

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ
УНІВЕРСИТЕТ
імені ТАРАСА ШЕВЧЕНКА**



ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Кафедра прикладних інформаційних систем

Лабораторна роботи №2

*Студента 5 курсу
групи ІС-11
ОП «Інформаційні системи»
Кадирова К. Б.*

Київ – 2023

```

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.pipeline import Pipeline
from sklearn.model_selection import GridSearchCV
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score, r2_score
from joblib import dump, load

# Завантаження датасету (приклад з використанням датасету Iris)
iris = pd.read_csv('iris.csv')

# Розділення на ознаки та цільову змінну
X = iris.drop('species', axis=1)
y = iris['species']

# Розділення на тренувальну та тестову вибірки
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Створення пайплайну
pipeline = Pipeline([
    ('scaler', StandardScaler()),
    ('pca', PCA()),
    ('classifier', LogisticRegression())
])

# Налаштування гіперпараметрів та відбір ознак за допомогою GridSearchCV
parameters = {
    'scaler__with_mean': [True, False],
    'pca__n_components': [2, 3],
    'classifier__C': [0.1, 1, 10]
}
grid_search = GridSearchCV(pipeline, parameters)
grid_search.fit(X_train, y_train)

# Оцінка якості на тестовій вибірці
y_pred = grid_search.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='weighted')
recall = recall_score(y_test, y_pred, average='weighted')

# Виведення результатів
print("Accuracy:", accuracy)
print("Precision:", precision)
print("Recall:", recall)

# Збереження навченого пайплайну
dump(grid_search, 'pipeline.joblib')

# Завантаження навченого пайплайну

```

```
loaded_pipeline = load('pipeline.joblib')
```