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

Практикум №6

з курсу «Аналіз даних в інформаційнних системах»

на тему: «Класифікація та кластеризація»

|  |  |
| --- | --- |
| Викладач:  Ліхоузова Т.А. | Виконав:  студент 2 курсу  групи ІП-11 Сідак Кирил з ФІОТ |

Київ-2023

ЗМІСТ

[1. ЗАВДАННЯ 3](#_Toc131621052)

[2. ОСНОВНЕ ЗАВДАННЯ 4](#_Toc131621053)

[3. ДОДАТКОВЕ ЗАВДАННЯ 5](#_Toc131621054)

[4. ВИСНОВОК 6](#_Toc131621055)

# ЗАВДАННЯ

**Основне завдання**

Для даних по титаніку titanic.csv побудувати модель, в якій можна визначити, чи виживе пасажир, заповнивши решту параметрів.

Використати декілька методів. Порівняти результати.

**Додаткове завдання**

Використовуючи файл Data2.csv

1. визначити, який регіон домінує в кластерах по ВВП на душу населення та щільності населення
2. вивести частотні гістограми всіх показників файла Data2.csv, використовуючи цикл
3. створити функцію, яка на вхід отримує два набори даних, перевіряє чи є лінійна залежність та виводить True чи False (будемо розуміти під «є лінійна залежність», якщо коефіцієнт кореляції по модулю більше 0,8)

# ОСНОВНЕ ЗАВДАННЯ

Звіт по основному завданню було оформлено за допомогою блокноту з кодом.

# ДОДАТКОВЕ ЗАВДАННЯ

Звіт по додатковому завданню було оформлено за допомогою блокноту з кодом.

# ВИСНОВОК

Отже, при виконанні даної лабораторної було досліджено набір даних про пасажирів Титаніку, створено 3 різні моделі для класифікації на основі цього набору даних для визначення, чи вижив пасажир, зокрема: KNN, SVM, логістична регресія. Крім того, у рамках додаткового завдання було на основі іншого набору даних визначено, який регіон домінує в кластерах по ВВП на душу населення та щільності населення за допомогою алгоритму K середніх, виведено частотні гістограми всіх показників цього датасету за допомогою циклу, а також створено функцію, яка на вхід отримує два набори даних, перевіряє чи є лінійна залежність та виводить True чи False, якщо коефіцієнт кореляції по модулю більше або менший чи рівний 0,8 відповідно.