Assignment 5 - Classification Model Experimentation

and Deployment Instructions

Dataset: Drug Classification

Link: https://shorturl.asia/xemB1

Problems:

This dataset contains information about drug classification based on patient general information and its diagnosis. Machine learning model is needed in order to predict the outcome of the drugs type that might be suitable for the patient.

Dataset Description:

There are 6 variables in this data set: 4 categorical variables and 2 continuous variables.

The following is the structure of the data set.

| Variable Name	Description
Age	Patient Age
Sex	Gender of patient (male or female)
BP	Levels of blood pressure (high, normal, or low)
Cholesterol	Levels of cholesterol (high or normal)
Na_to_K	Sodium to potassium ratio in blood
Drug	Type of drug

The machine learning models used in this assignment are:

1. K Neighbours (KNN)
2. Logistic Regression
3. Support Vector Machine (SVM)

K Neighbours (KNN)

```
              precision    recall  f1-score   support

           0     0.8333    1.0000    0.9091        15
           1     1.0000    0.8333    0.9091         6
           2     1.0000    1.0000    1.0000         3
           3     1.0000    0.8000    0.8889         5
           4     1.0000    0.9091    0.9524        11

    accuracy                         0.9250        40
   macro avg     0.9667    0.9085    0.9319        40
weighted avg     0.9375    0.9250    0.9253        40
```

K Neighbours accuracy is: 92.50%

Confusion Matrix: KNN
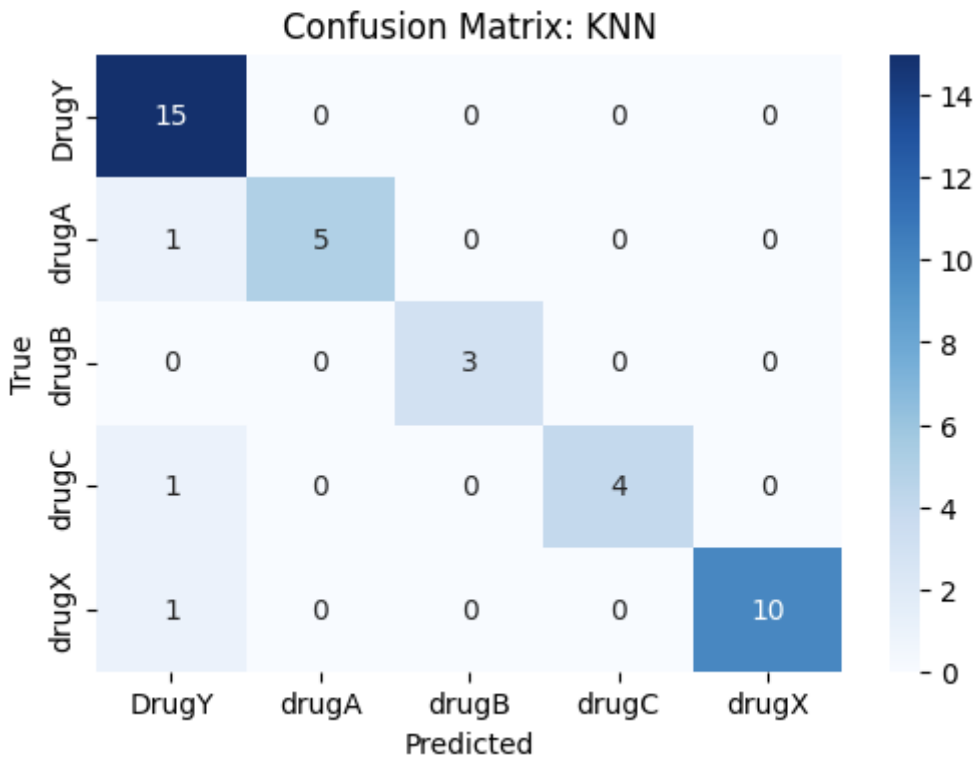
Logistic Regression

```
              precision    recall  f1-score   support

           0     1.0000    0.9333    0.9655        15
           1     0.8571    1.0000    0.9231         6
           2     1.0000    1.0000    1.0000         3
           3     1.0000    1.0000    1.0000         5
           4     1.0000    1.0000    1.0000        11

    accuracy                         0.9750        40
   macro avg     0.9714    0.9867    0.9777        40
weighted avg     0.9786    0.9750    0.9755        40
```

Logistic Regression accuracy is: 97.50%

Confusion Matrix: Logistic Regression
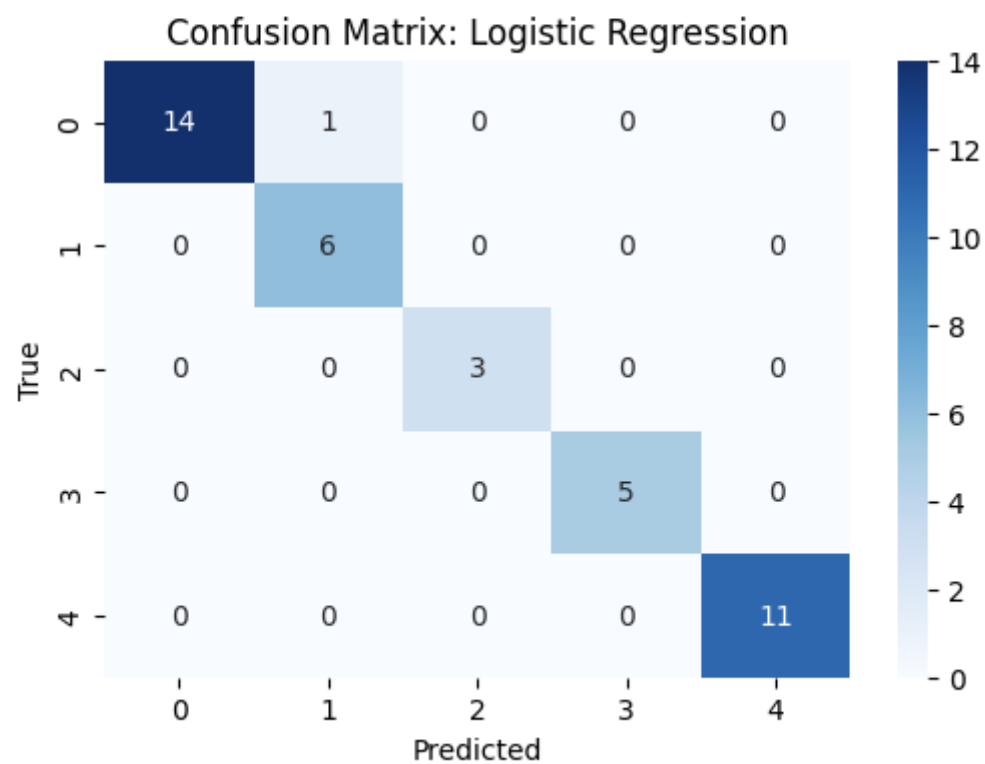
Support Vector Machine (SVM)

	precision	recall	f1-score	support
0	1.0000	1.0000	1.0000	15
1	1.0000	1.0000	1.0000	6
2	1.0000	1.0000	1.0000	3
3	1.0000	1.0000	1.0000	5
4	1.0000	1.0000	1.0000	11
accuracy			1.0000	40
macro avg	1.0000	1.0000	1.0000	40
weighted avg	1.0000	1.0000	1.0000	40

Support Vector Machines accuracy is: 100.00%

Confusion Matrix: SVM
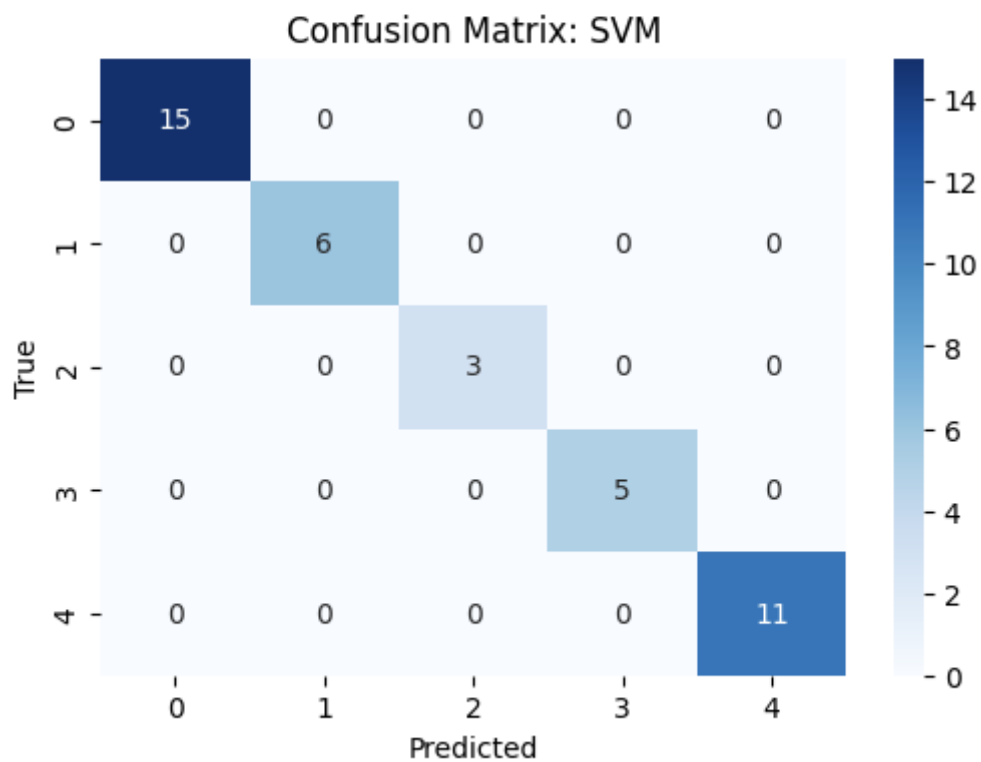
Model Selection

From the results, The Support Vector Machine (SVM) model was able to predict the drug class with 100% accuracy. We used SVM model to create a Streamlit application.

Hyperparameter Tuning (using SVM Model)

Grid Search ¶

```python
[62]: import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split , GridSearchCV
from sklearn.preprocessing import StandardScaler
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score
from sklearn.pipeline import Pipeline

# Create a pipeline with StandardScaler and SVM classifier
pipeline = Pipeline([
    ('scaler', StandardScaler()),
    ('svm', SVC())
])

# Define the hyperparameter grid to search
param_grid = {
    'svm__C': [0.1, 1, 10],
    'svm__kernel': ['linear', 'rbf'],
    'svm__gamma': [0.001, 0.01, 0.1, 1]
}

# Create a GridSearchCV object to perform hyperparameter tuning
grid_search = GridSearchCV(pipeline, param_grid, cv=5, n_jobs=-1)

# Fit the grid search to the data
grid_search.fit(X_train, y_train)

# Get the best parameters and the best estimator
best_params = grid_search.best_params_
best_estimator = grid_search.best_estimator_

print("Best Hyperparameters:", best_params)

# Evaluate the model on the test data
accuracy = best_estimator.score(X_test, y_test)
print("Test Accuracy:", accuracy)
```

```
Best Hyperparameters: {'svm__C': 10, 'svm__gamma': 0.001, 'svm__kernel': 'linear'}
Test Accuracy: 1.0
```

Github: https://github.com/wongsakorn-s/Predict_DrugType

🗋	app_predict_svm_drug.py	Add files via upload
🗋	requirements.txt	Add files via upload
🗋	svm_drug_model.pkl	Add files via upload

Streamlit: https://predictdrugtype-negekzpbz7rba7hggphcrf.streamlit.app

Drug Type Prediction

Enter Age:

51

15   75

Select Sex:

M

Select Blood Pressure Levels (BP):

LOW

Select Cholesterol Levels:

NORMAL

Na to Potassium Ration:

0.000   −   +

Predict

Prediction Result:

drugX