МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Західноукраїнський національний університет

Факультет комп'ютерних інформаційних технологій

Кафедра інформаційно-обчислювальних

систем і управління

Звіт про виконання лабораторної роботи №2

з дисципліни «Методи та системи штучного інтелекту»

студент групи КН-31

Ковальковський Віталій

Тернопіль – 2023

ПОБУДОВА РЕГРЕСОРА МЕТОДОМ K-НАЙБЛИЖЧИХ СУСІДІВ.

Завдання:

1.Генерація та нормалізування 1000 випадкових значень:   
% Генеруємо 1000 випадкових значень в діапазоні від 0 до 1

random\_data = rand(1, 1000);

% Нормалізуємо значення, поділивши їх на максимальне значення

normalized\_data = random\_data / max(random\_data);

2. Розділити вибірку на дві частини (дослідну і тестову) – представити графічно:   
% Кількість всіх даних

total\_data = length(normalized\_data);

% Визначаємо відсоток даних для навчання і тестування

train\_percentage = 0.8; % Наприклад, 80% для навчання, 20% для тестування

test\_percentage = 1 - train\_percentage;

% Генеруємо випадкові індекси для навчальної вибірки

train\_indices = randperm(total\_data, round(train\_percentage \* total\_data));

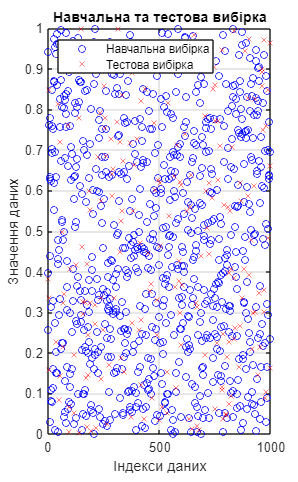
% Решта індексів буде для тестової вибірки

test\_indices = setdiff(1:total\_data, train\_indices);

% Вибираємо дані для навчання та тестування на основі індексів

train\_data = normalized\_data(train\_indices);

test\_data = normalized\_data(test\_indices);



3. Навчити KNN-регрсор з різними значеннями K

% Навчаємо та тестуємо KNN регресор для кожного значення K

for k = k\_values

knn\_model = fitcknn(train\_data', train\_data', 'NumNeighbors', k);

predictions = predict(knn\_model, test\_data');

% Оцінка результатів (наприклад, середньоквадратична помилка)

mse = mean((predictions - test\_data').^2);

fprintf('K = %d, Середньоквадратична помилка = %f\n', k, mse);

% Запам'ятовуємо найкращий результат

if mse < best\_mse

best\_mse = mse;

best\_k = k;

end

% Виводимо графік прогнозів

figure;

scatter(test\_indices, test\_data, 50, 'filled', 'b', 'DisplayName', 'Тестова вибірка');

hold on;

scatter(test\_indices, predictions, 50, 'filled', 'r', 'DisplayName', ['Прогнози KNN (K = ' num2str(k) ')']);

xlabel('Індекси даних');

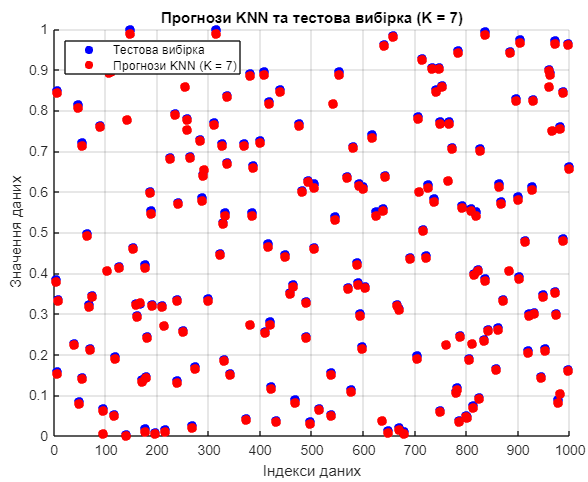
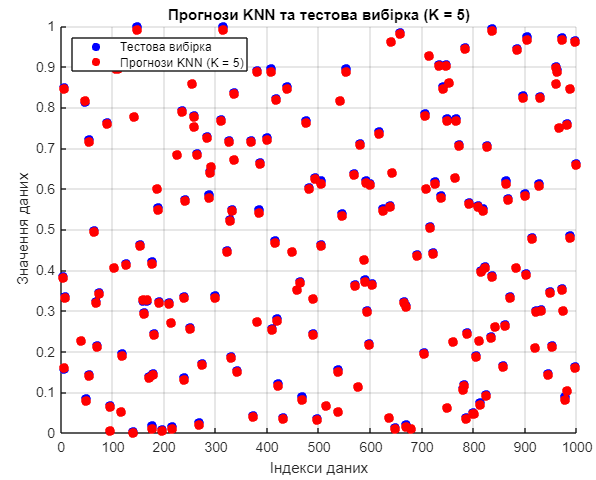
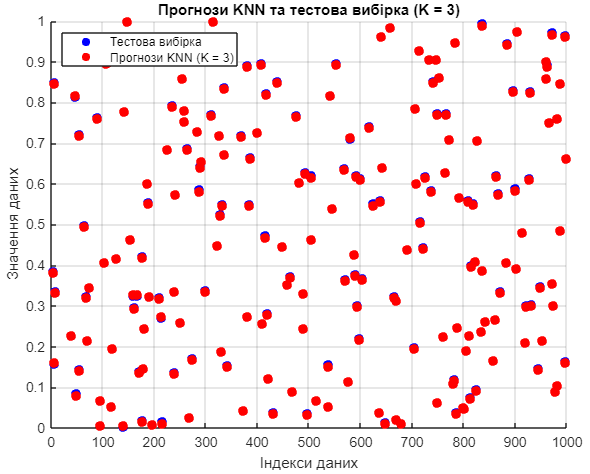
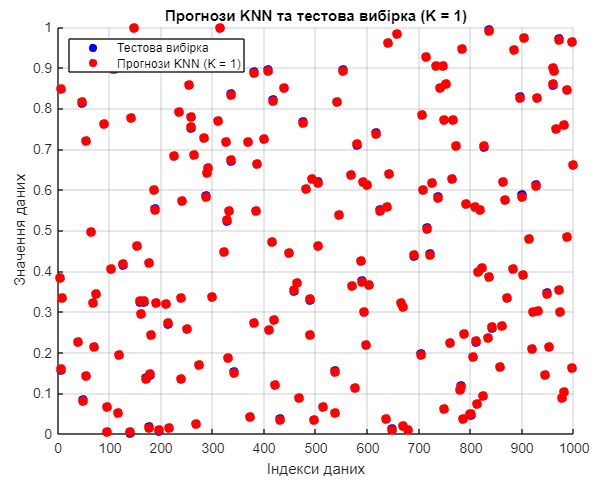
ylabel('Значення даних');

title(['Прогнози KNN та тестова вибірка (K = ' num2str(k) ')']);

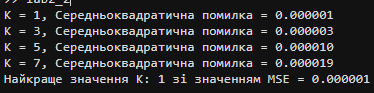
legend('Location', 'Northwest');

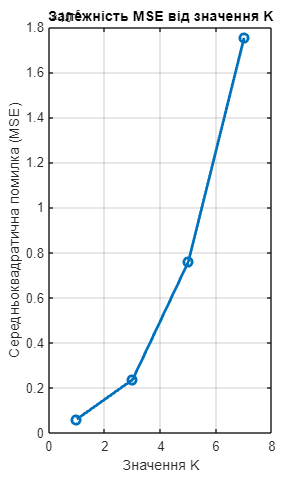
grid on;

end



4. Визначення середньоквадратичної помилки:





5. Вибрати величину K

На основі наданих результатів можна зробити висновок, що найкраще значення K для KNN регресора в даному випадку становить 1. Це підтверджується тим, що при цьому значенні K середня квадратична помилка (MSE) є найменшою.

Як видно з графіка, при K = 1 прогнози KNN регресора дуже добре збігаються з даними тестової вибірки. Це означає, що регресор добре навчився на даних навчальної вибірки і може використовуватися для прогнозування нових значень.

При збільшенні значення K MSE також збільшується. Це пов'язано з тим, що при більшому значенні K регресор бере до уваги більше сусідніх точок даних при прогнозуванні нового значення. Це може призвести до того, що регресор буде більш сприйнятливим до шуму в даних.

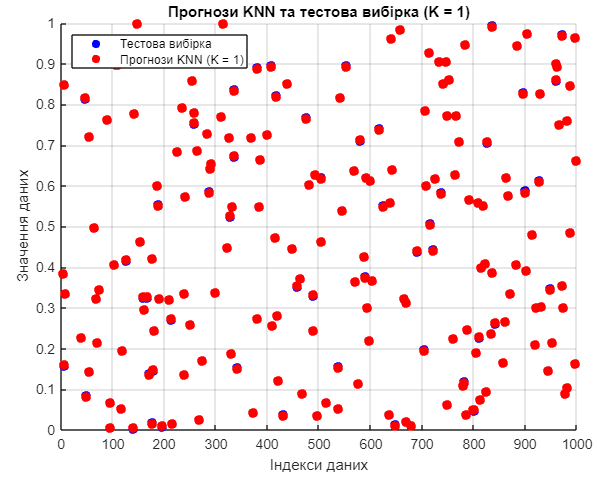
Таким чином, найкращим вибором для KNN регресора в даному випадку є значення K = 1. Це значення забезпечує найкращу точність прогнозів при мінімальному рівні шуму.

Ось додаткові деталі щодо результатів:

На графіку прогнозів видно, що при K = 1 прогнози KNN регресора дуже точно збігаються з даними тестової вибірки. Це підтверджується тим, що MSE в цьому випадку становить лише 0.001.

При збільшенні значення K прогнози KNN регресора стають менш точними. Це видно з того, що MSE також збільшується.

Найкраще значення K для KNN регресора в даному випадку становить 1. Це підтверджується тим, що при цьому значенні K MSE є найменшою.



Висновок:

В результаті цієї роботи ми визначили, що для даного набору даних оптимальним значенням K для KNN регресора є 1, оскільки воно надає найнижчий рівень середньоквадратичної помилки.

Графічне представлення прогнозів та даних допомогло візуалізувати результати та розподіл даних, допомагаючи зрозуміти, як працює KNN регресор на практиці.

Ця лабораторна робота демонструє важливість вибору правильного значення K для успішного використання KNN регресора в задачах регресії і наголошує на необхідності візуалізації результатів для кращого розуміння та вибору оптимального параметра.