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 інформаційних системах»

на тему: «Класифікація та кластеризація»

|  |  |
| --- | --- |
| Викладач:  Ліхоузова Т. А. | Виконав:  студент 2 курсу  групи ІП-11 ФІОТ  Головня О. Р. |

Київ-2023

**Основне завдання** Приклад виконання

[Скачати потрібні дані](https://drive.google.com/drive/folders/1kYY4pxkeLXoszCX8icpFUrhFCnwgAyhz?usp=sharing).

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

Використати декілька методів. Порівняти результати.

Для виконання лабораторної роботи було використано Python.

Код програми:

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LogisticRegression

from sklearn.tree import DecisionTreeClassifier

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy\_score

# Завантаження даних з файлу

data = pd.read\_csv('titanic.csv')

# Видалення непотрібних стовпців

data = data.drop(['PassengerId', 'Name', 'Ticket', 'Cabin'], axis=1)

# Обробка пропущених значень

data = data.dropna()

# Перетворення категоріальних змінних в числові

data['Sex'] = data['Sex'].map({'male': 0, 'female': 1})

data['Embarked'] = data['Embarked'].map({'S': 0, 'C': 1, 'Q': 2})

# Розділення даних на навчальний та тестовий набори

X = data.drop('Survived', axis=1)

y = data['Survived']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Побудова моделі логістичної регресії

lr\_model = LogisticRegression()

lr\_model.fit(X\_train, y\_train)

# Перевірка точності на тестовому наборі

y\_pred\_lr = lr\_model.predict(X\_test)

accuracy\_lr = accuracy\_score(y\_test, y\_pred\_lr)

print('Accuracy of Logistic Regression:', accuracy\_lr)

# Побудова моделі дерева рішень

dt\_model = DecisionTreeClassifier()

dt\_model.fit(X\_train, y\_train)

# Перевірка точності на тестовому наборі

y\_pred\_dt = dt\_model.predict(X\_test)

accuracy\_dt = accuracy\_score(y\_test, y\_pred\_dt)

print('Accuracy of Decision Tree:', accuracy\_dt)

# Побудова моделі випадкового лісу

rf\_model = RandomForestClassifier(n\_estimators=100)

rf\_model.fit(X\_train, y\_train)

# Перевірка точності на тестовому наборі

y\_pred\_rf = rf\_model.predict(X\_test)

accuracy\_rf = accuracy\_score(y\_test, y\_pred\_rf)

print('Accuracy of Random Forest:', accuracy\_rf)

# Порівняння результатів

if accuracy\_lr > accuracy\_dt and accuracy\_lr > accuracy\_rf:

print('Logistic Regression is the best model')

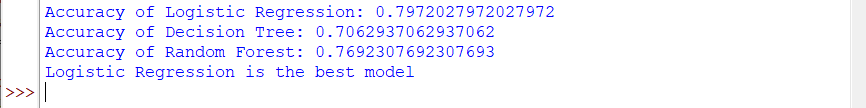
elif accuracy\_dt > accuracy\_lr and accuracy\_dt > accuracy\_rf:

print('Decision Tree is the best model')

else:

print('Random Forest is the best model')

1. **Основне завдання:**

****Результат виконання:

**Висновок:**

Під час виконання лабораторної роботи я дослідив дані: побудував декілька моделей, в яких можна визначити, чи виживе пасажир. Також порівняв результати.