DWDM ASSIGNMENT – 8

21BCE7371 RADHA KRISHNA GARG

CODE

```
FOR tennis.csv
# Modify the script to use the provided dataset
# Importing necessary libraries
import pandas as pd
import numpy as np
# Function to calculate entropy
def entropy(labels):
"""Calculate the entropy of a list of labels."""
unique labels, counts = np.unique(labels, return counts=True)
probabilities = counts / len(labels)
entropy value = -np.sum(probabilities * np.log2(probabilities))
return entropy value
def information gain(data, split_attribute_name, target_name):
"""Calculate the information gain for a given split attribute."""
total entropy = entropy(data[target name])
values, counts = np.unique(data[split attribute name], return counts=True)
weighted entropy = np.sum([(counts[i] / np.sum(counts)) *
entropy(data.where(data[split attribute name] == values[i]).dropna()
[target name])
for i in range(len(values))])
information gain value = total entropy - weighted entropy
return information_gain_value
 Function to calculate Gini index
def gini index(labels):
"""Calculate the Gini index of a list of labels."""
unique labels, counts = np.unique(labels, return counts=True)
probabilities = counts / len(labels)
gini_index value = 1 - np.sum(probabilities**2)
return gini index value
# Function to find the best splitting criterion
def find best split(data, target name, measure):
"""Find the best splitting criterion based on the specified measure."""
best measure value = 0
best split attribute = None
partitions = None
```

```
for column in data.columns[:-1]:
if measure == 'Information Gain':
current measure value = information gain(data, column, target name)
elif measure == 'Gini Index':
current_measure_value = gini_index(data[column])
if current measure value > best measure value:
best measure value = current measure value
best split attribute = column
if measure == 'Information Gain':
partitions = {value: data[data[best split attribute] == value] for value in
data[best split attribute].unique() }
elif measure == 'Gini Index':
partitions = {value: data[data[best split attribute] == value] for value in
np.unique(data[best_split_attribute])}
return best split attribute, partitions, best measure value
# Load the tennis dataset
tennis data = pd.DataFrame({
'Outlook': ['Sunny', 'Sunny', 'Overcast', 'Rain', 'Rain', 'Rain',
'Overcast', 'Sunny', 'Sunny', 'Rain', 'Sunny', 'Overcast', 'Overcast',
'Rain'],
'Temperature': ['Hot', 'Hot', 'Hot', 'Mild', 'Cool', 'Cool', 'Cool', 'Mild',
'Cool', 'Mild', 'Mild', 'Mild', 'Hot', 'Mild'],
'Humidity': ['High', 'High', 'High', 'Normal', 'Normal', 'Normal',
'High', 'Normal', 'Normal', 'Normal', 'High', 'Normal', 'High'],
'Wind': ['Weak', 'Strong', 'Weak', 'Weak', 'Weak', 'Strong', 'Strong',
'Weak', 'Weak', 'Weak', 'Strong', 'Strong', 'Weak', 'Strong'],
'Play Tennis': ['No', 'No', 'Yes', 'Yes', 'Yes', 'No', 'Yes', 'No', 'Yes',
'Yes', 'Yes', 'Yes', 'Yes', 'No']
})
# Example usage with tennis.csv
print("Information Gain:")
split criteria, data partitions, measure value =
find best split(tennis data, 'Play Tennis', 'Information Gain')
print("Best Splitting Criterion:", split criteria)
print("Data Partitions after Splitting:")
for value, partition in data partitions.items():
print("Partition for {}: \n{}".format(value, partition))
print("Information Gain Value:", measure value)
print("\nGini Index:")
split criteria, data partitions, measure value =
find best split(tennis data, 'Play Tennis', 'Gini Index')
print("Best Splitting Criterion:", split criteria)
print("Data Partitions after Splitting:")
for value, partition in data partitions.items():
print("Partition for {}: \n{}".format(value, partition))
```

OUTPUT

```
botk@botk:/media/botk/OS/Users/krish/Documents/RK/PROJECTS RK/DW
/extensions/ms-python.debugpy-2024.2.0-linux-x64/bundled/libs/de
py
Information Gain:
Best Splitting Criterion: Outlook
Data Partitions after Splitting:
Partition for Sunny:
   Outlook Temperature Humidity
                                   Wind Play Tennis
0
     Sunny
                   Hot
                           High
                                   Weak
                           High Strong
1
     Sunny
                   Hot
                                                 No
7
    Sunny
                  Mild
                           High
                                   Weak
                                                 No
8
                  Cool
                                   Weak
                                                Yes
     Sunny
                         Normal
    Sunny
                         Normal Strong
10
                  Mild
                                                Yes
Partition for Overcast:
     Outlook Temperature Humidity
                                     Wind Play Tennis
2
    0vercast
                    Hot
                             High
                                     Weak
                                                  Yes
6
                    Cool
                                                  Yes
    0vercast
                           Normal Strong
11 Overcast
                    Mild
                             High Strong
                                                  Yes
12
   0vercast
                                     Weak
                     Hot
                           Normal
                                                  Yes
Partition for Rain:
   Outlook Temperature Humidity
                                   Wind Play Tennis
3
                  Mild
                                   Weak
     Rain
                           High
                                                Yes
4
      Rain
                  Cool
                         Normal
                                   Weak
                                                Yes
5
     Rain
                  Cool
                         Normal Strong
                                                 No
                  Mild
9
     Rain
                         Normal
                                   Weak
                                                Yes
13
                  Mild
                           High Strong
     Rain
                                                 No
Information Gain Value: 0.24674981977443933
Gini Index:
Best Splitting Criterion: Outlook
Data Partitions after Splitting:
Partition for Overcast:
     Outlook Temperature Humidity
                                     Wind Play Tennis
2
                    Hot
                                     Weak
    Overcast
                             High
                                                  Yes
6
    0vercast
                    Cool
                           Normal
                                  Strong
                                                  Yes
11 Overcast
                    Mild
                             High Strong
                                                  Yes
12 Overcast
                     Hot
                                                  Yes
                           Normal
                                     Weak
Partition for Rain:
   Outlook Temperature Humidity
                                   Wind Play Tennis
3
                  Mild
                           High
                                   Weak
                                                Yes
      Rain
4
      Rain
                  Cool
                         Normal
                                   Weak
                                                Yes
5
                         Normal Strong
      Rain
                  Cool
                                                 No
                  Mild
9
     Rain
                         Normal
                                   Weak
                                                Yes
13
     Rain
                  Mild
                           High Strong
                                                 No
Partition for Sunny:
   Outlook Temperature Humidity
                                   Wind Play Tennis
0
    Sunny
                                   Weak
                   Hot
                           High
                                                 No
                           High Strong
1
     Sunny
                                                 No
                   Hot
                  Mild
     Sunny
                           High
                                   Weak
                                                 No
8
     Sunny
                  Cool
                         Normal
                                   Weak
                                                Yes
10
     Sunny
                  Mild
                         Normal Strong
                                                Yes
Gini Index Value: 0.6632653061224489
botk@botk:/media/botk/OS/Users/krish/Documents/RK/PROJECTS_RK/DW
```

FOR iris.csv

CODE

```
# Importing necessary libraries
import pandas as pd
import numpy as np
# Function to calculate entropy
def entropy(labels):
"""Calculate the entropy of a list of labels."""
unique labels, counts = np.unique(labels, return counts=True)
probabilities = counts / len(labels)
entropy value = -np.sum(probabilities * np.log2(probabilities))
return entropy value
# Function to calculate information gain
def information gain(data, split attribute name, target name):
"""Calculate the information gain for a given split attribute."""
total entropy = entropy(data[target name])
values, counts = np.unique(data[split_attribute_name], return_counts=True)
weighted entropy = np.sum([(counts[i] / np.sum(counts)) *
entropy(data.where(data[split attribute name] == values[i]).dropna()
[target name])
for i in range(len(values))])
information gain value = total entropy - weighted entropy
return information gain value
# Function to calculate Gini index
def gini index(labels):
"""Calculate the Gini index of a list of labels."""
unique labels, counts = np.unique(labels, return counts=True)
probabilities = counts / len(labels)
gini index value = 1 - np.sum(probabilities**2)
return gini index value
# Function to find the best splitting criterion
def find best split(data, target_name, measure):
"""Find the best splitting criterion based on the specified measure."""
best measure value = 0
best split attribute = None
partitions = None
for column in data.columns[:-1]:
if measure == 'Information Gain':
current measure value = information gain(data, column, target_name)
elif measure == 'Gini Index':
current measure value = gini index(data[column])
```

```
best split attribute = column
if measure == 'Information Gain':
partitions = {value: data[data[best split attribute] == value] for value in
data[best_split_attribute].unique() }
elif measure == 'Gini Index':
partitions = {value: data[data[best_split_attribute] == value] for value in
np.unique(data[best_split_attribute])}
return best split attribute, partitions, best measure value
# Load the iris dataset
iris data = pd.read csv('iris.csv')
# Example usage with iris.csv
print("Information Gain:")
split criteria, data partitions, measure value = find best split(iris data,
'Species', 'Information Gain')
print("Best Splitting Criterion:", split criteria)
print("Data Partitions after Splitting:")
for value, partition in data partitions.items():
print("Partition for {}: \n{}".format(value, partition))
print("Information Gain Value:", measure value)
print("\nGini Index:")
split criteria, data partitions, measure value = find best split(iris data,
'Species', 'Gini Index')
print("Best Splitting Criterion:", split criteria)
print("Data Partitions after Splitting:")
for value, partition in data partitions.items():
print("Partition for {}: \n{}".format(value, partition))
print("Gini Index Value:", measure value)
```

if current_measure_value > best_measure_value:
best measure value = current measure value

```
/extensions/ms-python.debugpy-2024.2.0-linux-x64/bundled/libs/debugpy/adapter/../../
[b\].py
Information Gain:
Best Splitting Criterion: Id
Data Partitions after Splitting:
Partition for 1:
  Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm
 1
                             3.5
                                                          0.2 Iris-setosa
               5.1
                                            1.4
Partition for 2:
  Id SepalLengthCm
                     SepalWidthCm PetalLengthCm PetalWidthCm
                                                          0.2 Iris-setosa
                             3.0
                                            1.4
Partition for 3:
  Id SepalLengthCm
                     SepalWidthCm PetalLengthCm PetalWidthCm
                                                          0.2 Iris-setosa
               4.7
                             3.2
                                            1.3
Partition for 4:
                                                                   Species
  Id SepalLengthCm
                     SepalWidthCm PetalLengthCm
                                                 PetalWidthCm
                4.6
                             3.1
                                            1.5
                                                          0.2 Iris-setosa
Partition for 5:
  Id SepalLengthCm
                     SepalWidthCm PetalLengthCm
                                                 PetalWidthCm
               5.0
                             3.6
                                            1.4
                                                          0.2 Iris-setosa
Partition for 6:
  Id SepalLengthCm
                     SepalWidthCm PetalLengthCm PetalWidthCm
                                                                   Species
                                                          0.4 Iris-setosa
                5.4
                             3.9
                                            1.7
Partition for 7:
  Id SepalLengthCm
                     SepalWidthCm PetalLengthCm PetalWidthCm
               4.6
                             3.4
                                            1.4
                                                          0.3 Iris-setosa
Partition for 8:
  Id SepalLengthCm
                     SepalWidthCm PetalLengthCm PetalWidthCm
                                                                   Species
                5.0
                             3.4
                                            1.5
                                                          0.2 Iris-setosa
Partition for 9:
  Id SepalLengthCm
                     SepalWidthCm PetalLengthCm PetalWidthCm
               4.4
                              2.9
                                            1.4
                                                          0.2 Iris-setosa
Partition for 10:
                     SepalWidthCm PetalLengthCm PetalWidthCm
  Id SepalLengthCm
                                                                   Species
9 10
                4.9
                              3.1
                                            1.5
                                                          0.1 Iris-setosa
Partition for 11:
   Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm
10 11
                              3.7
                                             1.5
                                                           0.2 Iris-setosa
                5.4
Partition for 12:
   Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm
                                                                   Species
11 12
                              3.4
                                             1.6
                                                           0.2 Iris-setosa
                 4.8
Partition for 13:
   Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm
                                                                   Species
12 13
                                                           0.1 Iris-setosa
                4.8
                              3.0
                                             1.4
Partition for 14:
   Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm
                                                                   Species
13 14
                 4.3
                                                           0.1 Iris-setosa
                              3.0
                                             1.1
Partition for 15:
   Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm
                                                                   Species
14 15
                 5.8
                                                           0.2 Iris-setosa
                              4.0
                                             1.2
Partition for 16:
   Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm
                                                                   Species
                                             1.5
  16
                 5.7
                              4.4
                                                          0.4 Iris-setosa
```

140 141 6	.7 3	.1 5	5.6 2	2.4 Iris-virginica			
Partition for 142: Id SepalLength	Cm SepalWidth	Cm PetalLength	nCm PetalWidth	nCm Species			
141 142 6 Partition for 143:	3.9	.1	0.1	2.3 Iris-virginica			
Id SepalLength	Cm SepalWidth	Cm PetalLength	nCm PetalWidth	nCm Species			
142 143 5 Partition for 144: Id SepalLength	.8 2	7	0.1	1.9 Iris-virginica			
Id SepalLength	Cm SepalWidth	Cm PetalLength	nCm PetalWidth	nCm Species			
143 144 6 Partition for 145: Id SepalLength	.8 3	.2	5.9	2.3 Iris-virginica			
Id SepalLength	Cm SepalWidth	Cm PetalLength	nCm PetalWidth	nCm Species			
144 145 6 Partition for 146:	.7 3	.3	5.7 2	2.5 Iris-virginica			
Id SepalLength	Cm SepalWidth	Cm PetalLength	nCm PetalWidth	nCm Species			
145 146 6 Partition for 147:	.7 3	.0	5.2 2	2.3 Iris-virginica			
Partition for 147:	Cm SenalWidth	Cm Detallenath	oCm DotalWidth	nCm Species			
146 147 6 Partition for 148:	.3 2	.5	5.0 1	l.9 Iris-virginica			
Partition for 148:							
Id SepalLength	Cm SepalWidth	Cm PetalLengti	ncm PetalWidtr 5.2	nCm Species 2.0 Tris-virginica			
147 148 6 Partition for 149: Id SepalLength				1115 VII gINICA			
Id SepalLength	Cm SepalWidth	Cm PetalLength	nCm PetalWidth	nCm Species			
148 149 6 Partition for 150:	3.2	.4	2.4	2.3 ITIS-VITGINICA			
Id SepalLength	Cm SepalWidth	Cm PetalLength	Cm PetalWidth	Cm Species			
149 150 5 Information Gain Valu	.9 e· 1 584962500	.0 .721156	5.1	l.8 Iris-virginica			
Partition for 150: Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm Species 149 150 5.9 3.0 5.1 1.8 Iris-virginica Information Gain Value: 1.584962500721156							
Gini Index:							
Best Splitting Criterion: Id Data Partitions after Splitting:							
Partition for 1: Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm Species							
o 1 SepallengthCm	SepalWidthCm 3.5	PetalLengthCm 1.4	PetalWidthCm 0.2	Species Tris-setosa			
0 1 5.1 Partition for 2:	3.3	2.1	0.2	1113 300030			
Id SepalLengthCm							
1 2 4.9 Partition for 3:	3.0	1.4	0.2	1112-261029			
Id SepalLengthCm	SepalWidthCm						
2 3 4.7 Partition for 4:	3.2	1.3	0.2	Iris-setosa			
Id SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species			
3 4 4.6	3.1	1.5	0.2	Iris-setosa			
Partition for 5: Id SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species			
4 5 5.0	3.6	1.4		Iris-setosa			
Partition for 6: Id SepalLengthCm	SenalWidthCm	Petallength(m	PetalWidthCm	Species			
5 6 5.4		1.7		Iris-setosa			
Partition for 7:	Copoliti deb Co	Dotall contbo	Dotal Width Co	Coories			
Id SepalLengthCm	separwiaincm	Petattengthum	PetatwiathCm	Species			

Partition for 134:						
Id SepalLengthCm 133 134 6.3 Partition for 135: Id SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species		
133 134 6.3	2.8	5.1	1.5	Iris-virginica		
Partition for 135:			5 . 1	- 1		
Id SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species		
134 135 6.1 Partition for 136:	2.6	5.6	1.4	Iris-virginica		
Partition for 136:						
Id SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species		
135 136 7.7	3.0	6.1	2.3	Iris-virginica		
Partition for 137:						
Id SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species		
136 137 6.3	3.4	5.6	2.4	Iris-virginica		
135 136 7.7 Partition for 137: Id SepalLengthCm 136 137 6.3 Partition for 138:						
Id SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species		
137 138 6.4	3.1	5.5	1.8	Iris-virginica		
137 138 6.4 Partition for 139: Id SepalLengthCm						
Id SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species		
138 139 6.0	3.0	4.8	1.8	Iris-virginica		
Partition for 140:						
Id SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species		
139 140 6.9	3.1	5.4	2.1	Iris-virginica		
138 139 6.0 Partition for 140: Id SepalLengthCm 139 140 6.9 Partition for 141:						
Id SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species		
140 141 6.7 Partition for 142:	3.1	5.6	2.4	Iris-virginica		
Partition for 142:						
Id SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species		
141 142 6.9 Partition for 143:	3.1	5.1	2.3	Iris-virginica		
Id SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species		
142 143 5.8 Partition for 144:	2.7	5.1	1.9	Iris-virginica		
Partition for 144:						
Id SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species		
143 144 6.8	3.2	5.9	2.3	Iris-virginica		
143 144 6.8 Partition for 145: Id SepalLengthCm 144 145 6.7 Partition for 146:			5 . 1			
Id SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species		
144 145 6.7	3.3	5.7	2.5	Iris-virginica		
Partition for 146:	Carallus deb car	Datall another	Data Ind Jahon	C		
Id SepalLengthCm	Sepatwidthcm	PetalLengthcm	Petalwidthcm	Species		
145 146 6.7 Partition for 147:	3.0	5.2	2.3	iris-virginica		
		PetalLengthCm		Species		
146 147 6.3	2.5	5.0	1.9	Iris-virginica		
Partition for 148:	Canallui deb ca	Datall another	Data lui debon	Casaisa		
Id SepalLengthCm				Species		
147 148 6.5	3.0	5.2	2.0	Iris-virginica		
Partition for 149:	ConalwidthCm	Dotall another	Dotal Width Cm	Consina		
		PetalLengthCm		Species		
148 149 6.2	3.4	5.4	2.3	Iris-virginica		
Partition for 150:	SonalWid+hCm	Dotall onathCm	Dotal WidthCm	Species		
Id SepalLengthCm 149 150 5.9	Sepatwidincm 3.0	PetalLengthCm 5.1		Iris-virginica		
Gini Index Value: 0.993		5.1	1.8	Ti 15-VII gillica		
		ocuments /DV /DDO	TECTS PK / DWDM	I AR¢		
botk@botk:/media/botk/OS/Users/krish/Documents/RK/PROJECTS_RK/DWDM LAB\$						