Лабараторна робота 15

import pandas as pd

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

from sklearn.preprocessing import LabelEncoder

from sklearn.neural\_network import MLPClassifier

from sklearn.ensemble import GradientBoostingClassifier

from sklearn.metrics import accuracy\_score

data = pd.read\_csv('C:\\Users\\baras\\OneDrive\\Рабочий стол\\Unik\\SSSHII\\LB15\\orange\_quality.csv')

label\_encoder = LabelEncoder()

data['Color'] = label\_encoder.fit\_transform(data['Color'])

data['Variety'] = label\_encoder.fit\_transform(data['Variety'])

data['Blemishes (Y/N)'] = label\_encoder.fit\_transform(data['Blemishes (Y/N)'])

X = data.drop('Quality (1-5)', axis=1)

y = data['Quality (1-5)'].round().astype(int)

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

mlp = MLPClassifier(hidden\_layer\_sizes=(10, 10), max\_iter=500, random\_state=42)

mlp.fit(X\_train, y\_train)

mlp\_predictions = mlp.predict(X\_test)

mlp\_accuracy = accuracy\_score(y\_test, mlp\_predictions)

print(f"Точність нейронної мережі: {mlp\_accuracy:.2f}")

gb = GradientBoostingClassifier(random\_state=42)

param\_grid = {

    'n\_estimators': [50, 100],

    'learning\_rate': [0.01, 0.1],

    'subsample': [0.7, 1.0],

    'max\_depth': [3, 5]

}

grid\_search = GridSearchCV(estimator=gb, param\_grid=param\_grid, cv=3, scoring='accuracy', n\_jobs=-1, verbose=1)

grid\_search.fit(X\_train, y\_train)

best\_params = grid\_search.best\_params\_

best\_score = grid\_search.best\_score\_

print(f"Найкращі параметри: {best\_params}")

print(f"Точність з найкращими параметрами: {best\_score:.2f}")

Вихід

C:\Python312\Lib\site-packages\sklearn\neural\_network\\_multilayer\_perceptron.py:690: ConvergenceWarning: Stochastic Optimizer: Maximum iterations (500) reached and the optimization hasn't converged yet.

warnings.warn(

Точність нейронної мережі: 0.61

Fitting 3 folds for each of 16 candidates, totalling 48 fits

Найкращі параметри: {'learning\_rate': 0.1, 'max\_depth': 3, 'n\_estimators': 50, 'subsample': 0.7}

Точність з найкращими параметрами: 0.72

**Висновок**

У ході виконання роботи було проведено порівняння ефективності двох моделей машинного навчання для класифікації якості апельсинів: нейронної мережі та ансамблю градієнтного бустингу. Результати експериментів підтвердили перевагу ансамблевих методів над базовою нейронною мережею.

1. **Нейронна мережа**:
   * Нейронна мережа досягла точності 61%, що є базовим показником для порівняння. Проте для табличних даних з числовими та категоріальними ознаками нейронні мережі можуть бути менш ефективними, оскільки вони зазвичай добре працюють на завданнях із більш складною структурою даних (наприклад, зображення або текст).
2. **Ансамбль градієнтного бустингу**:
   * Градієнтний бустинг продемонстрував вищу точність (72%) після підбору оптимальних гіперпараметрів, що підтверджує його здатність добре працювати з табличними даними.
   * Оптимізація гіперпараметрів (кількості дерев, швидкості навчання, розміру підвибірки та глибини дерев) дозволила покращити результати моделі. Це свідчить про важливість налаштування моделі для досягнення максимальної продуктивності.
3. **Переваги градієнтного бустингу**:
   * Ансамбль градієнтного бустингу виявився більш стійким до особливостей даних, таких як розподіл значень ознак, завдяки механізму послідовного коригування помилок попередніх дерев.
   * Завдяки можливості налаштування гіперпараметрів, градієнтний бустинг дозволяє досягти оптимального балансу між узагальнюючою здатністю моделі та запобіганням перенавчанню.
4. **Висновки**:
   * Градієнтний бустинг є ефективним методом для класифікаційних задач на табличних наборах даних. У цій роботі ансамбль показав суттєве покращення точності в порівнянні з нейронною мережею, що робить його кращим вибором для даного набору даних.
   * Подальше дослідження може включати більш детальний підбір гіперпараметрів або тестування інших алгоритмів ансамблю, таких як XGBoost або LightGBM, які також засновані на градієнтному бустингу і можуть надати ще кращі результати.

Отже, використання ансамблевих методів для аналізу якості продукції є виправданим підходом, особливо в задачах, де необхідно досягти високої точності прогнозування.