# Національний технічний університет України

*Додаток 1*

# «Київський політехнічний інститут імені Ігоря Сікорського»

**Навчально-науковий інститут атомної і теплової енергетики Кафедра цифрових технологій в енергетиці**

**ЗВІТ**

# з лабораторної роботи № 5

**з дисципліни «[Вступ до інтелектуального аналізу даних](https://classroom.google.com/u/0/c/NjM5MzA3MTQyNjEz" \t "https://classroom.google.com/u/0/c/NjM5MzA3MTQyNjEz/a/NjUzMjM5MTY4NzY2/_self)»**

# Тема: «Hierarchical Clustering and

# Dendrograms in Python»

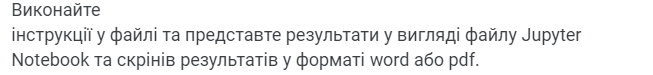
Варіант № 21

Виконав: Студент групи ТР-12

Руденко Владислав

Київ – 2023

**Завдання**



**Хід виконання та Результати**

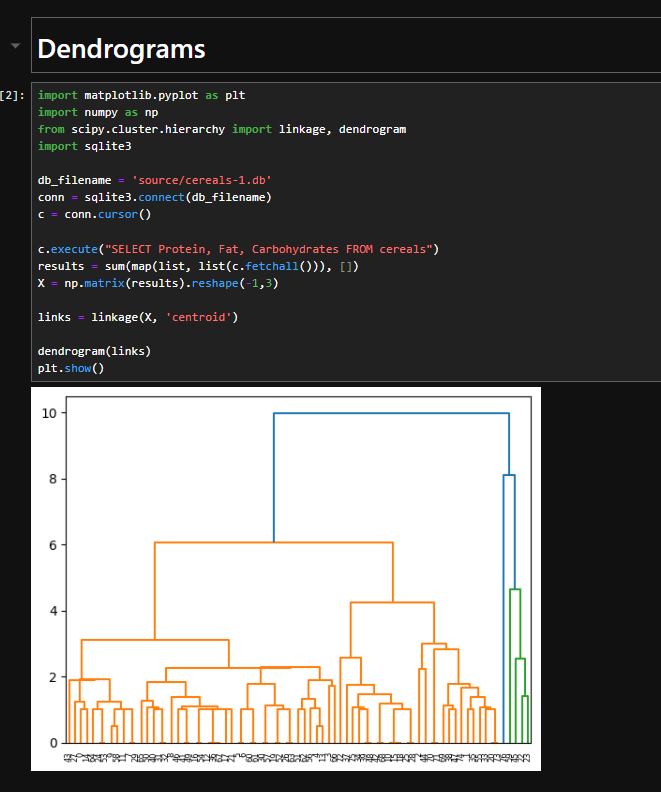
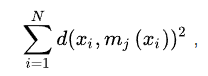


Table 1 K Means

Кластериза́ція ме́тодом k-сере́дніх  — популярний метод [кластеризації](https://uk.wikipedia.org/wiki/%D0%9A%D0%BB%D0%B0%D1%81%D1%82%D0%B5%D1%80%D0%B8%D0%B7%D0%B0%D1%86%D1%96%D1%8F" \o "Кластеризація), — впорядкування множини об'єктів у порівняно однорідні групи.

Мета методу — розділити n спостережень на k кластерів, так щоб кожне спостереження належало до кластера з найближчим до нього [середнім значенням](https://uk.wikipedia.org/wiki/%D0%A1%D0%B5%D1%80%D0%B5%D0%B4%D0%BD%D1%94_%D0%B0%D1%80%D0%B8%D1%84%D0%BC%D0%B5%D1%82%D0%B8%D1%87%D0%BD%D0%B5" \o "Середнє арифметичне). Метод базується на мінімізації суми квадратів відстаней між кожним спостереженням та центром його кластера, тобто функції



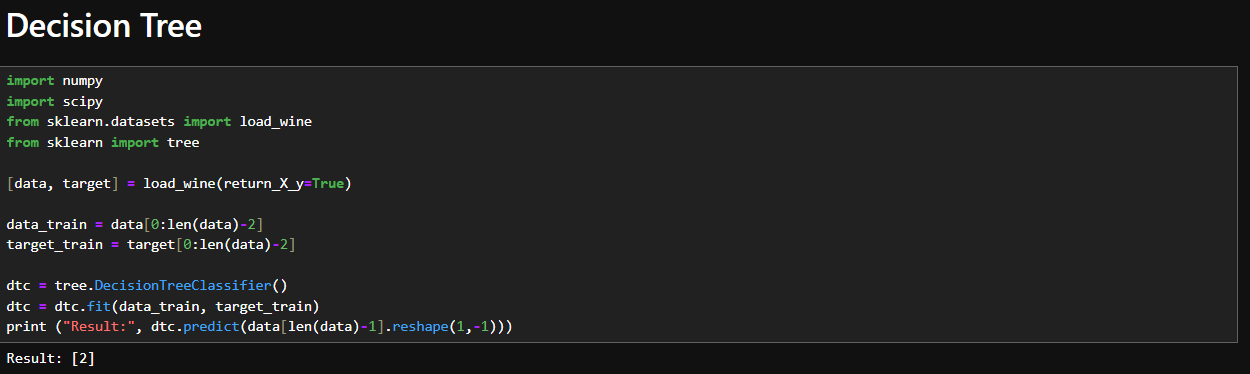
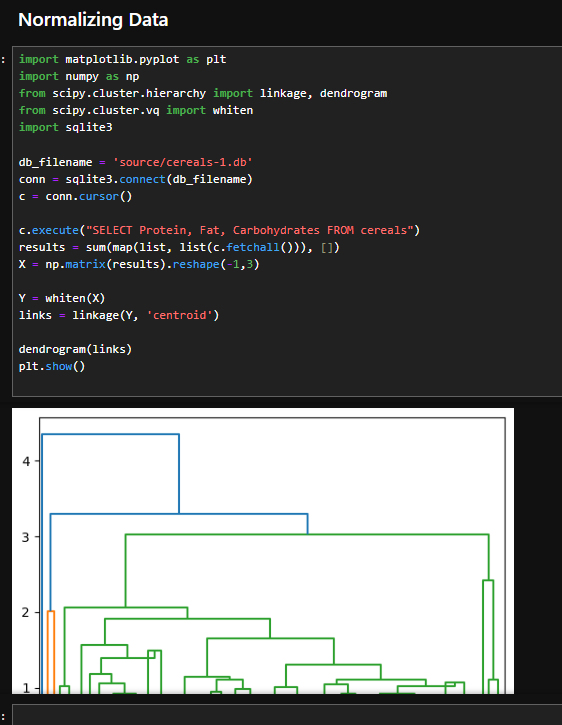


Table 2 Decision Tree

Дендрограма - це діаграма, що зображує дерево. Це схематичне зображення часто використовується в різних контекстах: в ієрархічній кластеризації вона ілюструє розташування кластерів, отриманих відповідними аналізами.



Нормалізація даних - це процес приведення вхідних даних до стандартної форми перед їхнім використанням у кластерному аналізі та побудові дендрограми.Нормалізація даних допомагає зробити аналіз більш об'єктивним і дозволяє правильно порівнювати об'єкти між собою в контексті їхньої схожості чи віддаленості. Це важливий етап у процесі кластерного аналізу, який допомагає виявити структуру даних та відображає залежності між об'єктами чи групами об'єктів.