

Лабораторна робота №5 ІАД

Виконав студент групи МІТ-31
Тимохін Роман Миколайович

Завдання

1. Створити набір даних за допомогою [DatasetGenerator](#).
2. Завантажити набір даних для класифікації з сайту [Kaggle](#). Кожен студент має обрати *унікальний набір даних*.
3. (Опціонально для отримання відмінної оцінки) Згенерувати набір даних для класифікації, використовуючи стандартні можливості Python (наприклад, з допомогою бібліотек NumPy або random).
4. Для отриманих наборів даних провести встановленні кроки моделювання.

Викласти результати на GitHub у власний репозиторій в основну гілку в папці [Lab5](#).

Код і результати

```
from sklearn.datasets import make_classification
import pandas as pd

# Генеруємо набір даних
X, y = make_classification(
    n_samples=1000, # кількість зразків
    n_features=10, # кількість ознак
    n_informative=5, # кількість інформативних ознак
    n_classes=3, # кількість класів
    random_state=42 # фіксація генерації
```

)

Зберігаємо у форматі DataFrame

```
data = pd.DataFrame(X, columns=[f'feature_{i}' for i in range(10)])
```

```
data['target'] = y
```

Збереження у файл

```
data.to_csv('generated_dataset.csv', index=False)
```

```
print("Набір даних створено та збережено як 'generated_dataset.csv'.")
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.ensemble import RandomForestClassifier
```

```
from sklearn.metrics import classification_report
```

Завантаження даних

```
data = pd.read_csv('generated_dataset.csv')
```

Поділ на ознаки та мітки

```
X = data.drop('target', axis=1)
```

```
y = data['target']
```

Поділ на тренувальний та тестовий набори

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,  
random_state=42)
```

Створення моделі

```
model = RandomForestClassifier(random_state=42)
```

```
model.fit(X_train, y_train)
```

```
# Прогноз
```

```
y_pred = model.predict(X_test)
```

```
# Оцінка результатів
```

```
print(classification_report(y_test, y_pred))
```

```
Набір даних створено та збережено як 'generated_dataset.csv'.
```

	precision	recall	f1-score	support
0	0.85	0.92	0.88	90
1	0.82	0.85	0.84	96
2	0.91	0.82	0.86	114
accuracy			0.86	300
macro avg	0.86	0.86	0.86	300
weighted avg	0.86	0.86	0.86	300

Висновок:

Ця робота демонструє ключові аспекти аналізу даних і моделювання, що є фундаментом для успішного вирішення задач класифікації.