

Лабораторна робота №5 ІАД

Завдання

1. Створити набір даних за допомогою DatasetGenerator
 - а. Перейдіть на [DatasetGenerator](#) і створіть набір даних. Використовуйте зручні вам параметри.
2. Завантажити набір даних для класифікації з сайту Kaggle
 - а. Виберіть унікальний набір даних для кожного студента на [Kaggle](#). Для цього вам потрібен акаунт на Kaggle.
3. Опціонально для отримання відмінної оцінки: Згенерувати набір даних для класифікації
 - а. Використовуйте стандартні можливості Python, такі як numpy або random.

```
import numpy as np
import pandas as pd

# Генерація випадкових даних
X = np.random.rand(100, 5) # 100 зразків, 5 ознак
y = np.random.randint(2, size=100) # 100 цільових значень (0 або 1)

# Створення DataFrame
data = pd.DataFrame(X, columns=[f'Feature{i}' for i in range(1, 6)])
data['Target'] = y

# Збереження в CSV файл
data.to_csv('generated_dataset.csv', index=False)
```

Провести моделювання для отриманих наборів даних з використанням регресійного аналізу

- Використовуйте бібліотеки scikit-learn для моделювання.

```

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

# Завантаження даних
data = pd.read_csv('generated_dataset.csv')
X = data.drop('Target', axis=1)
y = data['Target']

# Розподіл на тренувальний та тестовий набори
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Модель логістичної регресії
model = LogisticRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

# Оцінка моделі
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)

print(f'Accuracy: {accuracy}')
print(f'Precision: {precision}')
print(f'Recall: {recall}')
print(f'F1-Score: {f1}')

```

Висновок

У ході виконання лабораторної роботи було створено та використано кілька наборів даних для задач класифікації. Ми використовували DatasetGenerator, дані з Kaggle, а також генерували власні дані за допомогою бібліотек numpy та random. Проведено моделювання з використанням логістичної регресії, та результати було оцінено за допомогою метрик точності, precision, recall та F1-score. Усі результати було викладено на GitHub для подальшого аналізу та збереження.

Цей процес допоміг зрозуміти основні етапи аналізу даних та машинного навчання, а також надати практичні навички роботи з різними інструментами та платформами. Якщо у вас виникнуть якісь питання, звертайтеся, завжди радий допомогти