ML LAB REPORT

1BM18CS061

NIKHIL A S

WEEK 1:

I) FIND S ALGORITHM

```
# THIS PYTHON 3 ENVIRONMENT COMES WITH MANY HELPFUL ANALYTICS LIBRARIES INSTALLED
# IT IS DEFINED BY THE KAGGLE/PYTHON DOCKER IMAGE: HTTPS://GITHUB.COM/KAGGL
E/DOCKER-PYTHON
# FOR EXAMPLE, HERE'S SEVERAL HELPFUL PACKAGES TO LOAD

IMPORT NUMPY AS NP # LINEAR ALGEBRA
IMPORT PANDAS AS PD # DATA PROCESSING, CSV FILE I/O (E.G. PD.READ_CSV)

# INPUT DATA FILES ARE AVAILABLE IN THE READ-ONLY "../INPUT/" DIRECTORY
# FOR EXAMPLE, RUNNING THIS (BY CLICKING RUN OR PRESSING SHIFT+ENTER) WILL LIST ALL
FILES UNDER THE INPUT DIRECTORY

IMPORT OS
FOR DIRNAME, _, FILENAMES IN OS.WALK('/KAGGLE/INPUT'):
    FOR FILENAME IN FILENAMES:
        PRINT(OS.PATH.JOIN(DIRNAME, FILENAME))

# YOU CAN WRITE UP TO 20GB TO THE CURRENT DIRECTORY (/KAGGLE/WORKING/) THAT GETS
PRESERVED AS OUTPUT WHEN YOU CREATE A VERSION USING "SAVE & RUN ALL"  # YOU CAN ALSO
WRITE TEMPORARY FILES TO /KAGGLE/TEMP/, BUT THEY WON'T BE SAVED OUTSIDE OF THE
CURRENT SESSION
```

```
/KAGGLE/INPUT/DATASETCSV/DATA.CSV
```

```
data = pd.read_csv("/kaggle/input/datasetcsv/data.csv")
print("The entered data is \n")
print(data,"\n")
d = np.array(data)[:,:-1]
print("\n The attributes are: \n",d)
target = np.array(data)[:,-1]
print("\n The target is: ",target)
def training(c,t):
    for i, val in enumerate(t):
        if val == "Yes":
            specific_hypothesis = c[i].copy()
            break
    for i, val in enumerate(c):
        if t[i] == "Yes":
            for x in range(len(specific_hypothesis)):In
                if val[x] != specific_hypothesis[x]:
                    specific_hypothesis[x] = '?'
                else:
                    pass
    return specific_hypothesis
print("\n The final hypothesis is:",training(d,target))
```

OUTPUT:

```
THE ENTERED DATA IS

   WEATHER TEMPERATURE HUMIDITY GOES
0  SUNNY          WARM       MILD    YES
1   RAINY          COLD       MILD     NO
2  SUNNY      MODERATE     NOMAL    YES
3  SUNNY           COLD      HIGH     YES


 THE ATTRIBUTES ARE:
 [['SUNNY ' 'WARM ' 'MILD']
 ['RAINY' 'COLD' 'MILD']
 ['SUNNY ' 'MODERATE' 'NOMAL']
 ['SUNNY ' 'COLD' 'HIGH ']]

 THE TARGET IS:      ['YES' 'NO' 'YES' 'YES']

 THE FINAL HYPOTHESIS IS: ['SUNNY ' '?' '?']
```

WEEK 2:

II) CANDIDATE ELIMINATION ALGORITHM:

```
# THIS PYTHON 3 ENVIRONMENT COMES WITH MANY HELPFUL ANALYTICS LIBRARIES INS TALLED
# IT IS DEFINED BY THE KAGGLE/PYTHON DOCKER IMAGE: HTTPS://GITHUB.COM/KAGGL
E/DOCKER-PYTHON
# FOR EXAMPLE, HERE'S SEVERAL HELPFUL PACKAGES TO LOAD

IMPORT NUMPY AS NP # LINEAR ALGEBRA
IMPORT PANDAS AS PD # DATA PROCESSING, CSV FILE I/O (E.G. PD.READ_CSV)

# INPUT DATA FILES ARE AVAILABLE IN THE READ-ONLY "../INPUT/" DIRECTORY
# FOR EXAMPLE, RUNNING THIS (BY CLICKING RUN OR PRESSING SHIFT+ENTER) WILL LIST ALL
FILES UNDER THE INPUT DIRECTORY

IMPORT OS
```

```
FOR DIRNAME, _, FILENAMES IN OS.WALK('/KAGGLE/INPUT'):
    FOR FILENAME IN FILENAMES:
        PRINT(OS.PATH.JOIN(DIRNAME, FILENAME))

# YOU CAN WRITE UP TO 20GB TO THE CURRENT DIRECTORY (/KAGGLE/WORKING/) THAT GETS
PRESERVED AS OUTPUT WHEN YOU CREATE A VERSION USING "SAVE & RUN ALL"  # YOU CAN ALSO
WRITE TEMPORARY FILES TO /KAGGLE/TEMP/, BUT THEY WON'T BE SA VED OUTSIDE OF THE
CURRENT SESSION
/KAGGLE/INPUT/CANDIDATECSV/CANDIDATE.CSV

DATA = PD.READ_CSV("/KAGGLE/INPUT/CANDIDATECSV/CANDIDATE.CSV")
PRINT("ENTERED DATA IS")
PRINT(DATA)
CONCEPTS = NP.ARRAY(DATA)[:,:-1]
PRINT("\N THE ATTRIBUTES ARE: \N",D)
TARGET = NP.ARRAY(DATA)[:,-1] PRINT("\N
THE TARGET IS: ",TARGET)
```

```
ENTERED DATA IS
      SKY  AIRTEMP   HUMIDITY      WIND   WATE  FORECAST  ENJOYSPORT
                                             R
0   SUNNY     WARM    NORMAL    STRONG   WARM      SAME         YES
1   SUNNY     WARM      HIGH   STRONG    WARM      SAME         YES
2   RAINY     COLD      HIGH    STRONG   WARM    CHANGE          NO
3   SUNNY     WARM      HIGH    STRONG   COOL    CHANGE         YES

 THE ATTRIBUTES ARE:
 [['SUNNY' 'WARM' 'NORMAL' 'STRONG' 'WARM' 'SAME']
 ['SUNNY' 'WARM' 'HIGH' 'STRONG ' 'WARM' 'SAME']
 ['RAINY' 'COLD' 'HIGH' 'STRONG' 'WARM' 'CHANGE']
 ['SUNNY' 'WARM' 'HIGH' 'STRONG' 'COOL' 'CHANGE']]

 THE TARGET IS:     ['YES' 'YES' 'NO' 'YES']
```

```
#training function to implement candidate_elimination algorithm
def learn(concepts, target):
 specific_h = concepts[0].copy()
 print("\n Initialization of specific_h and general_h")
 print(specific_h)
 general_h = [["?" for i in range(len(specific_h))] for i in
range(len(specific_h))]
 print(general_h)
 for i, h in enumerate(concepts):
     if target[i] == "yes":
         for x in range(len(specific_h)):
             if h[x]!= specific_h[x]:
                 specific_h[x] ='?'
                 general_h[x][x] ='?'
             print(specific_h)
     print(specific_h)
     if target[i] == "no":
         for x in range(len(specific_h)):
             if h[x]!= specific_h[x]:
                 general_h[x][x] = specific_h[x]
             else:
                 general_h[x][x] = '?'
     print("\n Steps of Candidate Elimination Algorithm",i+1)
     print(specific_h)
     print(general_h)
 indices = [i for i, val in enumerate(general_h) if val ==
['?', '?', '?', '?', '?', '?']]
 for i in indices:
     general_h.remove(['?', '?', '?', '?', '?', '?'])
 return specific_h, general_h
s_final, g_final = learn(concepts, target)

#obtaining the final hypothesis
print("\nFinal Specific_h:", s_final, sep="\n")
print("\nFinal General_h:", g_final, sep="\n")
```

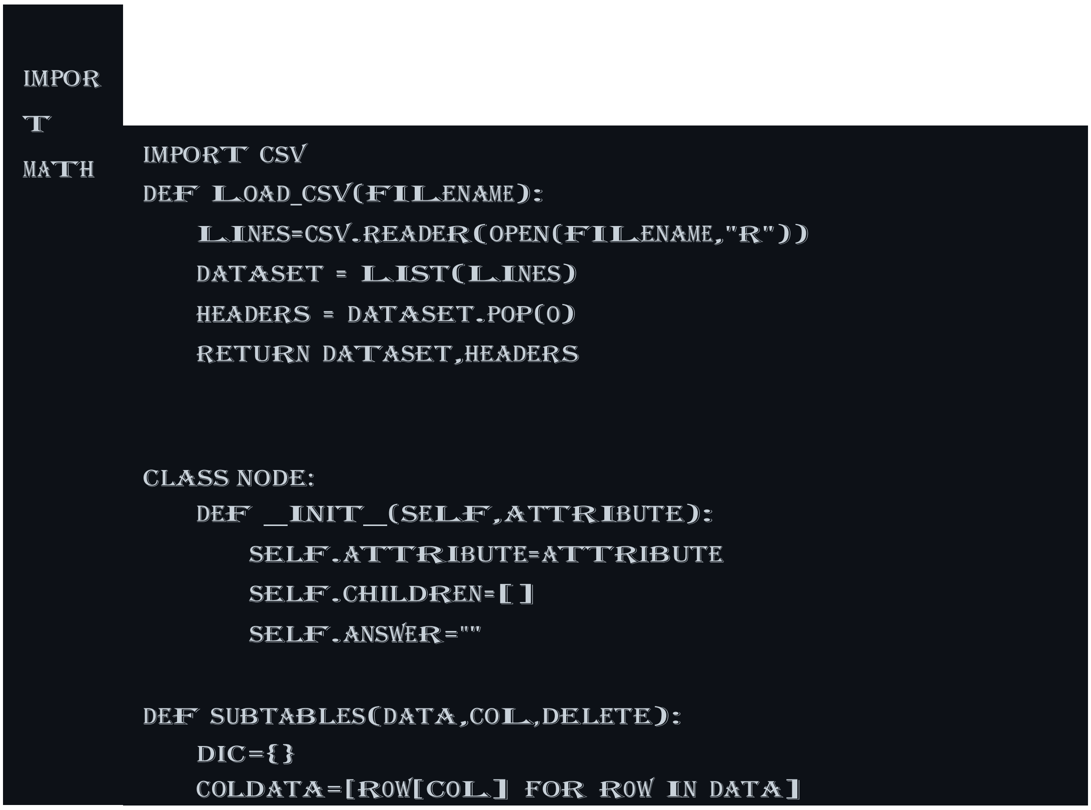
OUTPUT:

```
INITIALIZATION OF SPECIFIC_H AND GENERAL_H ['SUNNY'
'WARM' 'NORMAL' 'STRONG' 'WARM' 'SAME']
[['?', '?', '?', '?', '?', '?'], ['?', '?', '?', '?', '?', '?'], ['?',
'?', '?', '?', '?', '?'], ['?', '?', '?', '?', '?', '?'], ['?', '?', '?
', '?', '?', '?'], ['?', '?', '?', '?', '?', '?']]
['SUNNY' 'WARM' 'NORMAL' 'STRONG' 'WARM' 'SAME']
['SUNNY' 'WARM' 'NORMAL' 'STRONG' 'WARM' 'SAME']
['SUNNY' 'WARM' 'NORMAL' 'STRONG' 'WARM' 'SAME']
['SUNNY' 'WARM' 'NORMAL' 'STRONG' 'WARM' 'SAME']
['SUNNY' 'WARM' 'NORMAL' 'STRONG' 'WARM' 'SAME']
['SUNNY' 'WARM' 'NORMAL' 'STRONG' 'WARM' 'SAME']
['SUNNY' 'WARM' 'NORMAL' 'STRONG' 'WARM' 'SAME']

 STEPS OF CANDIDATE ELIMINATION ALGORITHM 1 ['SUNNY'
'WARM' 'NORMAL' 'STRONG' 'WARM' 'SAME']
[['?', '?', '?', '?', '?', '?'], ['?', '?', '?', '?', '?', '?'], ['?',
'?', '?', '?', '?', '?'], ['?', '?', '?', '?', '?', '?'], ['?', '?', '?
', '?', '?', '?'], ['?', '?', '?', '?', '?', '?']]
['SUNNY' 'WARM' 'NORMAL' 'STRONG' 'WARM' 'SAME']
['SUNNY' 'WARM' 'NORMAL' 'STRONG' 'WARM' 'SAME']
['SUNNY' 'WARM' '?' 'STRONG' 'WARM' 'SAME']
['SUNNY' 'WARM' '?' '?' 'WARM' 'SAME']
['SUNNY' 'WARM' '?' '?' 'WARM' 'SAME']
['SUNNY' 'WARM' '?' '?' 'WARM' 'SAME']
['SUNNY' 'WARM' '?' '?' 'WARM' 'SAME']



STEPS OF CANDIDATE ELIMINATION ALGORITHM 2
['SUNNY' 'WARM' '?' '?' 'WARM' 'SAME']
[['?', '?', '?', '?', '?', '?'], ['?', '?', '?', '?', '?', '?'], ['?',
'?', '?', '?', '?', '?'], ['?', '?', '?', '?', '?', '?'], ['?', '?', '?
', '?', '?', '?'], ['?', '?', '?', '?', '?', '?']]
['SUNNY' 'WARM' '?' '?' 'WARM' 'SAME']

 STEPS OF CANDIDATE ELIMINATION ALGORITHM 3
['SUNNY' 'WARM' '?' '?' 'WARM' 'SAME']
```

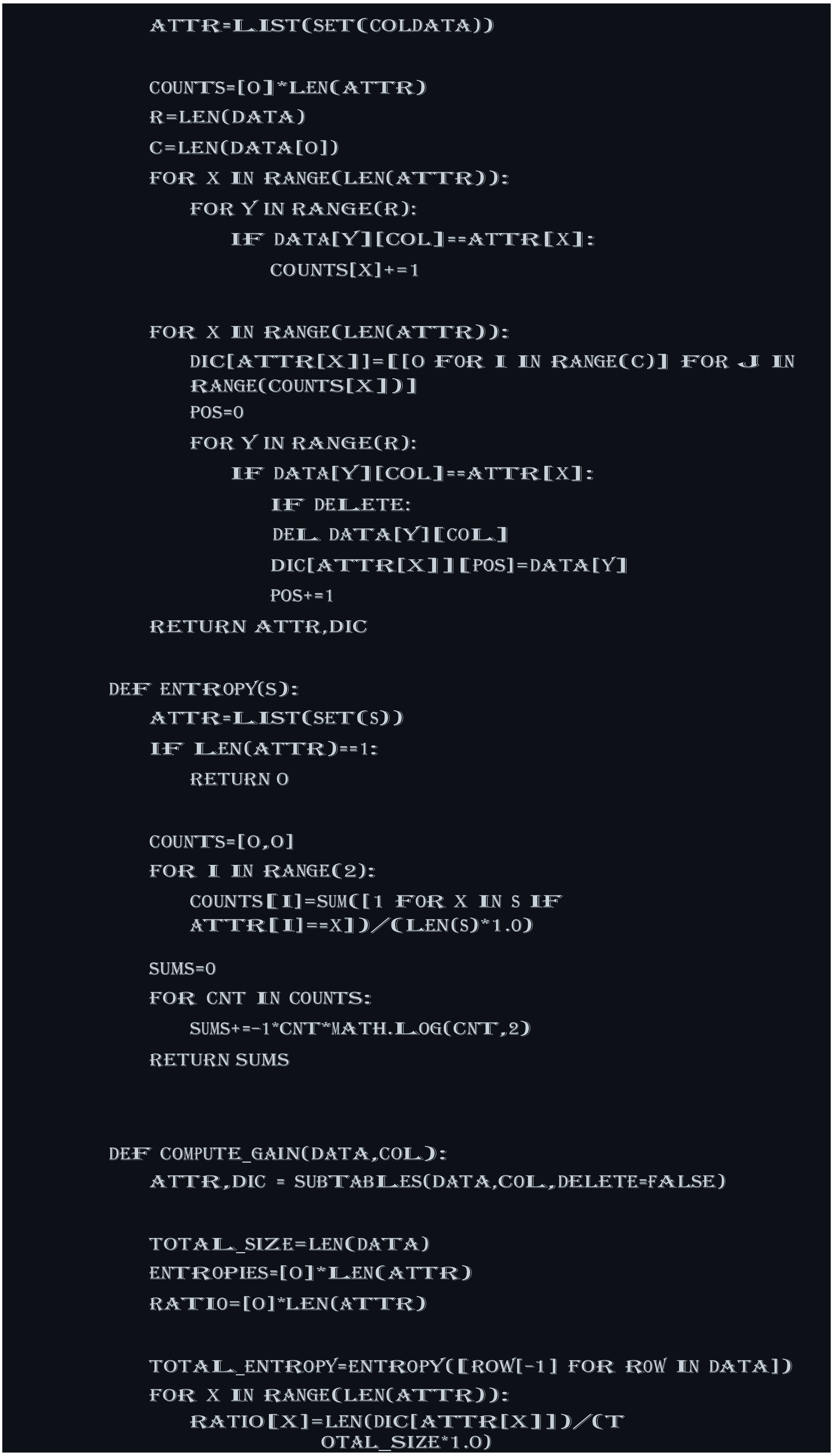
```
[['SUNNY', '?', '?', '?', '?', '?'], ['?', 'WARM', '?', '?', '?', '?'],
['?', '?', '?', '?', '?', '?'], ['?', '?', '?', '?', '?', '?'], ['?', '
?', '?', '?', '?', '?'], ['?', '?', '?', '?', '?', 'SAME']]
['SUNNY' 'WARM' '?' '?' 'WARM' 'SAME']
['SUNNY' 'WARM' '?' '?' 'WARM' 'SAME']
['SUNNY' 'WARM' '?' '?' 'WARM' 'SAME']
['SUNNY' 'WARM' '?' '?' 'WARM' 'SAME']
['SUNNY' 'WARM' '?' '?' '?' 'SAME']
['SUNNY' 'WARM' '?' '?' '?' '?']
['SUNNY' 'WARM' '?' '?' '?' '?']

  STEPS OF CANDIDATE ELIMINATION ALGORITHM 4
['SUNNY' 'WARM' '?' '?' '?' '?']
[['SUNNY', '?', '?', '?', '?', '?'], ['?', 'WARM', '?', '?', '?', '?'],
['?', '?', '?', '?', '?', '?'], ['?', '?', '?', '?', '?', '?'], ['?', '
?', '?', '?', '?', '?'], ['?', '?', '?', '?', '?', '?']]

FINAL SPECIFIC_H:
['SUNNY' 'WARM' '?' '?' '?' '?']

FINAL GENERAL_H:
[['SUNNY', '?', '?', '?', '?', '?'], ['?', 'WARM', '?', '?', '?', '?']]
```

WEEK 3:

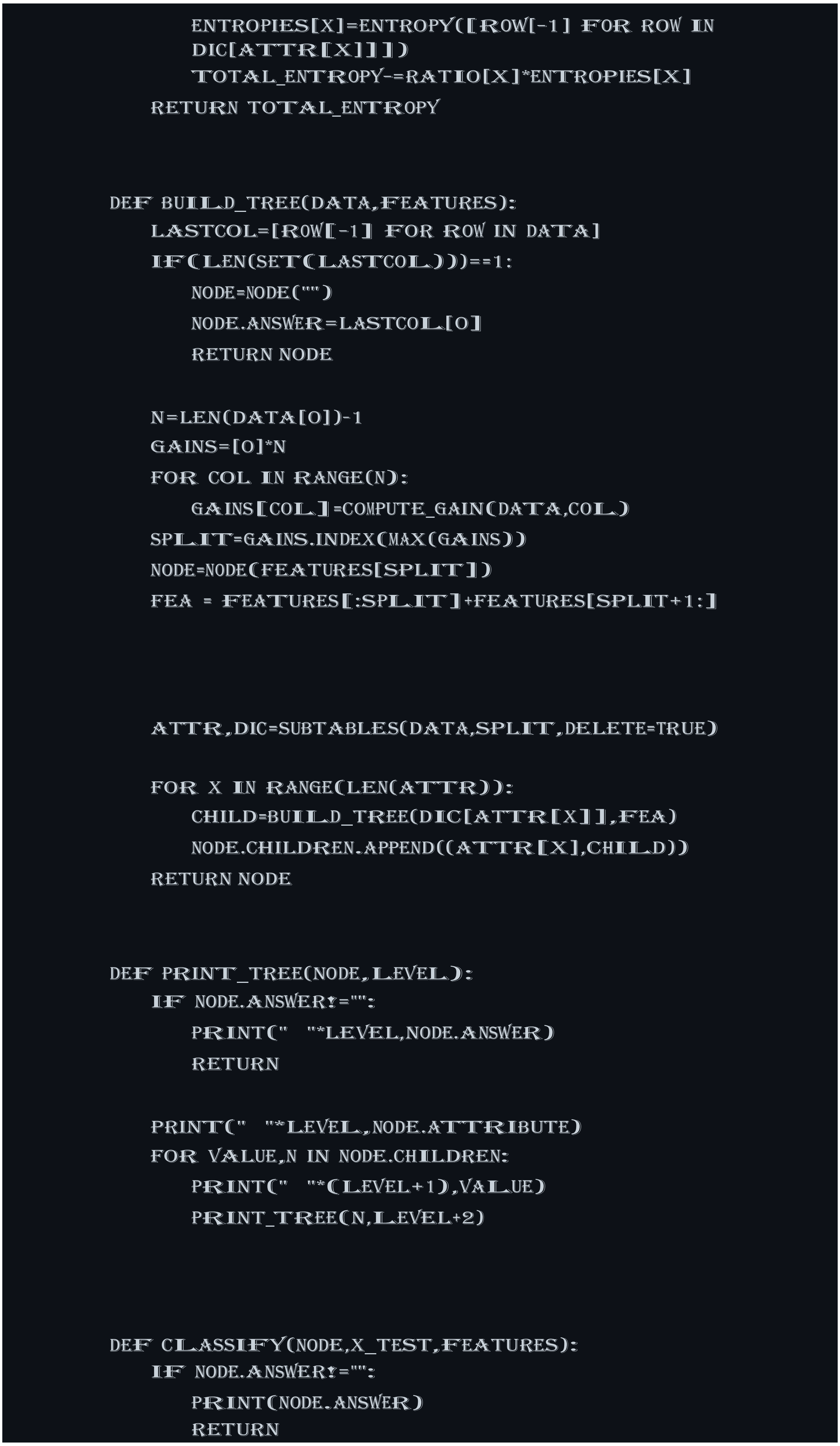
ID3 ALGORITHM:



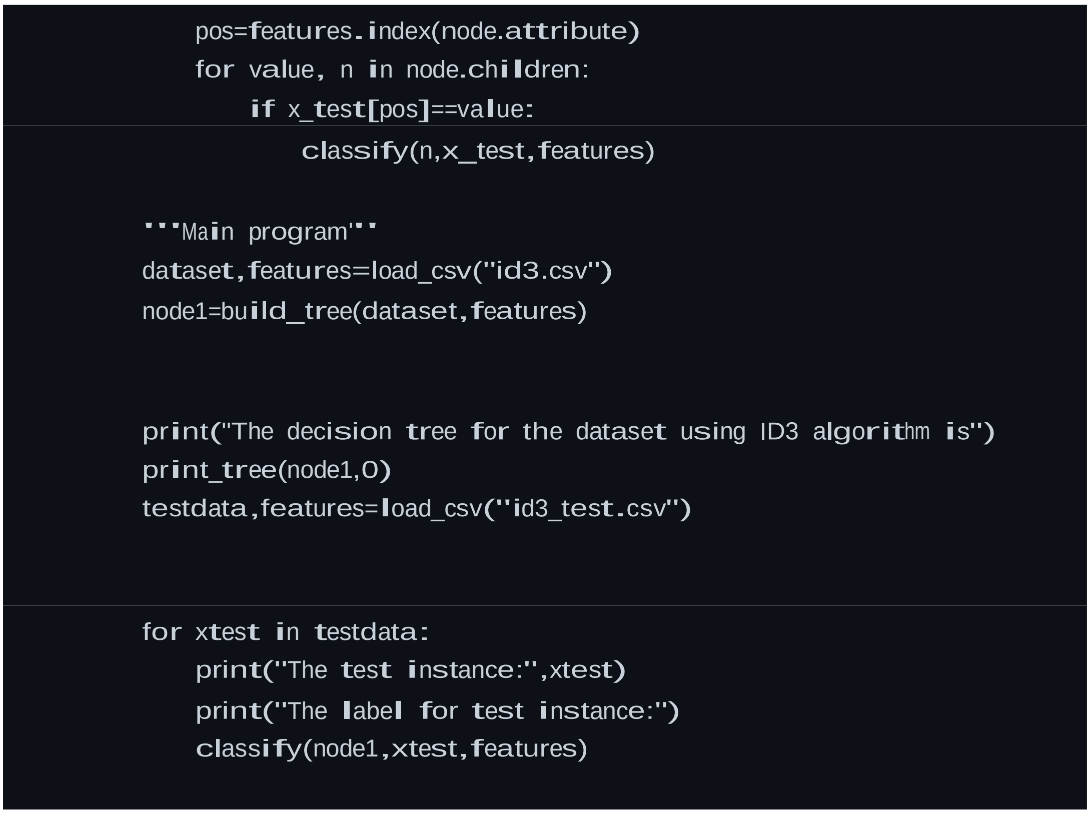
```
IMPOR
T
MATH
IMPORT CSV
DEF LOAD_CSV(FILENAME):
    LINES=CSV.READER(OPEN(FILENAME,"R"))
    DATASET = LIST(LINES)
    HEADERS = DATASET.POP(0)
    RETURN DATASET,HEADERS



CLASS NODE:
    DEF __INIT__(SELF,ATTRIBUTE):
        SELF.ATTRIBUTE=ATTRIBUTE
        SELF.CHILDREN=[]
        SELF.ANSWER=""

DEF SUBTABLES(DATA,COL,DELETE):
    DIC={}
    COLDATA=[ROW[COL] FOR ROW IN DATA]
```



```
    ATTR=LIST(SET(COLDATA))

    COUNTS=[0]*LEN(ATTR)
    R=LEN(DATA)
    C=LEN(DATA[0])
    FOR X IN RANGE(LEN(ATTR)):
        FOR Y IN RANGE(R):
            IF DATA[Y][COL]==ATTR[X]:
                COUNTS[X]+=1

    FOR X IN RANGE(LEN(ATTR)):
        DIC[ATTR[X]]=[[0 FOR I IN RANGE(C)] FOR J IN
        RANGE(COUNTS[X])]
        POS=0
        FOR Y IN RANGE(R):
            IF DATA[Y][COL]==ATTR[X]:
                IF DELETE:
                DEL DATA[Y][COL]
                DIC[ATTR[X]][POS]=DATA[Y]
                POS+=1
    RETURN ATTR,DIC

DEF ENTROPY(S):
    ATTR=LIST(SET(S))
    IF LEN(ATTR)==1:
        RETURN 0

    COUNTS=[0,0]
    FOR I IN RANGE(2):
        COUNTS[I]=SUM([1 FOR X IN S IF
        ATTR[I]==X])/(LEN(S)*1.0)

    SUMS=0
    FOR CNT IN COUNTS:
        SUMS+=-1*CNT*MATH.LOG(CNT,2)
    RETURN SUMS



DEF COMPUTE_GAIN(DATA,COL):
    ATTR,DIC = SUBTABLES(DATA,COL,DELETE=FALSE)

    TOTAL_SIZE=LEN(DATA)
    ENTROPIES=[0]*LEN(ATTR)
    RATIO=[0]*LEN(ATTR)

    TOTAL_ENTROPY=ENTROPY([ROW[-1] FOR ROW IN DATA])
    FOR X IN RANGE(LEN(ATTR)):
        RATIO[X]=LEN(DIC[ATTR[X]])/(T
                     OTAL_SIZE*1.0)
```

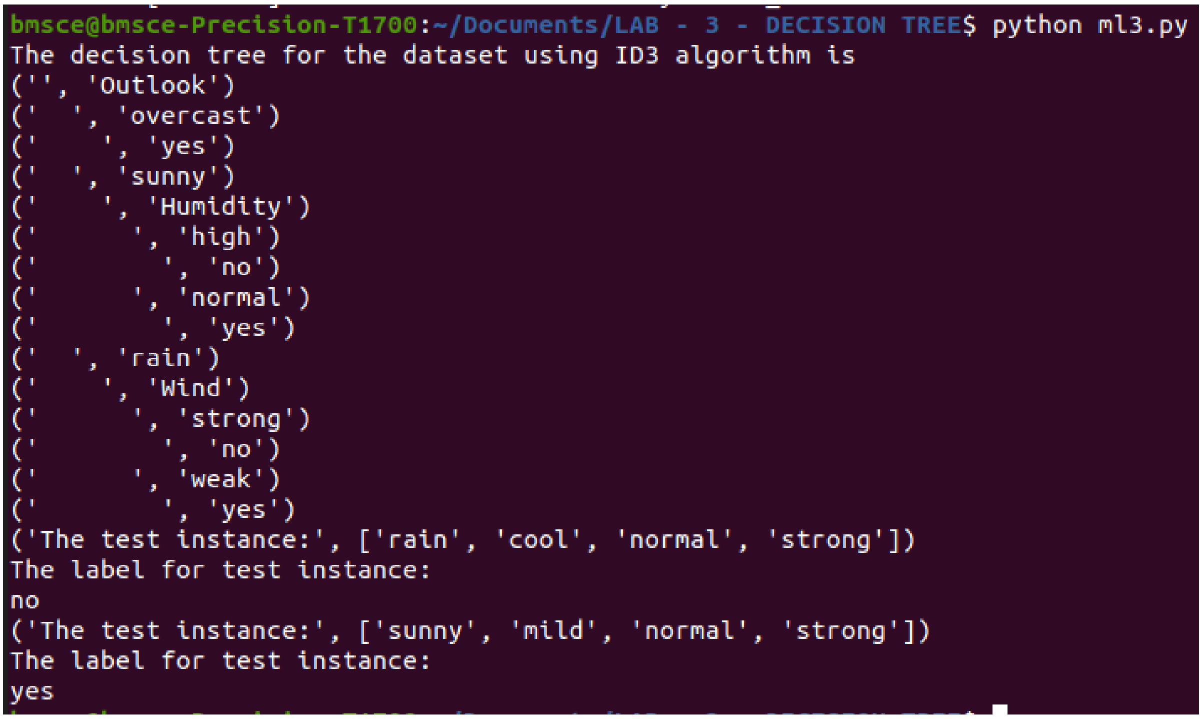


```
        ENTROPIES[X]=ENTROPY([ROW[-1] FOR ROW IN
        DIC[ATTR[X]]])
        TOTAL_ENTROPY-=RATIO[X]*ENTROPIES[X]
    RETURN TOTAL_ENTROPY


DEF BUILD_TREE(DATA,FEATURES):
    LASTCOL=[ROW[-1] FOR ROW IN DATA]
    IF(LEN(SET(LASTCOL)))==1:
        NODE=NODE("")
        NODE.ANSWER=LASTCOL[0]
        RETURN NODE

    N=LEN(DATA[0])-1
    GAINS=[0]*N
    FOR COL IN RANGE(N):
        GAINS[COL]=COMPUTE_GAIN(DATA,COL)
    SPLIT=GAINS.INDEX(MAX(GAINS))
    NODE=NODE(FEATURES[SPLIT])
    FEA = FEATURES[:SPLIT]+FEATURES[SPLIT+1:]



    ATTR,DIC=SUBTABLES(DATA,SPLIT,DELETE=TRUE)

    FOR X IN RANGE(LEN(ATTR)):
        CHILD=BUILD_TREE(DIC[ATTR[X]],FEA)
        NODE.CHILDREN.APPEND((ATTR[X],CHILD))
    RETURN NODE


DEF PRINT_TREE(NODE,LEVEL):
    IF NODE.ANSWER!="":
        PRINT("  "*LEVEL,NODE.ANSWER)
        RETURN

    PRINT("  "*LEVEL,NODE.ATTRIBUTE)
    FOR VALUE,N IN NODE.CHILDREN:
        PRINT("  "*(LEVEL+1),VALUE)
        PRINT_TREE(N,LEVEL+2)



DEF CLASSIFY(NODE,X_TEST,FEATURES):
    IF NODE.ANSWER!="":
        PRINT(NODE.ANSWER)
        RETURN
```



```
    pos=features.index(node.attribute)
    for value, n in node.children:
        if x_test[pos]==value:
            classify(n,x_test,features)

'''Main program'''
dataset,features=load_csv("id3.csv")
node1=build_tree(dataset,features)


print("The decision tree for the dataset using ID3 algorithm is")
print_tree(node1,0)
testdata,features=load_csv("id3_test.csv")


for xtest in testdata:
    print("The test instance:",xtest)
    print("The label for test instance:")
    classify(node1,xtest,features)
```

OUTPUT:



```
bmsce@bmsce-Precision-T1700:~/Documents/LAB - 3 - DECISION TREE$ python ml3.py
The decision tree for the dataset using ID3 algorithm is
('', 'Outlook')
('  ', 'overcast')
('    ', 'yes')
('  ', 'sunny')
('    ', 'Humidity')
('      ', 'high')
('        ', 'no')
('      ', 'normal')
('        ', 'yes')
('  ', 'rain')
('    ', 'Wind')
('      ', 'strong')
('        ', 'no')
('      ', 'weak')
('        ', 'yes')
('The test instance:', ['rain', 'cool', 'normal', 'strong'])
The label for test instance:
no
('The test instance:', ['sunny', 'mild', 'normal', 'strong'])
The label for test instance:
yes
```

WEEK 4:

IV) NAÏVE BAYES CLASSIFIER:

```
import pandas as pd

data = pd.read_csv('PlayTennis.csv')
data.head()
```

```
Y = LIST(DATA['PLAYTENNIS'].VALUES)
```

```
X = DATA.ILOC[:,1:].VALUES
```

```
PRINT(F'TARGET VALUES: {Y}')
```

```
PRINT(F'FEATURES: \N{X}')
```

```
TARGET VALUES: ['NO', 'NO', 'YES', 'YES', 'YES', 'NO', 'YES', 'NO', 'YES', 'YES', 'YES', 'YES', 'YES', 'NO']
FEATURES:
[['SUNNY' 'HOT' 'HIGH' 'WEAK']
 ['SUNNY' 'HOT' 'HIGH' 'STRONG']
 ['OVERCAST' 'HOT' 'HIGH' 'WEAK']
 ['RAIN' 'MILD' 'HIGH' 'WEAK']
 ['RAIN' 'COOL' 'NORMAL' 'WEAK']
 ['RAIN' 'COOL' 'NORMAL' 'STRONG'] ['OVERCAST'
 'COOL' 'NORMAL' 'STRONG'] ['SUNNY' 'MILD' 'HIGH'
 'WEAK']
 ['SUNNY' 'COOL' 'NORMAL' 'WEAK']
 ['RAIN' 'MILD' 'NORMAL' 'WEAK']
 ['SUNNY' 'MILD' 'NORMAL' 'STRONG']
 ['OVERCAST' 'MILD' 'HIGH' 'STRONG']
 ['OVERCAST' 'HOT' 'NORMAL' 'WEAK']
 ['RAIN' 'MILD' 'HIGH' 'STRONG']]
```

```
y_train = y[:8]
y_val = y[8:]

X_train = X[:8]
X_val = X[8:]

print(f"Number of instances in training set: {len(X_train)}")
print(f"Number of instances in testing set: {len(X_val)}")
```

```
NUMBER OF INSTANCES IN TRAINING SET: 8
NUMBER OF INSTANCES IN TESTING SET: 6
```

```
class NaiveBayesClassifier:


    def __init__(self, X, y):

        self.X, self.y = X, y

        self.N = len(self.X)
        SELF.DIM = LEN(SELF.X[0])

        SELF.ATTRS = [[] FOR _ IN RANGE(SELF.DIM)]

        SELF.OUTPUT_DOM = {}

        SELF.DATA = []

        FOR I IN RANGE(LEN(SELF.X)):
            FOR J IN RANGE(SELF.DIM):
                IF NOT SELF.X[I][J] IN SELF.ATTRS[J]:
                    SELF.ATTRS[J].APPEND(SELF.X[I][J])

            IF NOT SELF.Y[I] IN SELF.OUTPUT_DOM.KEYS():
                SELF.OUTPUT_DOM[SELF.Y[I]] = 1

            ELSE:
                SELF.OUTPUT_DOM[SELF.Y[I]] += 1

            SELF.DATA.APPEND([SELF.X[I], SELF.Y[I]])
        DEF CLASSIFY(SELF, ENTRY):

        SOLVE = NONE
        MAX_ARG = -1

        FOR Y IN SELF.OUTPUT_DOM.KEYS():

            PROB = SELF.OUTPUT_DOM[Y]/SELF.N

            FOR I IN RANGE(SELF.DIM):
                CASES = [X FOR X IN SELF.DATA IF X[0][I] == ENTRY[I] AN
D X[1] == Y]
                N = LEN(CASES)
                PROB *= N/SELF.N

            IF  PROB  >  MAX_ARG:
                MAX_ARG = PROB
                SOLVE = Y

        RETURN SOLVE

NBC = NAIVEBAYESCLASSIFIER(X_TRAIN, Y_TRAIN)

TOTAL_CASES = LEN(Y_VAL)

GOOD = 0
BAD = 0
PREDICTIONS = []

FOR I IN  RANGE(TOTAL_CASES): PREDICT =
    NBC.CLASSIFY(X_VAL[I])
    PREDICTIONS.APPEND(PREDICT)

    IF Y_VAL[I] == PREDICT:
        GOOD += 1
```

```
ELSE:
    BAD += 1
```

```
print('Predicted values:', predictions)
print('Actual values:', y_val)
print()
print('Total number of testing instances in the dataset:', total_cases)
print('Number of correct predictions:', good)
print('Number of wrong predictions:', bad)
print()
print('Accuracy of Bayes Classifier:', good/total_cases)
```

OUTPUT:

```
PREDICTED VALUES: ['NO', 'YES', 'NO', 'YES', 'YES', 'NO']
ACTUAL VALUES: ['YES', 'YES', 'YES', 'YES', 'YES', 'NO']

TOTAL NUMBER OF TESTING INSTANCES IN THE DATASET: 6
NUMBER OF CORRECT PREDICTIONS: 4
NUMBER OF WRONG PREDICTIONS: 2

ACCURACY OF BAYES CLASSIFIER: 0.6666666666666666
```

WEEK 5:

V) BAYESIAN NETWORK:

# THIS PYTHON 3 ENVIRONMENT COMES WITH MANY HELPFUL ANALYTICS LIBRARIES INSTALLED # IT IS DEFINED BY THE KAGGLE/PYTHON DOCKER IMAGE: HTTPS://GITHUB.COM/KAGGLE/DOCKER-PYTHON # FOR EXAMPLE, HERE'S SEVERAL HELPFUL PACKAGES TO LOAD **IMPORT NUMPY AS NP**

# LINEAR ALGEBRA

IMPORT PANDAS AS PD

IMPORT PGMPY AS PGMPY FROM PGMPY.ESTIMATORS

IMPORT MAXIMUMLIKELIHOODESTIMATOR **FROM PGMPY.MODELS**

IMPORT BAYESIANMODEL **FROM PGMPY.INFERENCE**

IMPORT VARIABLEELIMINATION

IMPORT OS FOR DIRNAME, _, FILENAMES **IN** OS.WALK('/KAGGLE/INPUT'):

FOR FILENAME **IN** FILENAMES: PRINT(OS.PATH.JOIN(DIRNAME, FILENAME)) *# YOU CAN WRITE UP TO 20GB TO THE CURRENT DIRECTORY (/KAGGLE/WORKING/) THAT GETS PRESERVED AS OUTPUT WHEN YOU CREATE A VERSION USING "SAVE & RUN ALL" # YOU CAN ALSO WRITE TEMPORARY FILES TO /KAGGLE/TEMP/, BUT THEY WON'T BE SAVED OUTSIDE OF THE CURRENT SESSION*

```
#read Cleveland Heart Disease data
heartDisease = pd.read_csv("/kaggle/input/bayesiannetwork/heart.csv")
heartDisease = heartDisease.replace('?',np.nan)
#display the data
print('Sample instances from the dataset are given below')
print(heartDisease.head())
```

```
#display the Attributes names and datatyes
print('\n Attributes and datatypes')
print(heartDisease.dtypes)
#Creat Model- Bayesian Network
model = BayesianModel([('age','heartdisease'),('sex','heartdisease'),('
exang','heartdisease'),('cp','heartdisease'),('heartdisease','restecg')
,('heartdisease','chol')])
#Learning CPDs using Maximum Likelihood Estimators
print('\n Learning CPD using Maximum likelihood estimators')
model.fit(heartDisease,estimator=MaximumLikelihoodEstimator)
# Inferencing with Bayesian Network
print('\n Inferencing with Bayesian Network:')
HeartDiseasetest_infer = VariableElimination(model)
#computing the Probability of HeartDisease given restecg
print('\n 1.Probability of HeartDisease given evidence= restecg :1')
q1=HeartDiseasetest_infer.query(variables=['heartdisease'],evidence={'r
estecg':1})
print(q1)
#computing the Probability of HeartDisease given cp
print('\n 2.Probability of HeartDisease given evidence= cp:2 ')
q2=HeartDiseasetest_infer.query(variables=['heartdisease'],evidence={'c
p':2})
print(q2)
```

OUTPUT:

```
SAMPLE INSTANCES FROM THE DATASET ARE GIVEN BELOW
    AGE   SEX   CP   TRESTBPS   CHOL   FBS   RESTECG   THALACH   EXANG
        OLDPEAK SLOPE              \
0   63     1    1        145    233     1         2       150       0      2.3
3
1   67     1    4        160    286     0         2       108       1      1.5
2
2   67     1    4        120    229     0         2       129       1      2.6
2
3   37     1    3        130    250     0         0       187       0      3.5
3
4   41     0    2        130    204     0         2       172       0      1.4
1

   CA THAL   HEARTDISEASE
0   0     6             0
1   3     3             2
2   2     7             1
3   0     3             0
4   0     3             0

 ATTRIBUTES AND DATATYPES
AGE                  INT64
SEX                  INT64
CP                   INT64
TRESTBPS             INT64
CHOL                 INT64
FBS                  INT64
RESTECG              INT64
THALACH              INT64
```

```
EXANG                   INT64
OLDPEAK               FLOAT64
SLOPE                   INT64
CA                     OBJECT
THAL                   OBJECT
HEARTDISEASE            INT64
DTYPE: OBJECT

 LEARNING CPD USING MAXIMUM LIKELIHOOD ESTIMATORS
FINDING ELIMINATION ORDER: :          0%|                | 0/5 [00:00<?, ?IT/S] 0%|  |
  0/5 [00:00<?, ?IT/S]
ELIMINATING: AGE:        0%|               | 0/5 [00:00<?, ?IT/S] ELIMINATING: CHOL:
                          0%|               | 0/5 [00:00<?, ?IT/S] ELIMINATING: CP:
                         0%|               | 0/5 [00:00<?, ?IT/S] ELIMINATING: SEX:
                         0%|               | 0/5 [00:00<?, ?IT/S] ELIMINATING:
EXANG: 100%|██████████| 5/5 [00:00<00:00, 189.65IT/S]
FINDING ELIMINATION ORDER: : 100%|██████████| 5/5 [00:00<00:00, 132.81IT/S] FINDING
ELIMINATION ORDER: :          0%|                | 0/5 [00:00<?, ?IT/S]
  0%|              | 0/5 [00:00<?, ?IT/S]
ELIMINATING: AGE:        0%|               | 0/5 [00:00<?, ?IT/S] ELIMINATING: CHOL:
                          0%|               | 0/5 [00:00<?, ?IT/S] ELIMINATING:
RESTECG:                 0%|               | 0/5 [00:00<?, ?IT/S] ELIMINATING: SEX:
                         0%|               | 0/5 [00:00<?, ?IT/S] ELIMINATING:
EXANG: 100%|██████████| 5/5 [00:00<00:00, 230.00IT/S]
 INFERENCING WITH BAYESIAN NETWORK:
1. PROBABILITY OF HEARTDISEASE GIVEN EVIDENCE= RESTECG :1
+-----------------+---------------------+
| HEARTDISEASE    |   PHI(HEARTDISEASE) |
+=================+=====================+
| HEARTDISEASE(0) |              0.1012 |
+-----------------+---------------------+
| HEARTDISEASE(1) |              0.0000 |
+-----------------+---------------------+
| HEARTDISEASE(2) |              0.2392 |
+-----------------+---------------------+
| HEARTDISEASE(3) |              0.2015 |
+-----------------+---------------------+
| HEARTDISEASE(4) |              0.4581 |
+-----------------+---------------------+

 2. PROBABILITY OF HEARTDISEASE GIVEN EVIDENCE= CP:2
+-----------------+---------------------+
| HEARTDISEASE    |   PHI(HEARTDISEASE) |
+=================+=====================+
| HEARTDISEASE(0) |              0.3610 |
+-----------------+---------------------+
| HEARTDISEASE(1) |              0.2159 |
+-----------------+---------------------+
| HEARTDISEASE(2) |              0.1373 |
+-----------------+---------------------+
| HEARTDISEASE(3) |              0.1537 |
+                 +                     +
| HEARTDISEASE(4) |              0.1321
```

```
+-------------------------+------------------------------+
```

WEEK 6:

VI) INFERRING FROM BAYESIAN MODEL:

FROM PGMPY.MODELS IMPORT BAYESIANMODEL

FROM PGMPY.FACTORS.DISCRETE IMPORT TABULARCPD

```
cancer_model = BayesianModel([('Pollution','Cancer'),
                             ('Smoker','Cancer'),
                             ('Cancer','Xray'),
                             ('Cancer','Dyspnoea')])
print('Bayesian network nodes are:')
print("\t",cancer_model.nodes())
print('Bayesian network edges are:')
print("\t",cancer_model.edges())

cpd_poll = TabularCPD(variable='Pollution',variable_card=2,values=[[0.9
],[0.1]])
cpd_smoke = TabularCPD(variable='Smoker',variable_card=2,values=[[0.3],
[0.7]])
cpd_cancer = TabularCPD(variable='Cancer',variable_card=2,values=[[0.03
,0.05,0.001,0.02],
                                                                  [0.97,
0.95,0.999,0.98]],
                       evidence=['Smoker','Pollution'],
                       evidence_card=[2,2])
cpd_xray = TabularCPD(variable='Xray',variable_card=2,values=[[0.9,0.2]
,[0.1,0.8]],
                     evidence=['Cancer'],evidence_card=[2])
cpd_dysp = TabularCPD(variable='Dyspnoea',variable_card=2,values=[[0.65
,0.3],[0.35,0.7]],
                     evidence=['Cancer'],evidence card=[2])
```

FROM PGMPY.INFERENCE IMPORT VARIABLEELIMINATION

```
BAYESIAN NETWORK NODES ARE:
        ['POLLUTION', 'CANCER', 'SMOKER', 'XRAY', 'DYSPNOEA'] BAYESIAN
NETWORK EDGES ARE:
        [('POLLUTION', 'CANCER'), ('CANCER', 'XRAY'), ('CANCER', 'DYSPN OEA'), ('SMOKER',
'CANCER')]
```

```
cancer_model.add_cpds(cpd_poll,cpd_smoke,cpd_cancer,cpd_xray,cpd_dysp)
print('Model generated by adding cpts(cpds)')
print('Checking correctness of model:',end='')
print(cancer_model.check_model())
```

```
MODEL GENERATED BY ADDING CPTS(CPDS)
CHECKING CORRECTNESS OF MODEL:TRUE
```

```
PRINT('ALL LOCAL DEPENCIES ARE AS FOLLOWS')
```

```
CANCER_MODEL.GET_INDEPENDENCIES()
```

ALL LOCAL DEPENCIES ARE AS FOLLOWS

OUT[10]:

```
(POLLUTION ⊥ SMOKER)
(POLLUTION ⊥ DYSPNOEA, XRAY | CANCER) (POLLUTION ⊥
XRAY | DYSPNOEA, CANCER) (POLLUTION ⊥ DYSPNOEA |
CANCER, XRAY) (POLLUTION ⊥ DYSPNOEA, XRAY |
CANCER, SMOKER) (POLLUTION ⊥ XRAY | DYSPNOEA,
CANCER, SMOKER) (POLLUTION ⊥ DYSPNOEA | CANCER,
XRAY, SMOKER) (SMOKER ⊥ POLLUTION)
(SMOKER ⊥ DYSPNOEA, XRAY | CANCER)
(SMOKER ⊥  XRAY | DYSPNOEA,CANCER)
(SMOKER ⊥ DYSPNOEA, XRAY | POLLUTION, CANCER)
(SMOKER ⊥ DYSPNOEA | CANCER, XRAY)
(SMOKER ⊥ XRAY | DYSPNOEA, POLLUTION, CANCER)
(SMOKER ⊥ DYSPNOEA | POLLUTION, CANCER, XRAY)
(XRAY ⊥ DYSPNOEA, POLLUTION, SMOKER | CANCER)
(XRAY ⊥ POLLUTION, SMOKER | DYSPNOEA, CANCER)
(XRAY ⊥ DYSPNOEA, SMOKER | POLLUTION, CANCER)
(XRAY ⊥ DYSPNOEA, POLLUTION | CANCER, SMOKER)
(XRAY ⊥ SMOKER | DYSPNOEA, POLLUTION, CANCER)
(XRAY ⊥ POLLUTION | DYSPNOEA, CANCER, SMOKER)
(XRAY ⊥ DYSPNOEA | POLLUTION, CANCER, SMOKER)
(DYSPNOEA ⊥ POLLUTION, XRAY, SMOKER | CANCER)
(DYSPNOEA ⊥ XRAY, SMOKER | POLLUTION, CANCER)
(DYSPNOEA ⊥ POLLUTION, SMOKER | CANCER, XRAY)
(DYSPNOEA ⊥ POLLUTION, XRAY | CANCER, SMOKER)
(DYSPNOEA ⊥ SMOKER | POLLUTION, CANCER, XRAY)
(DYSPNOEA ⊥ XRAY | POLLUTION, CANCER, SMOKER)
(DYSPNOEA ⊥ POLLUTION | CANCER, XRAY, SMOKER)
```

```
print('Displaying CPDs')
print(cancer_model.get_cpds('Pollution'))
print(cancer_model.get_cpds('Smoker'))
print(cancer_model.get_cpds('Cancer'))
print(cancer_model.get_cpds('Xray'))
print(cancer_model.get_cpds('Dyspnoea'))
```

```
DISPLAYING    CPDS
+------------------------------+- ---------+
| POLLUTION(0) | 0.9 |
+------------------------------+- ---------+
| POLLUTION(1) | 0.1 |
+------------------------------+- ---------+
+-------------------------+---------.-+
| SMOKER(0) | 0.3 |
+-------------------------+---------.-+
| SMOKER(1) | 0.7 |
+-------------------------+---------.-+
+-------------------------+------------------------------+- ------------------- ---------+---------------------------------+- ------------------------------
-+
| SMOKER        | SMOKER(0)        | SMOKER(0)        | SMOKER(1)        | SMOKER(1)
|
+---------------+------------------+- ---------------- ---------+------------------+- ------------------
-+
| POLLUTION | POLLUTION(0) | POLLUTION(1) | POLLUTION(0) | POLLUTION(1)
|
+---------------+------------------+- ---------------- ---------+------------------+- ------------------
-+
| CANCER(0) | 0.03             | 0.05             | 0.001            | 0.02
|
+---------------+------------------+- ---------------- ---------+------------------+- ------------------
-+
| CANCER(1) | 0.97             | 0.95             | 0.999            | 0.98
|
+---------------+------------------+- ---------------- ---------+------------------+- ------------------
-+
+-----------+ ---------------+ ---------------+
| CANCER    | CANCER(0) | CANCER(1) |
+-----------+ ---------------+ ---------------+
| XRAY(0) | 0.9            | 0.2            |
+-----------+ ---------------+ ---------------+
| XRAY(1) | 0.1            | 0.8            |
+-----------+ ---------------+ ---------------+
+----------------+ ---------------+ ---------------+
| CANCER         | CANCER(0) | CANCER(1) |
+----------------+ ---------------+ ---------------+
| DYSPNOEA(0) | 0.65           | 0.3            |
+----------------+ ---------------+ ---------------+
| DYSPNOEA(1) | 0.35           | 0.7            |
+----------------+ ---------------+ ---------------+
```

```python
cancer_infer=VariableElimination(cancer_model)
print('\n Inferencing with bayesian network')
print("\n Probability of Cancer given smoker")
q=cancer_infer.query(variables=['Cancer'],evidence={'Smoker':1})
print(q)

print("\n Probability of Cancer given smoker,pollution")
q=cancer_infer.query(variables=['Cancer'],evidence={'Smoker':1,'Polluti
on':1})
print(q)
```

OUTPUT:

```
FINDING ELIMINATION ORDER: :          0%|          | 0/3 [00:00<?, ?IT/S] 0%|   |
   0/3 [00:00<?, ?IT/S]
ELIMINATING: DYSPNOEA:        0%|          | 0/3 [00:00<?, ?IT/S]
ELIMINATING: POLLUTION:        0%|          | 0/3 [00:00<?, ?IT/S]
ELIMINATING: XRAY: 100%|██████████| 3/3 [00:00<00:00, 359.52IT/S]


   0%|          | 0/2 [00:00<?, ?IT/S]
FINDING ELIMINATION ORDER: :          0%|          | 0/2 [00:00<?, ?IT/S] 0%|   |


   0/2 [00:00<?, ?IT/S]
```

```
ELIMINATING: DYSPNOEA:           0%|                    | 0/2 [00:00<?, ?IT/S]

ELIMINATING: XRAY: 100%|██████████| 2/2 [00:00<00:00, 333.49IT/S]A
  INFERENCING WITH BAYESIAN NETWORK

  PROBABILITY OF CANCER GIVEN SMOKER
+-----------+---------------+
| CANCER    |   PHI(CANCER) |
+===========+===============+
| CANCER(0) |        0.0029 |
+-----------+---------------+
| CANCER(1) |        0.9971 |
+-----------+---------------+
PROBABILITY OF CANCER GIVEN SMOKER,POLLUTION
+-----------+---------------+
| CANCER    |   PHI(CANCER) |
+===========+===============+
| CANCER(0) |        0.0200 |
+-----------+---------------+
| CANCER(1) |        0.9800 |
+-----------+---------------+
```