1. EXPERIMENT NO: 7

2. TITLE: **BAYESIAN NETWORK**

3. LEARNING OBJECTIVES:

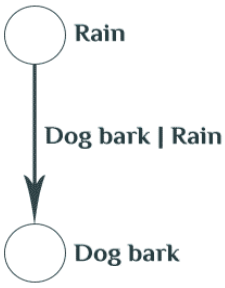
- Make use of Data sets in implementing the machine learning algorithms.
- Implement ML concepts and algorithms in Python

4. AIM:

- Write a program to construct a Bayesian network considering medical data. Use this model to demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use Java/Python ML library classes/API.

5. THEORY:

- Bayesian networks are very convenient for representing similar probabilistic relationships between multiple events.
- Bayesian networks as graphs - People usually represent Bayesian networks as directed graphs in which each node is a hypothesis or a random process. In other words, something that takes at least 2 possible values you can assign probabilities to. For example, there can be a node that represents the state of the dog (barking or not barking at the window), the weather (raining or not raining), etc.
- The arrows between nodes represent the conditional probabilities between them — how information about the state of one node changes the probability distribution of another node it's connected to.



6. PROCEDURE / PROGRAMME :

Program for the Illustration of Baysian Belief networks using 5 nodes using Lung cancer data. (The Conditional probabilities are given)

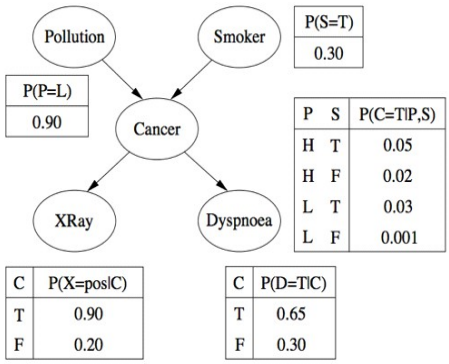
```
from pgmpy.models import BayesianModel
from pgmpy.factors.discrete import TabularCPD
from pgmpy.inference import VariableElimination

#Define a Structure with nodes and edge
cancer_model = BayesianModel([('Pollution', 'Cancer'),
                              ('Smoker', 'Cancer'),
                              ('Cancer', 'Xray'),
                              ('Cancer', 'Dyspnoea')])

print('Baysian network nodes are:')
print('\t',cancer_model.nodes())
print('Baysian network edges are:')
print('\t',cancer_model.edges())

#Creation of Conditional Probability Table

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])

# Associating the parameters with the model structure.
cancer_model.add_cpds(cpd_poll, cpd_smoke, cpd_cancer, cpd_xray, cpd_dysp)
print('Model generated by adding conditional probability disttributions(cpds)')

# Checking if the cpds are valid for the model.
print('Checking for Correctness of model : ', end='' )
print(cancer_model.check_model())

'''print('All local idependencies are as follows')
cancer_model.get_independencies()
'''
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'))

##Inferencing with Bayesian Network

# Computing the probability of Cancer given smoke.
cancer_infer = VariableElimination(cancer_model)

print('\nInferencing with Bayesian Network');

print('\nProbability of Cancer given Smoker')
q = cancer_infer.query(variables=['Cancer'], evidence={'Smoker': 1})
print(q['Cancer'])

print('\nProbability of Cancer given Smoker,Pollution')
q = cancer_infer.query(variables=['Cancer'], evidence={'Smoker': 1,'Pollution': 1})
print(q['Cancer'])
```

Program as per the Syllubus

```
import numpy as np
import pandas as pd
import csv
from pgmpy.estimators import MaximumLikelihoodEstimator
from pgmpy.models import BayesianModel
from pgmpy.inference import VariableElimination

#Read the attributes
lines = list(csv.reader(open('data7_names.csv', 'r')));
attributes = lines[0]
#Read Cleveland Heart dicease data
heartDisease = pd.read_csv('data7_heart.csv', names = attributes)
heartDisease = heartDisease.replace('?', np.nan)
```

```
# Display the data
#print('Few examples from the dataset are given below')
#print(heartDisease.head())
#print('\nAttributes and datatypes')
#print(heartDisease.dtypes)

# Model Baysian Network
model = BayesianModel([('age', 'trestbps'), ('age', 'fbs'), ('sex', 'trestbps'), ('sex', 'trestbps'),
                       ('exang', 'trestbps'),('trestbps','heartdisease'),('fbs','heartdisease'),
                      ('heartdisease','restecg'),('heartdisease','thalach'),('heartdisease','chol')])

# Learning CPDs using Maximum Likelihood Estimators
print('\nLearning CPDs using Maximum Likelihood Estimators...');
model.fit(heartDisease, estimator=MaximumLikelihoodEstimator)

# Inferencing with Bayesian Network
print('\nInferencing with Bayesian Network:')
HeartDisease_infer = VariableElimination(model)

# Computing the probability of bronc given smoke.
print('\n1.Probability of HeartDisease given Age=20')
q = HeartDisease_infer.query(variables=['heartdisease'], evidence={'age': 28})
print(q['heartdisease'])

print('\n2. Probability of HeartDisease given chol (Cholestoral) =100')
q = HeartDisease_infer.query(variables=['heartdisease'], evidence={'chol': 100})
print(q['heartdisease'])
```

7. RESULTS & CONCLUSIONS:

Dataset

data7_names.csv (14 attributes)

```
age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,oldpeak,
slope,ca,thal,heartdisease
```

data7_heart.csv (5 instances out of 303)

```
63.0,1.0,1.0,145.0,233.0,1.0,2.0,150.0,0.0,2.3,3.0,0.0,6.0,0
67.0,1.0,4.0,160.0,286.0,0.0,2.0,108.0,1.0,1.5,2.0,3.0,3.0,2
67.0,1.0,4.0,120.0,229.0,0.0,2.0,129.0,1.0,2.6,2.0,2.0,7.0,1
37.0,1.0,3.0,130.0,250.0,0.0,0.0,187.0,0.0,3.5,3.0,0.0,3.0,0
41.0,0.0,2.0,130.0,204.0,0.0,2.0,172.0,0.0,1.4,1.0,0.0,3.0,0
```

Output

```
Learing CPDs using Maximum Likelihood Estimators...
Inferencing with Bayesian Network:
1.Probability of HeartDisease given Age=20
```

heartdisease	phi(heartdisease)
heartdisease_0	0.6791
heartdisease_1	0.1212
heartdisease_2	0.0810
heartdisease_3	0.0939
heartdisease_4	0.0247

2. Probability of HeartDisease given chol (Cholestoral) =100

heartdisease	phi(heartdisease)
heartdisease_0	0.5400
heartdisease_1	0.1533
heartdisease_2	0.1303
heartdisease_3	0.1259
heartdisease_4	0.0506

8. LEARNING OUTCOMES :

- The student will be able to apply baysian network for the medical data and demonstrate the diagnosis of heart patients using standard Heart Disease Data Set.

9. APPLICATION AREAS:

- Applicable in prediction and classification
- Gene Regulatory Networks
- Medicine
- Biomonitoring
- Document Classification
- Information Retrieval
- Semantic Search

10. REMARKS: