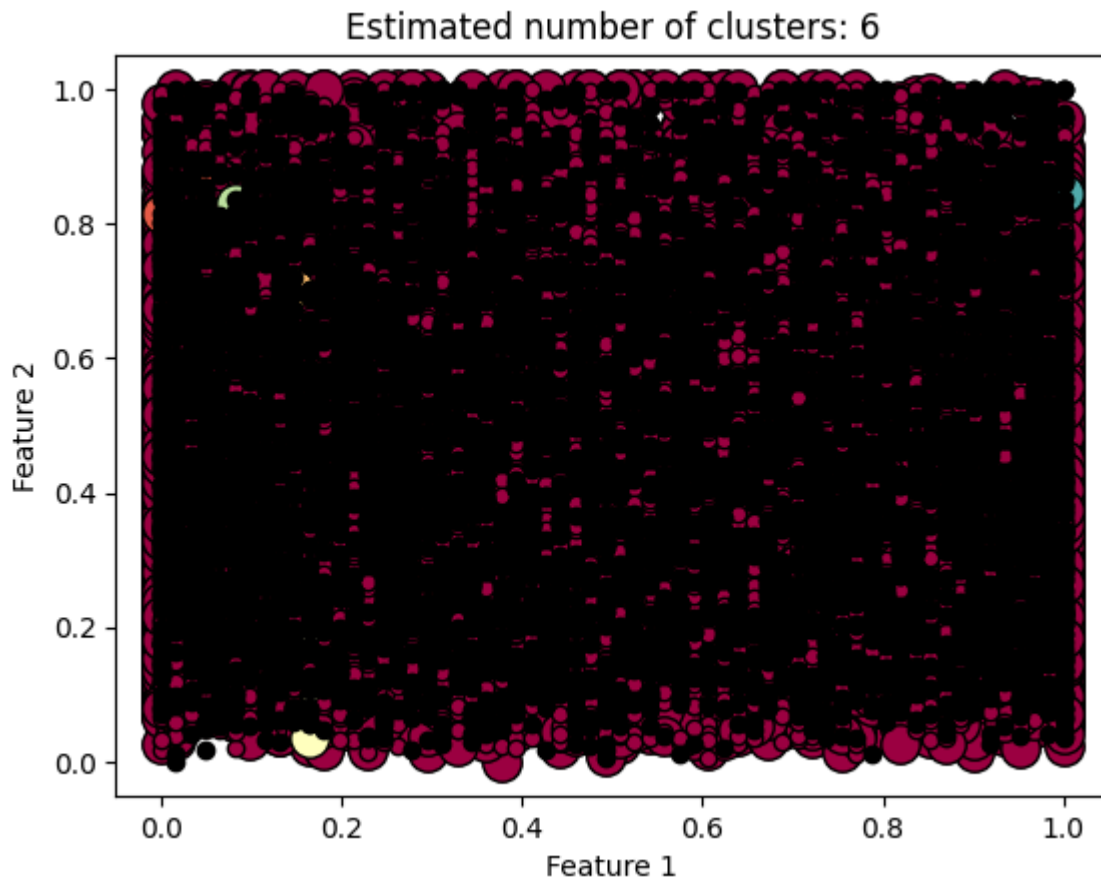
Кластер №0 - самые молодые и низкий BMI

Кластер №1 - средний возраст и умеренно низкий BMI

Кластер №2 - Самые старшие с низким BMI

Кластер №3 - взрослый возраст и высокий BMI

Кластер №4 - молодые с высоким BMI



Распределения объектов по кластерам методом DBSCAN

Вывод

Как видно из кластеров полученными рассматриваемыми методами мы увидели некоторые скрытые структуры, распределяемые по большей мере по возрастному признаку и индекса массы тела. Из статистических данных видно, что половой признак никак не оказал влияния на распределение объектов по кластерам. Признак заболевания незначительно, но отличается между группами самых пожилых и самых молодых. Считаем, что лабораторная работа №3 выполнена в полной мере.