Lab 11 Tasks

Name: Zainab Bibi

Sap ID: 46462

Batch: BSCS-6th semester

Lab Instructor: Ayesha Akram

Task1.

```
import pandas as pd
from sklearn.naive_bayes import GaussianNB
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.preprocessing import LabelEncoder

# Step 1: Load data
dataset = pd.read_csv("C:/Users/Raja/Downloads/public-data.csv")

# Step 2: Handle missing values
dataset.replace( to_replace: " ?", pd.NA, inplace=True)
dataset.dropna(inplace=True)

# Step 3: Encode object columns
encoder = LabelEncoder()
for col in dataset.columns:
    if dataset[col].dtype == "object":
        dataset[col] = encoder.fit_transform(dataset[col])
```

```
# Step 4: Prepare features and label                                    ✓1
X_data = dataset.drop( labels: "Salary", axis=1)  # All columns except Salary
y_data = dataset["Salary"]                # Salary column

# Step 5: Train-test split
X_train, X_test, y_train, y_test = train_test_split( *arrays: X_data, y_data, test_size=0.2, random_state=99)

# Step 6: Train model
gnb = GaussianNB()
gnb.fit(X_train, y_train)

# Step 7: Predict and display
y_pred = gnb.predict(X_test)

print("Accuracy Score:", accuracy_score(y_test, y_pred))
print("\nActual:   ", y_test.values)
print("Predicted:", y_pred)
```

Output:

```
Accuracy Score: 0.7546445570397666

Actual:    [0 0 0 ... 0 0 0]
Predicted: [0 0 0 ... 0 0 0]
```
