```
#SVM - Decision Trees Program
#A0262453H Homework 2(SVM)
#Application - Questions
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, precision_score, recall_score
import numpy as np

# Load dataset
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/wdbc.data"
data = pd.read_csv(url, header=None)

# Preprocessing
X = data.iloc[:, 2:].values
y = data.iloc[:, 1].apply(lambda x: 1 if x == 'M' else 0).values

# Initialize arrays to store results
results_dt1 = []
results_dt2 = []

# Repeat the process 20 times
for _ in range(20):
    # Split the data
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=None)

    # Train DT1
    dt1 = DecisionTreeClassifier(criterion='entropy')
    dt1.fit(X_train, y_train)
    y_train_pred_dt1 = dt1.predict(X_train)
    y_test_pred_dt1 = dt1.predict(X_test)

    # Evaluate DT1
    acc_train_dt1 = accuracy_score(y_train, y_train_pred_dt1)
    acc_test_dt1 = accuracy_score(y_test, y_test_pred_dt1)
    prec_dt1 = precision_score(y_test, y_test_pred_dt1)
    recall_dt1 = recall_score(y_test, y_test_pred_dt1)
    results_dt1.append([acc_train_dt1, acc_test_dt1, prec_dt1, recall_dt1])

    # Train DT2 with a limited tree size
    dt2 = DecisionTreeClassifier(criterion='entropy', max_depth=4) # Example of limited tree depth
    dt2.fit(X_train, y_train)
    y_train_pred_dt2 = dt2.predict(X_train)
    y_test_pred_dt2 = dt2.predict(X_test)

    # Evaluate DT2
    acc_train_dt2 = accuracy_score(y_train, y_train_pred_dt2)
    acc_test_dt2 = accuracy_score(y_test, y_test_pred_dt2)
    prec_dt2 = precision_score(y_test, y_test_pred_dt2)
    recall_dt2 = recall_score(y_test, y_test_pred_dt2)
    results_dt2.append([acc_train_dt2, acc_test_dt2, prec_dt2, recall_dt2])

# Compute average results
avg_results_dt1 = np.mean(results_dt1, axis=0)
avg_results_dt2 = np.mean(results_dt2, axis=0)

print("DT1 (IG): ", avg_results_dt1)
print("DT2 (IG with limited size): ", avg_results_dt2)
```

```
DT1 (IG):  [1.         0.92748538 0.91554468 0.89219763]
DT2 (IG with limited size):  [0.98366834 0.92923977 0.91996553 0.89221434]
```