

# Засіб аналізу даних у задачах кібербезпеки з використанням кластеризації

Виконала:

студентка

групи ІБС-17 м

Гнідовська Ю. Ю.

Керівник:

к.т.н, проф. каф. ЗІ

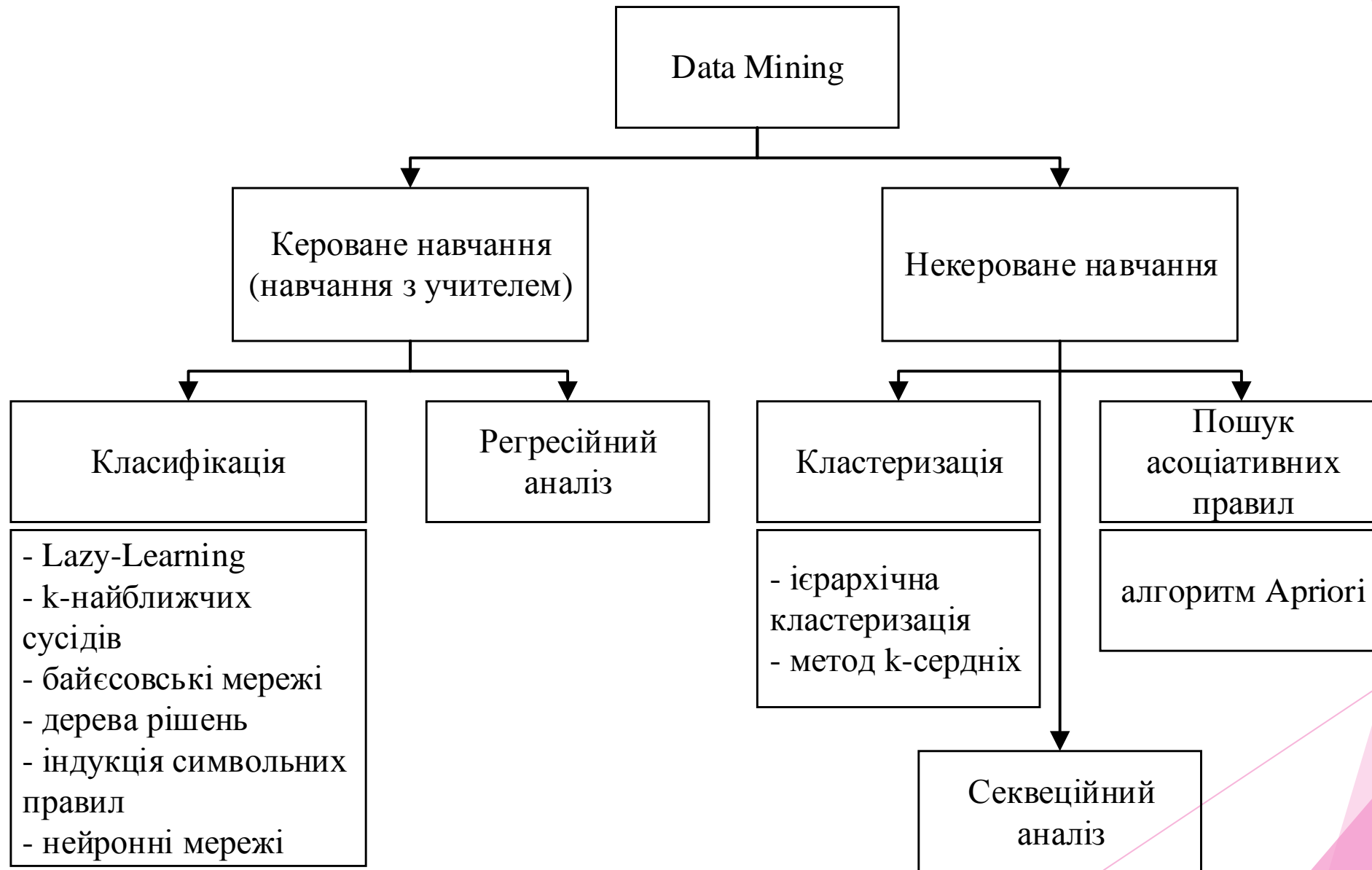
Кондратенко Н. Р.

- ▶ **Актуальність** теми магістерської кваліфікаційної роботи пов'язана зі зростанням кіберзагроз інформаційної безпеки, що виникають через наявність різного роду вразливостей та пошуку ефективних методів аналізу даних для розв'язання задач кібербезпеки.
- ▶ **Метою** магістерської кваліфікаційної роботи є підвищення ефективності прийняття рішень в задачах кібербезпеки, що вирішуються за допомогою нейромереж, завдяки побудові еталонної вибірки, що пришвидшує час навчання нейромережі.
- ▶ **Об'єкт** дослідження - процес аналізу даних у задачах кібербезпеки з використанням нейромережевої моделі кластер-регресійної апроксимації.
- ▶ **Предмет** дослідження - модель кластер-регресійної апроксимації.

# *Постановка задачі:*

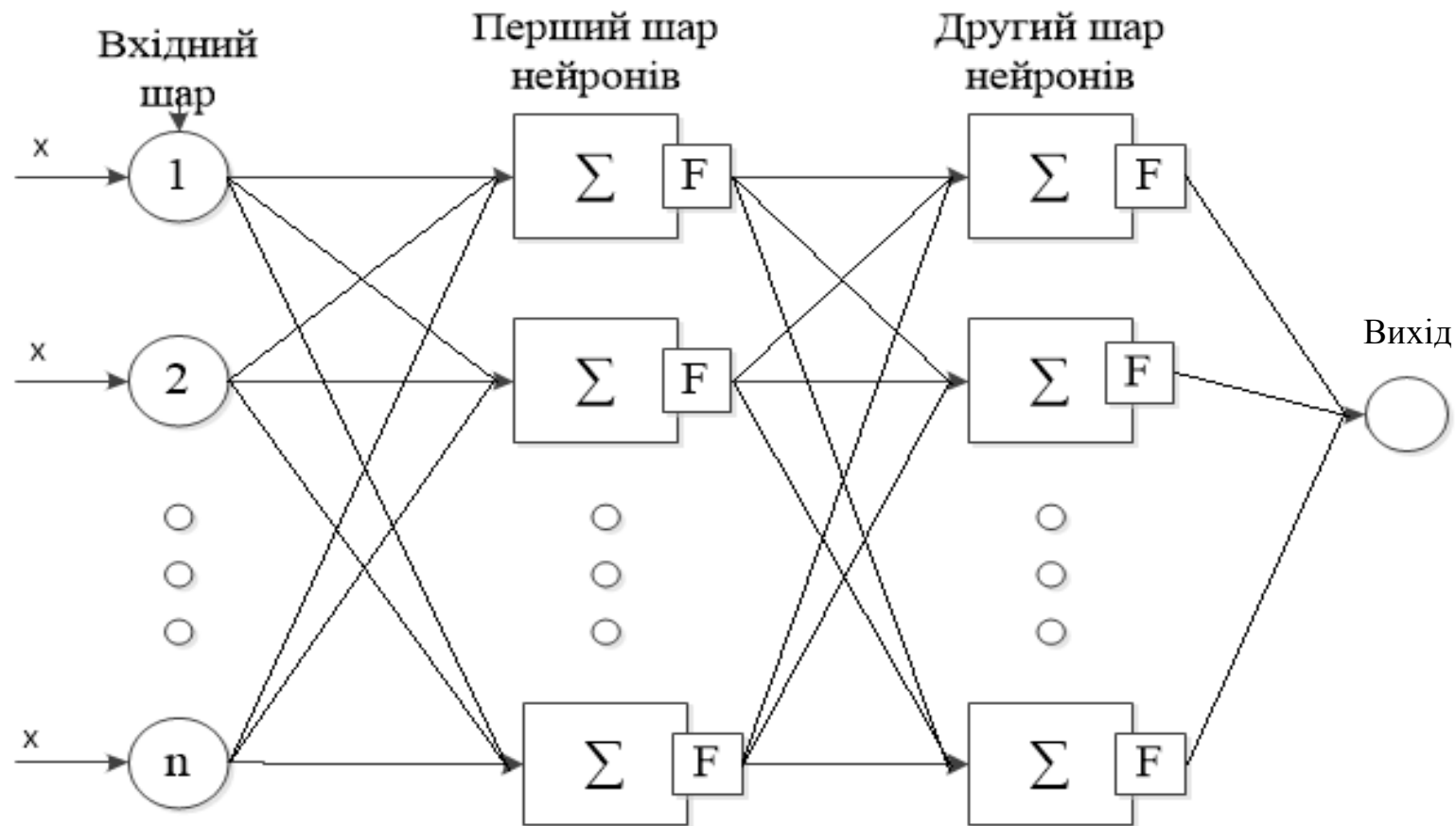
- ▶ виконати огляд літературних джерел;
- ▶ проаналізувати сучасні методи аналізу даних;
- ▶ підготувати та сформулювати вибірку даних;
- ▶ розробити підхід, який поєднує елементи регресійного та кластерного аналізу;
- ▶ провести експериментальні дослідження реалізованої методики;
- ▶ обґрунтувати економічну доцільність.

# Методи аналізу даних



- ▶ відбір вибірки об'єктів для кластеризації.
- ▶ визначення безлічі змінних, за якими будуть оцінюватися об'єкти у вибірці.
- ▶ обчислення значень міри схожості між об'єктами. Застосування методу кластерного аналізу для створення груп схожих об'єктів (кластерів).
- ▶ представлення результатів аналізу

# Багатошаровий перцептрон



# Модель апроксимації точкових даних

$$x = \{x_j^s\},$$

де  $x_j^s$ , – значення  $j$ -тої ознаки  $s$ -го елемента,  $s=1,2,\dots,S$ ;

$S$  - кількість елементів в навчальній вибірці, і кожному елементу предствлено значення прогнозованого параметра  $y^s$ .

Тоді залежність  $y$  від  $x$  можна записати у вигляді:

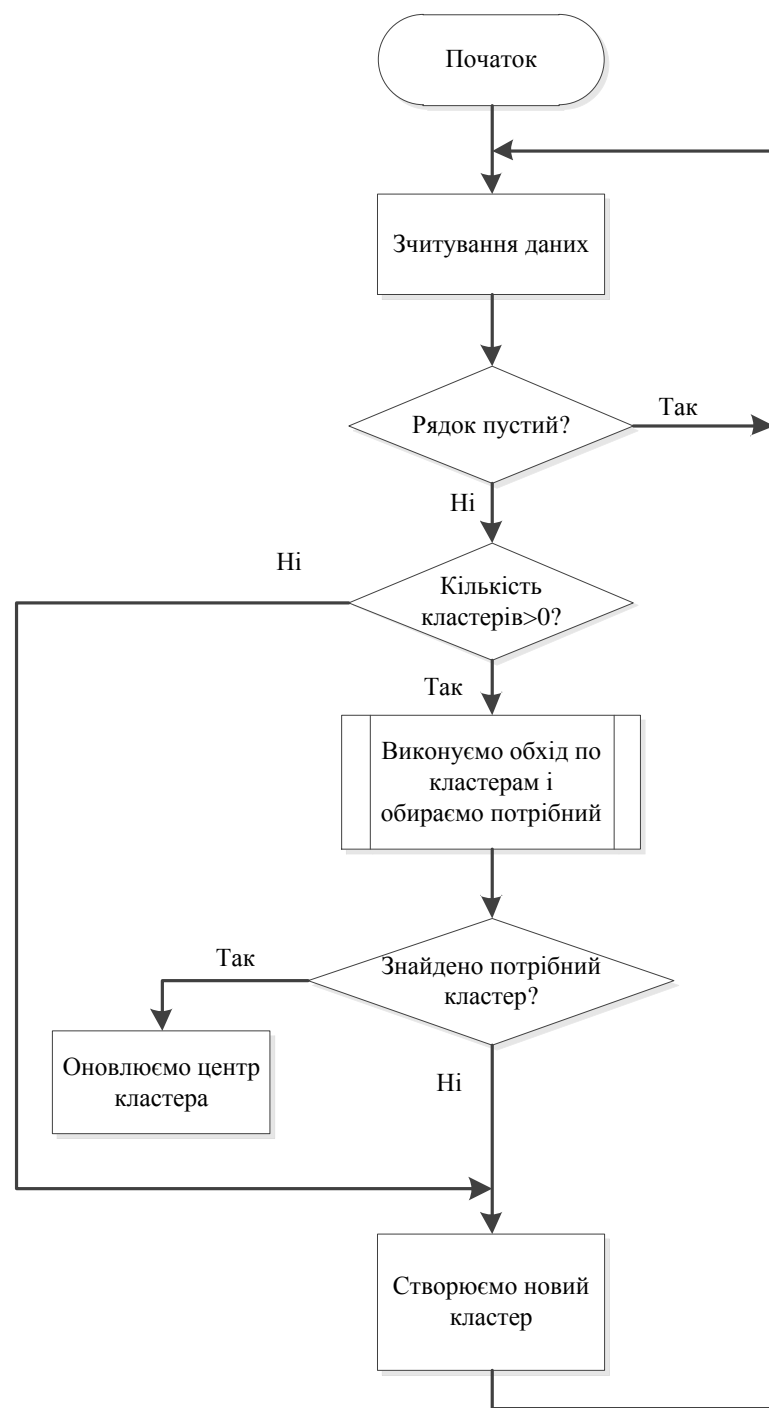
$$y = f(w, x) + Err,$$

де  $f(w, x)$  - деяка функція, вид якої задається алгоритмом апроксимації або користувачем;

$w$  – множина параметрів, що дозволяють налаштувати (навчити) функцію на розв'язання конкретної апроксимаційної задачі.

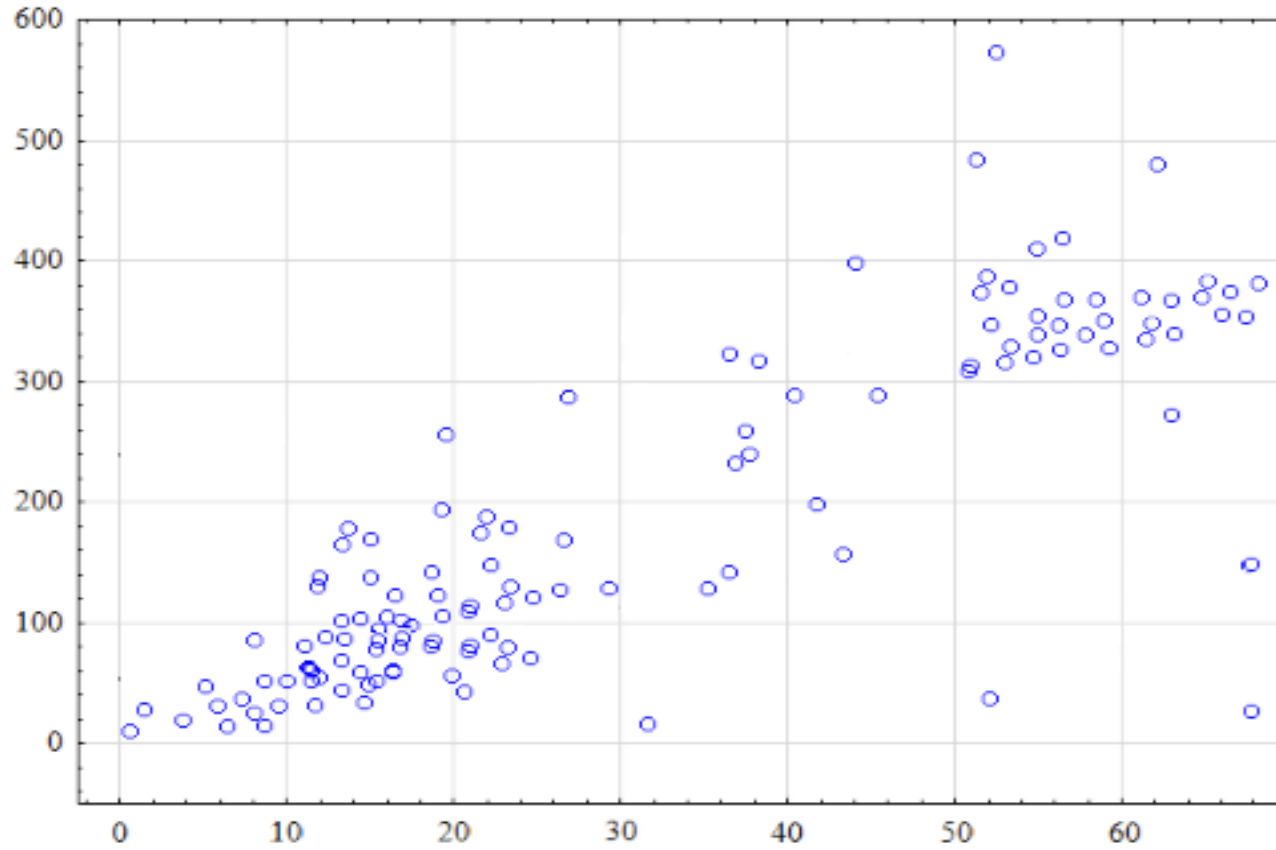
$Err$  – деяка помилка, що виникає через неповну відповідність виду апроксимуючої функції виду реально існуючій залежності та похибок у визначенні значень параметрів апроксимуючої функції.

$$X = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_m^1 \\ x_1^2 & x_2^2 & \dots & x_m^2 \\ \dots & \dots & \dots & \dots \\ x_1^n & x_2^n & \dots & x_m^n \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix},$$

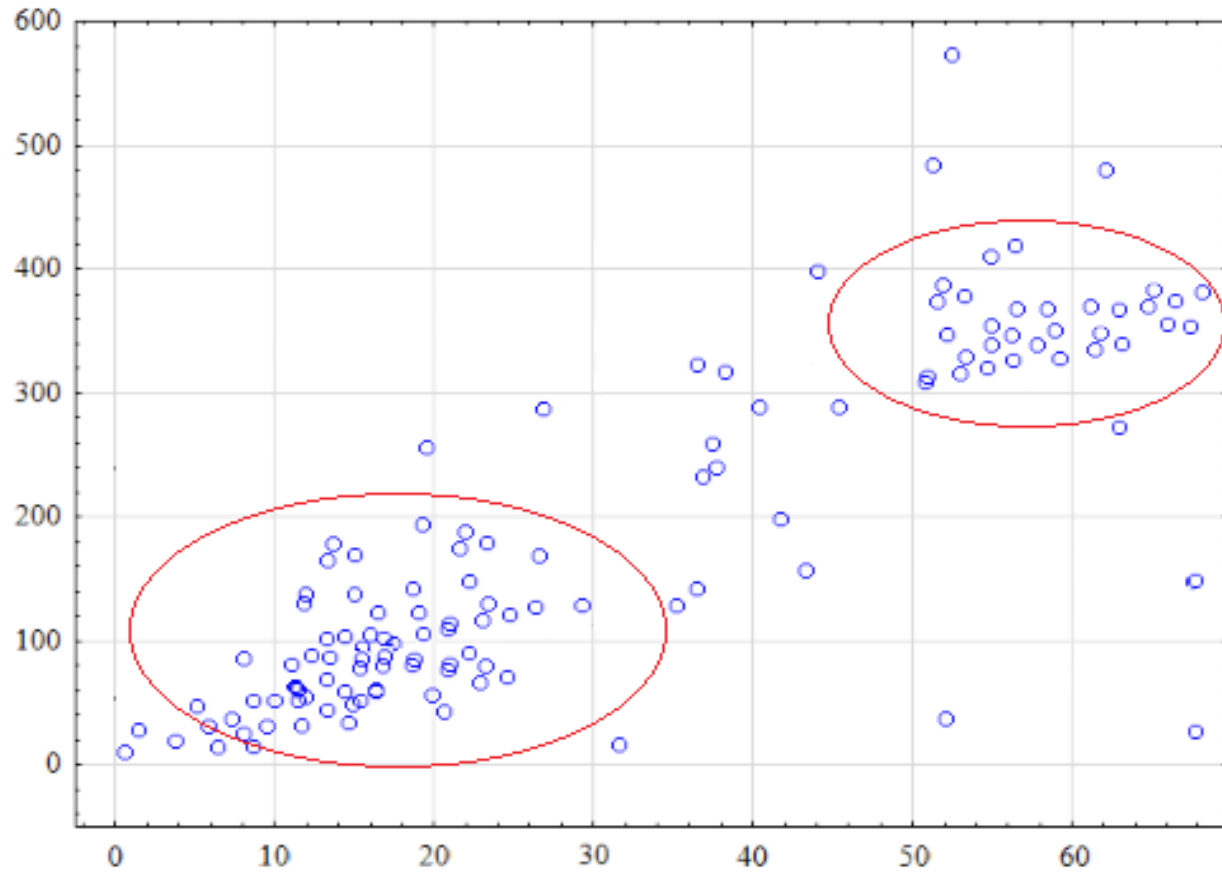




## Результати виконання кластеризації навчальної вибірки



## Результати виконання кластеризації навчальної вибірки



Коефіцієнт кореляції:  $r_{xy} = \frac{c_{xy}}{\sigma_x \sigma_y}$ .

Лінійне наближення  $\bar{y}_x$  від  $x$ :

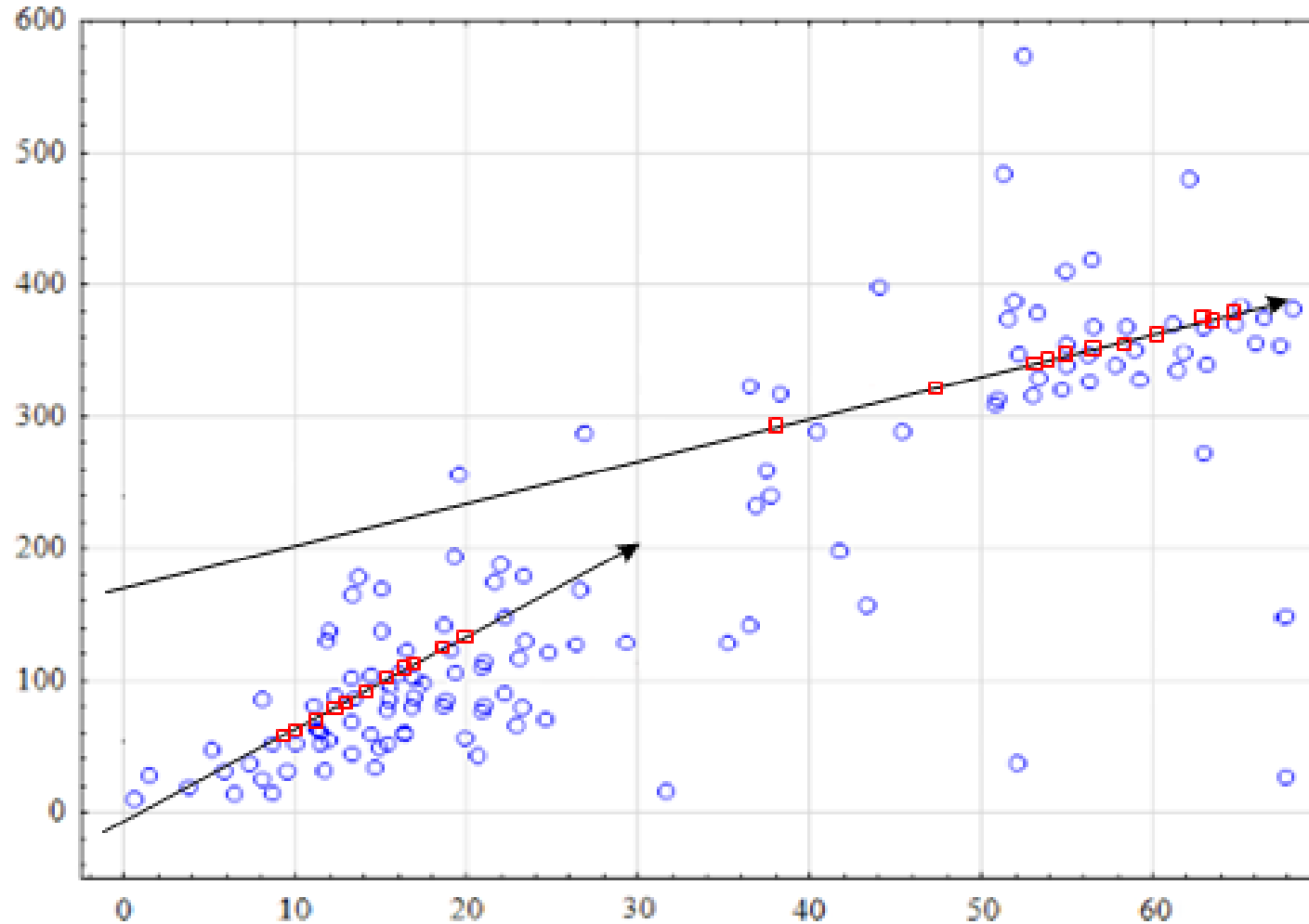
$$\bar{y}_x - \bar{y} = r_{xy} * \frac{\sigma_y}{\sigma_x} (x - \bar{x}), \text{ або } \bar{y}_x = ax + b.$$

Лінійне наближення  $\bar{x}_y$  від  $y$ :

$$\bar{x}_y - \bar{x} = r_{xy} * \frac{\sigma_x}{\sigma_y} (y - \bar{y}), \text{ або } \bar{x}_y = cy + d.$$

Рівняння лінії регресія для першого кластеру:  $\bar{y}_i = 314,42 + 1,01x_i$ .

Рівняння лінії регресія для другого кластеру:  $\bar{y}_i = 63,37 + 1,01x_i$



# **Обґрунтування економічної доцільності:**

- ▶ Проведено оцінку комерційного потенціалу розробки;
- ▶ Спрогнозовано витрати на виконання наукової роботи – 61149,33 грн.;
- ▶ Спрогнозовано чистий прибуток від впровадження результатів розробки – за 3 роки 322987,4 грн.;
- ▶ Термін окупності – 3,3 роки.

***Дякую за увагу!***