

## **ПАРАЛЕЛЬНА РЕАЛІЗАЦІЯ АЛГОРИТМУ K-MEANS З ПОСЛІДОВНИМ ЗАПУСКОМ АКТИВНИХ ЦЕНТРОЇДІВ**

Лаховець С.М.

Науковий керівник – доц., к.т.н. Ткаченко О.М.

Об'єктом даного дослідження є кластеризація даних. Предметом – паралельна реалізація алгоритму k-means.

Побудова систем аналізу даних є важливим напрямком розвитку інформаційних технологій. Кластеризація даних є відомою науковою та практичною задачею, а саме, як розподілити експериментально отримані набори векторів по групах, або кластерах. Кластеризація часто використовується, зокрема, при статистичному аналізі даних, векторній квантизації, розпізнаванні образів тощо.

Алгоритм k-means є найбільш розповсюдженим і найкраще дослідженим серед усіх методів кластеризації. Він мінімізує спотворення, розподіляючи дані між регіонами, що не перетинаються та ідентифікуються за їх центрами. Поширеність методу k-середніх зумовлено його головними перевагами: простоті, гнучкості, швидкій збіжності.

В ході роботи був паралельно реалізований алгоритм k-means з послідовним запуском активних центроїдів. Для цього було детально розглянуто переваги та недоліки стандартного методу кластеризації k-means, визначення шляхи його покращення, та розроблено спосіб розпаралелення алгоритма.

За результатами дослідження було створено програму для кластеризації даних з інтерфейсом командного рядка, що отримує данні та параметри на вході, й видає кластеризовані данні на виході. Програма показує кращу швидкодію за Matlab, та може використовувати для обчислень всі наявні ядра на процесорі.