Machine Intelligence
Lab Assignment - 3

Name : Adarsh Subhas Nayak
SRN : PES1UG20CS620
Roll No : 54
Date : 26-08-2022

Code :

PES1UG20CS620.py U

Week 3 > PES1UG20CS620.py > ...

```python
'''
Assume df is a pandas dataframe object of the dataset given
'''
import numpy as np
import pandas as pd
import random

'''Calculate the entropy of the enitre dataset'''
# input:pandas_dataframe
# output:int/float
def get_entropy_of_dataset(df):
    # TODO
    entropy = 0
    column_values = df[[df.columns[-1]]].values
    a, unique_count = np.unique(column_values, return_counts = True)
    no_of_instances = len(column_values)
    if no_of_instances <= 1:
        return 0

    probs_array = []
    for i in range(0, len(unique_count)):
        probs = unique_count[i]/no_of_instances
        probs_array.append(probs)

    for probabilities in probs_array:
        if(probabilities!=0):
            entropy = entropy - (probabilities * np.log2(probabilities))
    return entropy

```

PES1UG20CS620.py U

Week 3 > PES1UG20CS620.py > get_entropy_of_dataset

```python
'''Return avg_info of the attribute provided as parameter'''
# input:pandas_dataframe,str   {i.e the column name ,ex: Temperature in the Play tennis dataset}
# output:int/float
def get_avg_info_of_attribute(df, attribute):
    # TODO
    average_info_of_attribute = 0
    attribute_values = df[attribute].values
    unique_attribute_values,unique_attribute_array = np.unique(attribute_values,return_counts = True)
    no_of_instances = len(attribute_values)

    for attribute_value in unique_attribute_values:
        sliced_dataframe = df[df[attribute] == attribute_value]
        instances = sliced_dataframe[[sliced_dataframe.columns[-1]]].values
        instances_unique_values,instances_unique_counts = np.unique(instances, return_counts = True)
        total_count_in_an_instance = len(instances)

        entropy_of_attribute_value = 0
        for i in instances_unique_counts:
            j = i/total_count_in_an_instance
            if j != 0:
                entropy_of_attribute_value = entropy_of_attribute_value - (j*np.log2(j))
        average_info_of_attribute = average_info_of_attribute + entropy_of_attribute_value*(total_count_in_an_instance/no_of_instances)
    return(abs(average_info_of_attribute))


```

PES1UG20CS620.py U

Week 3 > PES1UG20CS620.py > get_entropy_of_dataset

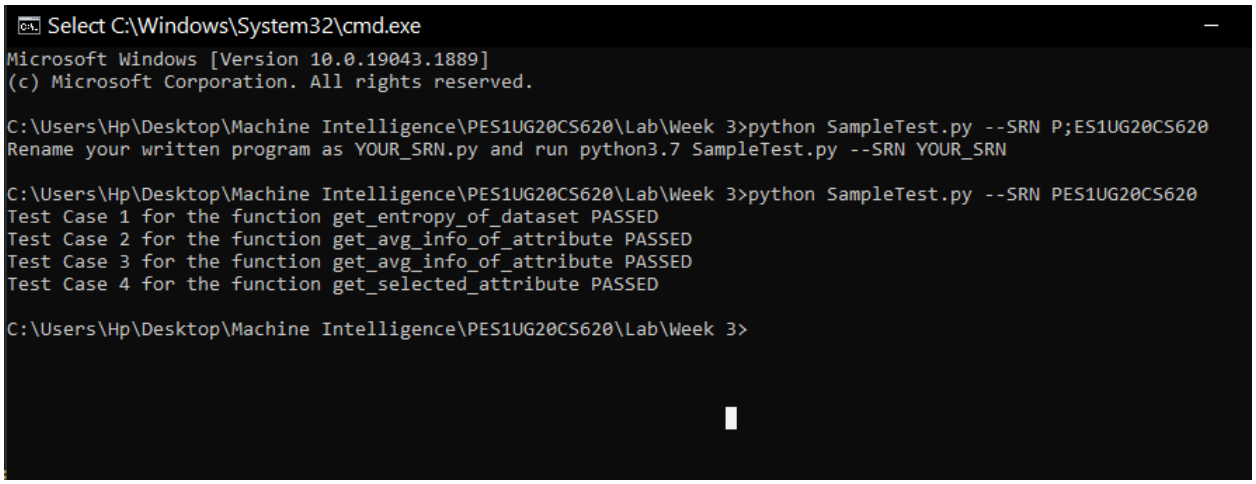
```python
'''Return Information Gain of the attribute provided as parameter'''
# input:pandas_dataframe,str
# output:int/float
def get_information_gain(df, attribute):
    # TODO
    information_gain = 0
    entropy_of_dataset = get_entropy_of_dataset(df)
    entropy_of_attribute = get_avg_info_of_attribute(df, attribute)
    information_gain = entropy_of_dataset - entropy_of_attribute
    return information_gain

#input: pandas_dataframe
#output: ({dict},'str')
def get_selected_attribute(df):
    '''
    Return a tuple with the first element as a dictionary which has IG of all columns
    and the second element as a string with the name of the column selected
    example : ({'A':0.123,'B':0.768,'C':1.23} , 'C')
    '''
    # TODO
    max_information_gain = 0
    information_gain_of_all_attributes = {}
    selected_attribute = ''

    for attribute in df.columns[:-1]:
        information_gain_of_an_attribute = get_information_gain(df, attribute)
        if information_gain_of_an_attribute > max_information_gain:
            max_information_gain = information_gain_of_an_attribute
            selected_attribute = attribute
        information_gain_of_all_attributes[attribute] = information_gain_of_an_attribute
    return (information_gain_of_all_attributes, selected_attribute)

```

Output :


Select C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19043.1889]
(c) Microsoft Corporation. All rights reserved.
C:\Users\Hp\Desktop\Machine Intelligence\PES1UG20CS620\Lab\Week 3>python SampleTest.py --SRN P;ES1UG20CS620
Rename your written program as YOUR_SRN.py and run python3.7 SampleTest.py --SRN YOUR_SRN
C:\Users\Hp\Desktop\Machine Intelligence\PES1UG20CS620\Lab\Week 3>python SampleTest.py --SRN PES1UG20CS620
Test Case 1 for the function get_entropy_of_dataset PASSED
Test Case 2 for the function get_avg_info_of_attribute PASSED
Test Case 3 for the function get_avg_info_of_attribute PASSED
Test Case 4 for the function get_selected_attribute PASSED
C:\Users\Hp\Desktop\Machine Intelligence\PES1UG20CS620\Lab\Week 3>