

КЛАСТЕРНИЙ АНАЛІЗ ЗАХВОРЮВАНОСТІ НА COVID-19
В УКРАЇНІ

*Федулов Кирило Андрійович, студент групи 365,
Базілевич Ксенія Олексіївна*, доцент кафедри математичного
моделювання та штучного інтелекту
Національний аерокосмічний університет ім. М.Є. Жуковського «ХАІ»*

Дослідницька спільнота усього світу намагається дослідити медичний, економічний та соціологічний вплив пандемії COVID-19, Україна не є виключенням, а задача аналізу захворюваності є актуальною. Для вирішення цієї задачі у роботі запропоновано використовувати методи Data Mining, зокрема кластерний аналіз. Основна ідея кластерного аналізу – виділення серед безлічі даних груп, всередині яких елементи в певній мірі схожі. В рамках такого аналізу відбувається певна класифікація досліджуваних даних за рахунок розподілу їх по групах. Ці групи впорядковані ієрархічно і структуру таких одержаних після аналізу кластерів можна представити у вигляді дерева.

У даній роботі було розглянуто існуючі методи кластеризації, проведено їх аналіз та порівняння, і реалізовано метод к-середніх. Основні переваги методу: простота використання, швидкість використання, зрозумілість і прозорість алгоритму. Метод к-середніх розбиває безліч елементів векторного простору на заздалегідь відоме число кластерів k . Необхідно мінімізувати середньоквадратичне відхилення на точках кожного кластера. Основна ідея полягає у тому, що на кожній ітерації обчислюють центр мас для кожного кластера, отриманого на попередньому кроці, потім вектори розбиваються на кластери знову відповідно до того, який з нових центрів виявився ближчим за обраною метрикою. Алгоритм завершується, коли на певній ітерації не відбувається зміни кластерів. У результаті дослідження на основі розглянутої математичної моделі була побудована алгоритмічна модель, що реалізована у програмних модулях на Python. Розрахунки було проведено на даних університету Джона Хопкінса [1].

Список літератури:

1. Johns Hopkins University & Medicine [Electronic resource]: Coronavirus resource center, Access Mode:<https://coronavirus.jhu.edu/map.html>

** Виконано в рамках проекту Національного фонду досліджень України 2020.02/0404 «Розробка інтелектуальних технологій оцінки епідемічної ситуації для підтримки прийняття управлінських рішень у сфері біобезпеки населення».*