Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Кафедра штучного інтелекту

Дисципліна: «Інтелектуальний аналіз даних»

ЛАБОРАТОРНА РОБОТА №1

«ПОПЕРЕДНЯ ОБРОБКА ДАНИХ»

Варіант 4

Виконав: Прийняла:

ст.гр. ІТШІ-17-1 ст. викл. каф. ШІ  
Гура А. О. Дейнеко А.О.

Харків - 2019

Хід роботи

1. Вибірка мого варіанта: liver\_disorders.csv.
2. Опис початкових даних.  
   Атрибути:  
   1. mcv(65 – 103)

2. alkphos(23 – 138)

3. sgpt(4 – 155)

4. sgot(5 – 82)

5. gammagt(5 – 297)

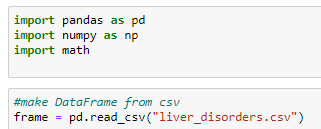
6. drinks(0.0 – 20.0)

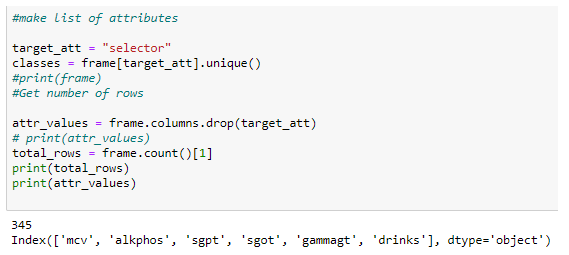
Цільовий атрибут – selector

345 кортежів × 7 атрибутів

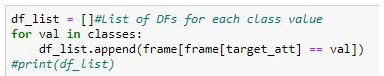
1. Первинний аналіз данних.

Спочатку, під’єднаємо необхідні бібліотеки та зчитуємо файл формату csv, перетворюємо його на датафрейм.

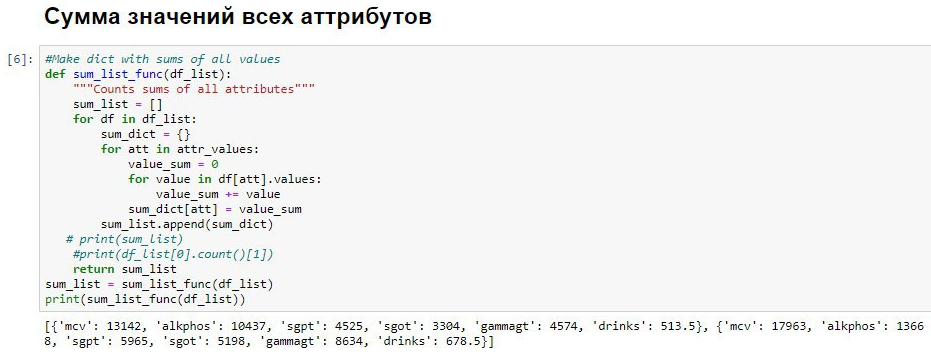


Після цього виділяю цільовий атрибут та лист з усіма іншими атрибутами.  


Далі розбиваю усю вибірку на підвиборки згідно зі значенням цільового атрибуту. В мене вийшло 2 виборки бо цільовий атрибут має 2 унік. значення.



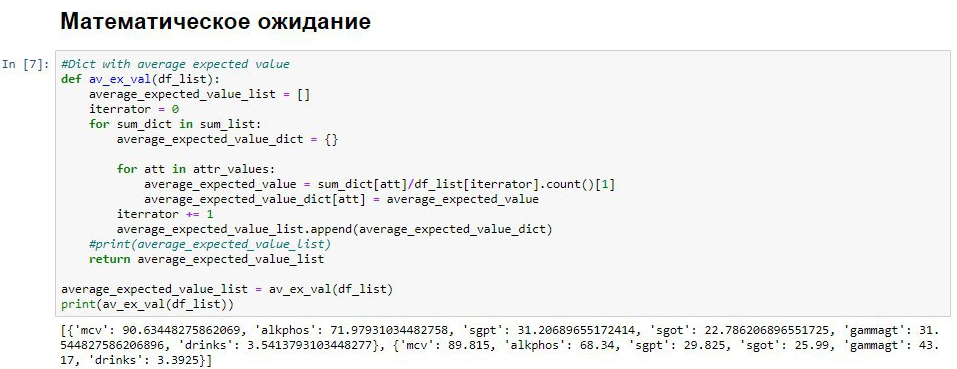
Після цього знаходжу суму усіх значень для кожного атрибуту для кожної підвибірки.



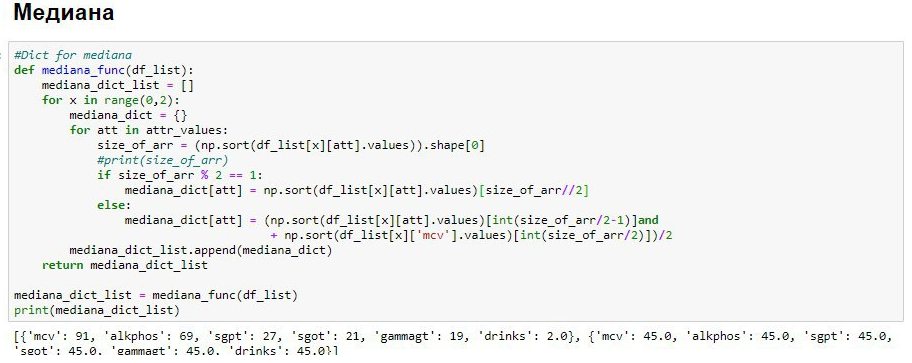
Потім я знаходив математичне очікування за формулою:

Изображение выглядит как объект, часы

Автоматически созданное описание



Після цього я для відсортованих вибірок я знаходжу медіану:

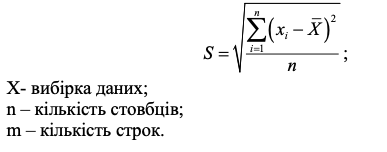


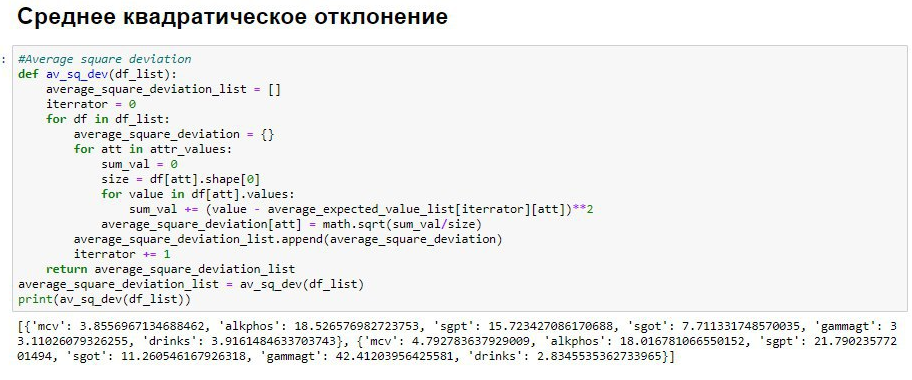
Напівсума «крайніх» спостережень за формулою:





Середнє квадратичне відхилення від середнього за формулою:

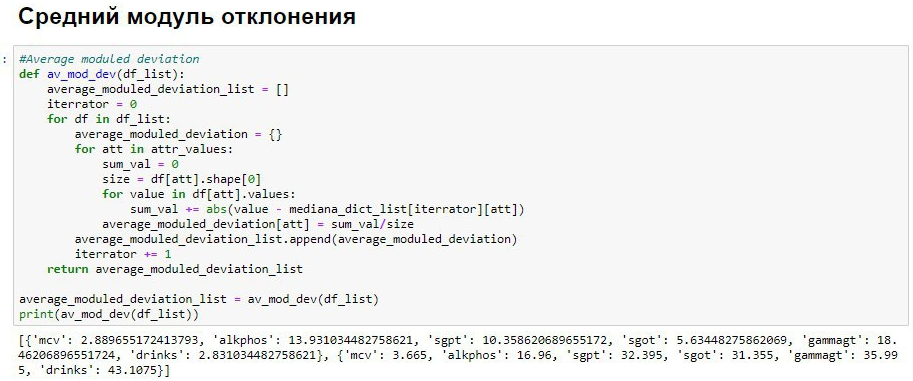




Середній модуль відхилень за формулою:

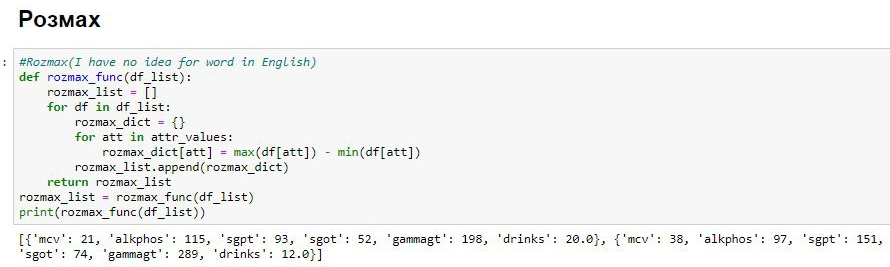
Изображение выглядит как объект

Автоматически созданное описание

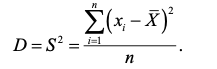


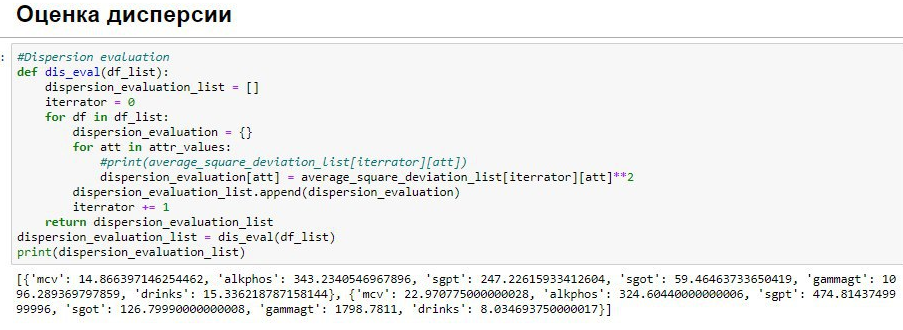
Розмах за формулою:



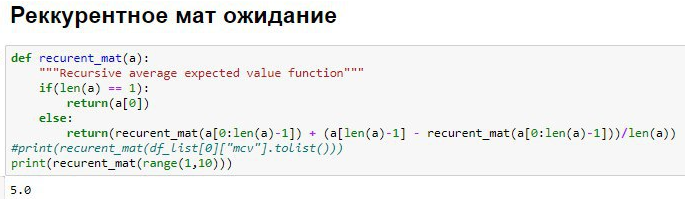


Оцінка дисперсії за формулою:

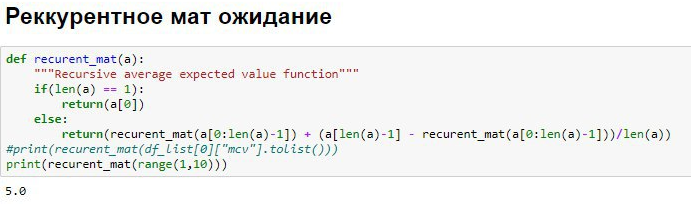




Рекурентне математичне очікування:



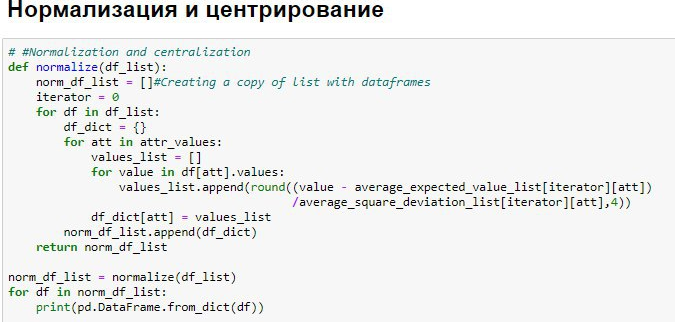
Рекурентна медіана:



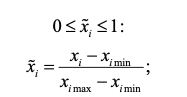
Нормування та центрування за формулою:

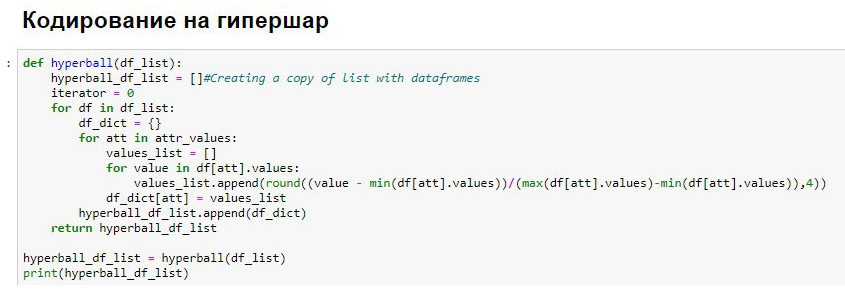
Изображение выглядит как объект, часы

Автоматически созданное описание



Кодування на гіперкулю за формулою:

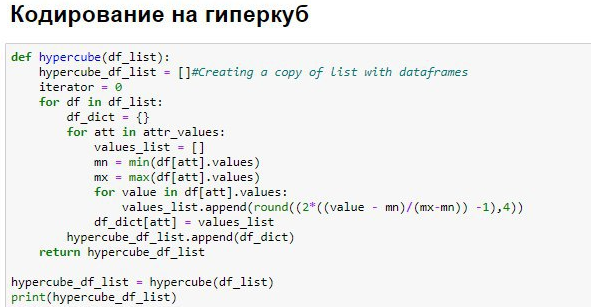




Кодування на гіперкуб за формулою:

Изображение выглядит как объект

Автоматически созданное описание



1. Висновки:

В ході виконання лабораторної роботи я ознайомився з основними методами попередньої обробки даних. Я попрацював з вибіркою liver\_disorders.csv та знайшов для кожного з атрибутів: середнє значення («маточікування»); медіану; напівсуму «крайніх» спостережень; середнє квадратичне відхилення від середнього; середній модуль відхилень; розмах; максимальне та мінімальне значення; провели нормування та центрування; виконали кодування на гіперкулю та кодування на гіперкуб.