**Міністерство освіти і науки України**

**Національний технічний університет України «Київський політехнічний**

**інститут імені Ігоря Сікорського"**

**Факультет інформатики та обчислювальної техніки**

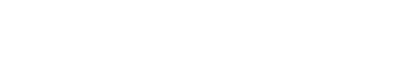
**Кафедра інформатики та програмної інженерії**

# Звіт

з лабораторної роботи № 4 з дисципліни

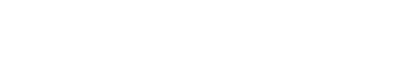
«Класифікація методом k найближчих сусідів і набір даних Digits,

частина 1»



**Виконав** *ІП-15, Дацьо Іван*

(шифр, прізвище, ім'я, по батькові)



**Перевірив**



прізвище, ім'я, по батькові



*Нестерук А.*

Київ 2023

# Завдання:

1. Для дослідження даних, візуалізуйте їх. Виведіть зображення перших 36 цифр з набору

2. Розбийте дані на навчальні та тестові, за замовчуванням train\_test\_split резервує 75% даних для навчання і 25% для тестування, змініть це.

3. Створити та навчити модель

4. Виконайте прогнозування класів

5. Порівняйте прогнозовані цифри з очікуваними для перших 36 тестових зразків.

6. Поясніть результат, застосуйте метрики точності моделі.

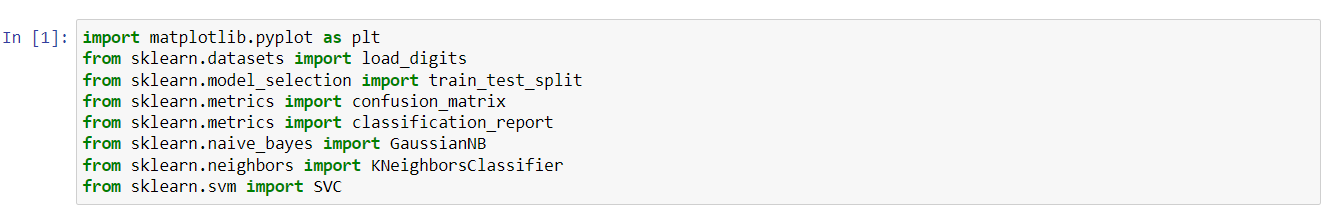
7. Виведіть звіт класифікації

8. Використайте декілька моделей KNeighborsClassifier, SVC и GaussianNB для пошуку найкращої

9. Налаштуйте гіперпараметр K в KNeighborsClassifier

# Виконання

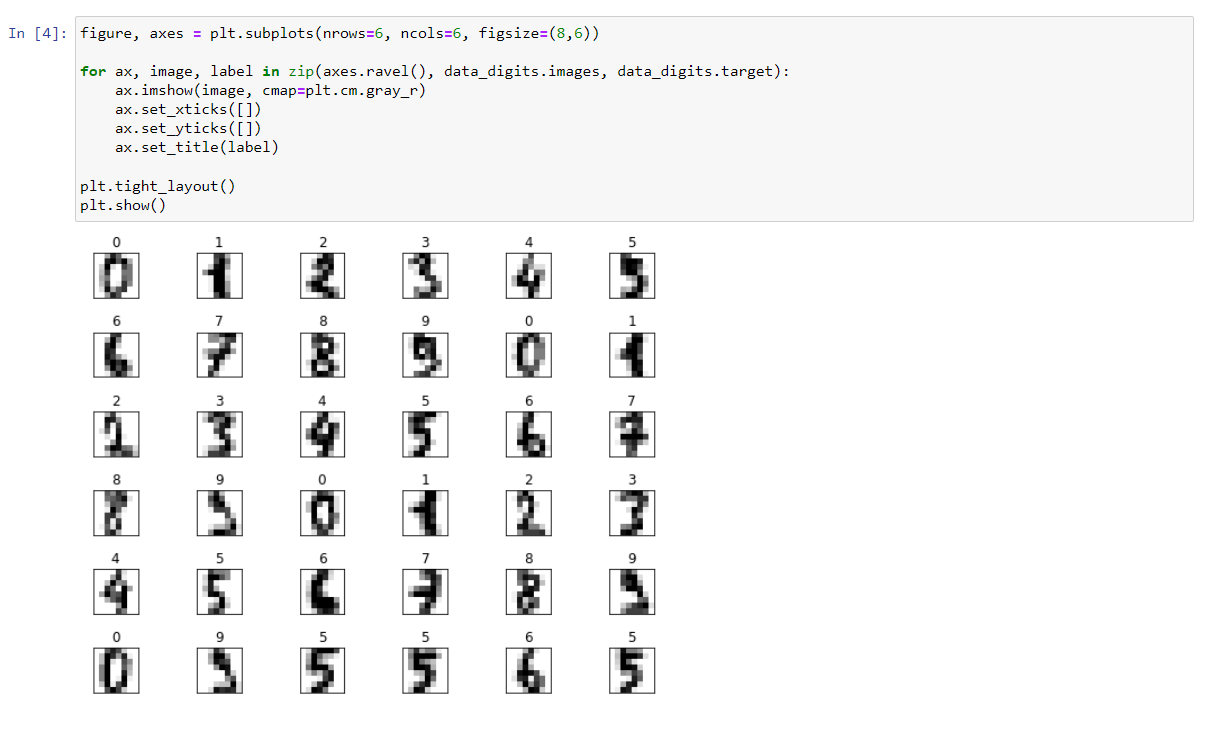
1. Імпортуємо необхідні пакети:



Завантажуємо вбудований в бібліотеку scikiklearn набір даних:



Візуалізуємо перших 36 цифр з набору даних:



1. Розіб'ємо дані із попереднього пункту на навчальні та тестові(75% до 25%):



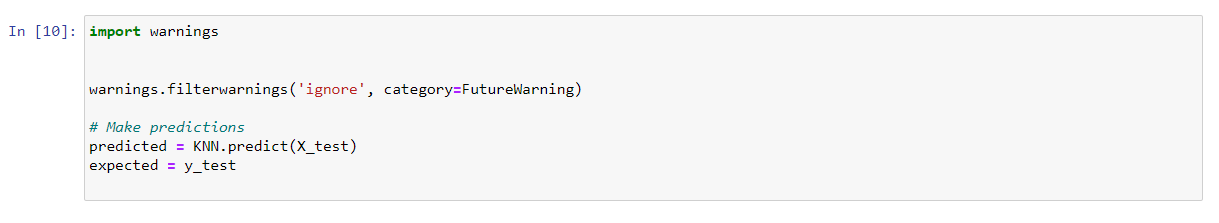
Перевіримо розбиття даних:



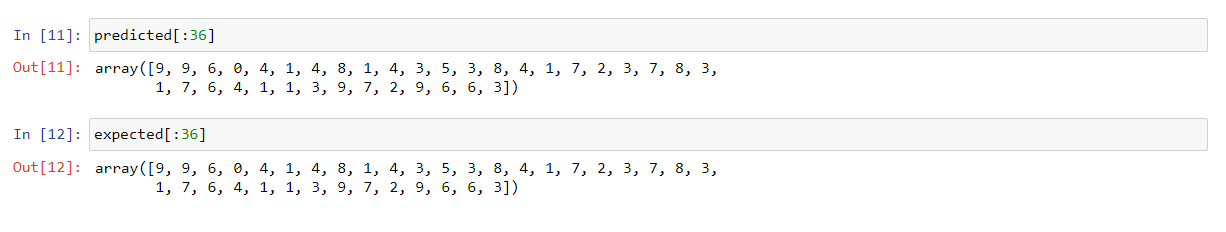
1. Створюємо та навчаємо модель на основі алгориту k найближчих сусідів.



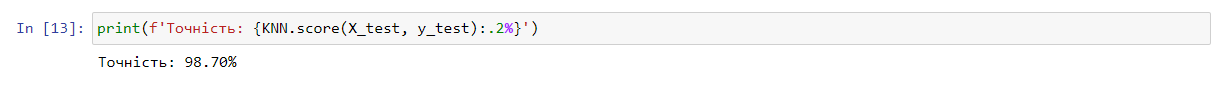
1. Виконаємо прогнозування класів



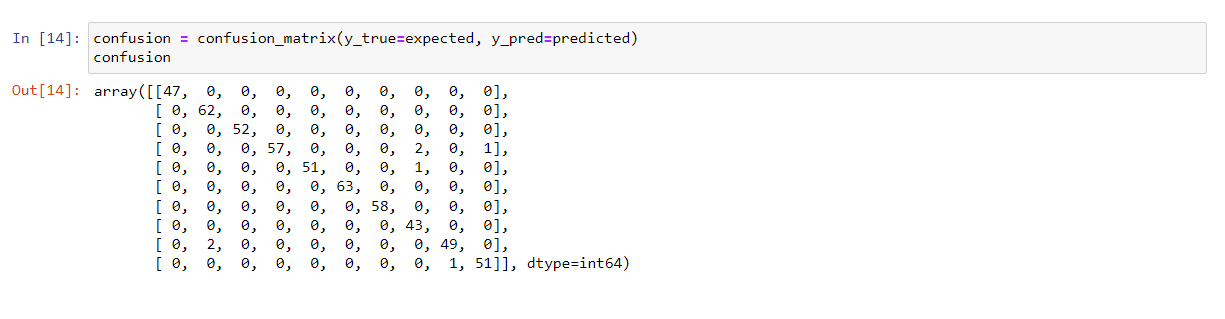
1. Порівняємо прогнозовані цифри з очікуваними для перших 36



1. Оцінимо якість моделі використовуючи метод score

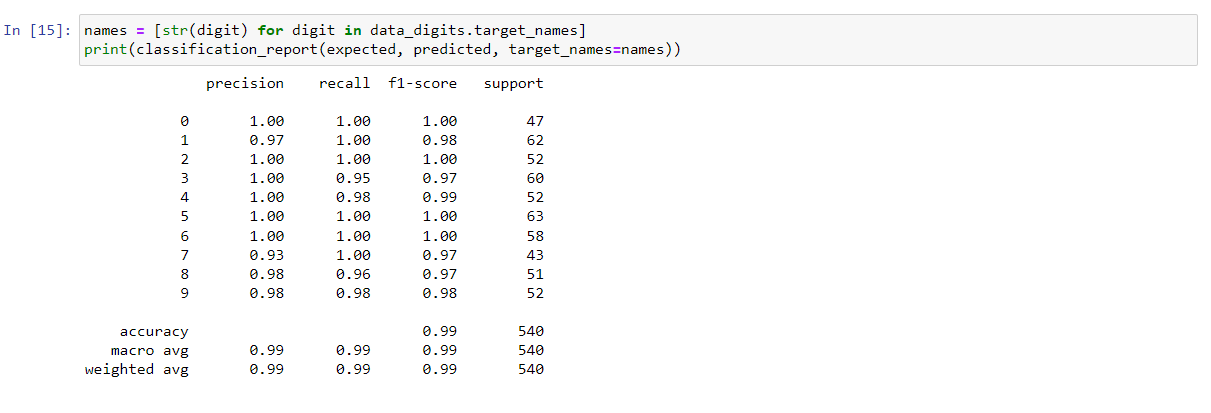


Створимо та виведемо матрицю невідповідностей



Можна створити висновок дана модель є достатньо точною.

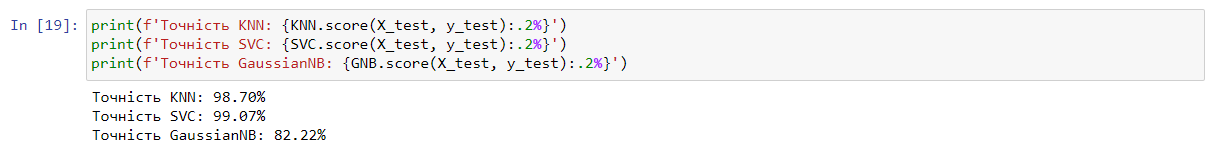
1. Використовуючи функцію sklearn.metrics виведемо звіт класифікації



1. Для пошуку кращої моделі створимо та навчимо моделі SVC та GaussianNB

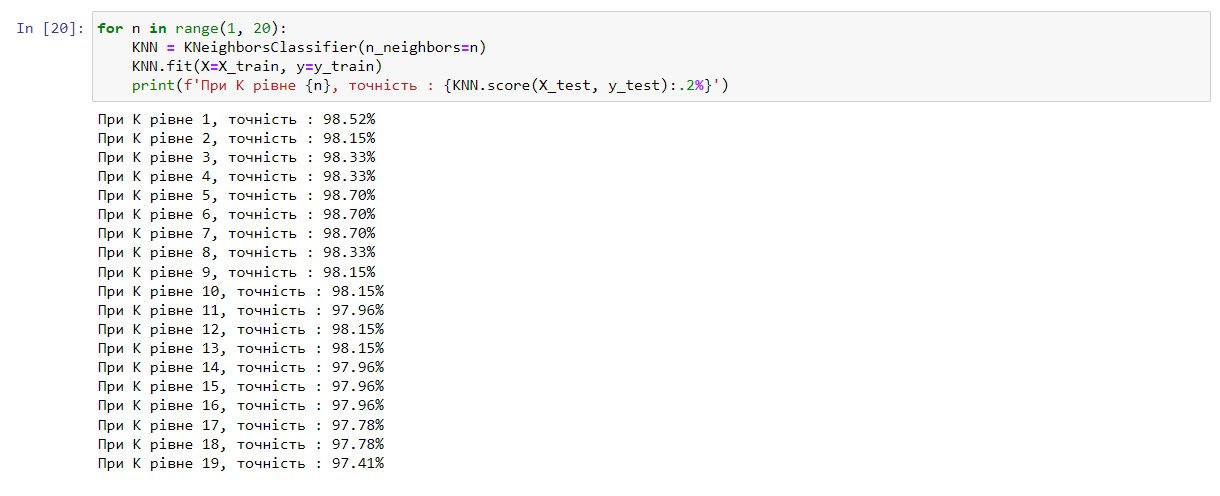


Оцінимо якість моделей використовуючи метод score



Як бачимо в даному випадку найгірше себе показав GaussianNB

1. Поекспериментуємо з різними значеннями k і порівняємо їх ефективність



В даному випадку найкращими параметрами стали декілька, а саме k = 5, k = 6, k = 7. Отже значення параметру за замовчуванням (k=5) є ефективним.

**Висновок**

Виконавши дану лабораторну я навчився реалізовувати основні етапи машинного навчання, а саме вибір даних для навчання моделі, завантаження та аналіз даних для навчання і тестування, вибір і побудова моделі, навчання модулі, навчання моделі, формування прогнозів, проведення оцінки результатів, налаштування параметрів моделі, опрацьовування декількох класифікаційних моделей. Все це було досліджено на основі даних із бібліотеки scikiklearn, цифри.