```
Name: Melwyn D Souza
Student Number: R00209495
Date: 4/12/2021
Module: COMP9016 - Knowledge Representation
Lecturer: Dr Ruairi O'Reilly
Course: MSc in Artificial Intelligence
Assignment - 2
```

Question 1.1

Axioms

grandchild(z,x) <==> for_all x,y,z (child(z,y) ∧ child(y,x))

greatgrandparent(w,z) <==> for_all w,x,y,z ((parent(w,x) ∧ parent(x,y) ∧ parent(y,z))

brother(x,y) <==> for_all a,b,x,y (parent(a,x) ∧ parent(b,x) ∧ parent(a,y) ∧ parent(b,y) ∧ male(x) ∧ x!=y)

sister(x,y) <==> for_all a,b,x,y (parent(a,x) ∧ parent(b,x) ∧ parent(a,y) ∧ parent(b,y) ∧ female(x) ∧ x!=y)

son(x,y) <==> for_all x,y (child(x,y) ∧ male(x))

daughter(x,y) <==> for_all x,y (child(x,y) ∧ female(x))

firstcousin(x,y) <==> for_all a,b,p,q,x,y (parent(a,x) ∧ parent(p,a) ∧ parent(b,y) ∧ parent(q,a) ∧ parent(p,b) ∧ parent(q,b) ∧ (a!=b) ∧ (x!=y) ∧ (p!=q))

aunt(x,y) <==> for_all a,b,x,y (parent(a,y) ∧ brother(b,a) ∧ married(b,x)) ∨ (parent(a,y) ∧ sister(x,a))

uncle(x,y) <==> for_all a,b,x,y (parent(a,y) ∧ sister(b,a) ∧ married(b,x)) ∨ (parent(a,y) ∧ brother(x,a))

brotherinlaw(x,y) <==> for_all a,x,y (male(x) ∧ married(x,a) ∧ (brother(a,y) ∨ sister(a,y)))

sisterinlaw(x,y) <==> for_all a,x,y (female(x) ∧ married(x,a) ∧ (brother(a,y) ∨ sister(a,y)))d

ancestor(x,y) <==> for_all x,y,z (parent(x,z) ∧ ancestor(z,y)) ∨ parent(x,y)

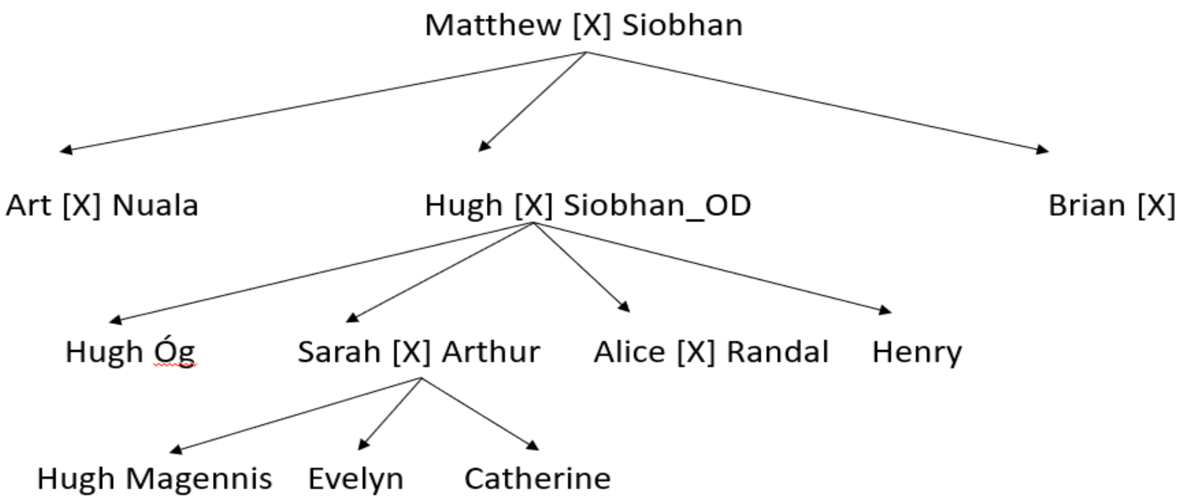

Fig. 1

From the figure 1, which is the family tree of Matthew[X]Siobhan

Hugh's Grandchildren are Hugh Magennis, Evelyn, Catherine

Hugh_OG's brothers-in-law are Arthur and Randall

Catherine's great-grandparents are Matthew [X] Siobhan

Evelyn's Ancestors are Sarah[X]Arthur, Hugh[X]Siobhan_OD, Matthew[X]Siobhan

Question 1.2

The Bayesian network that is implemented for question 1.2 is as shown below

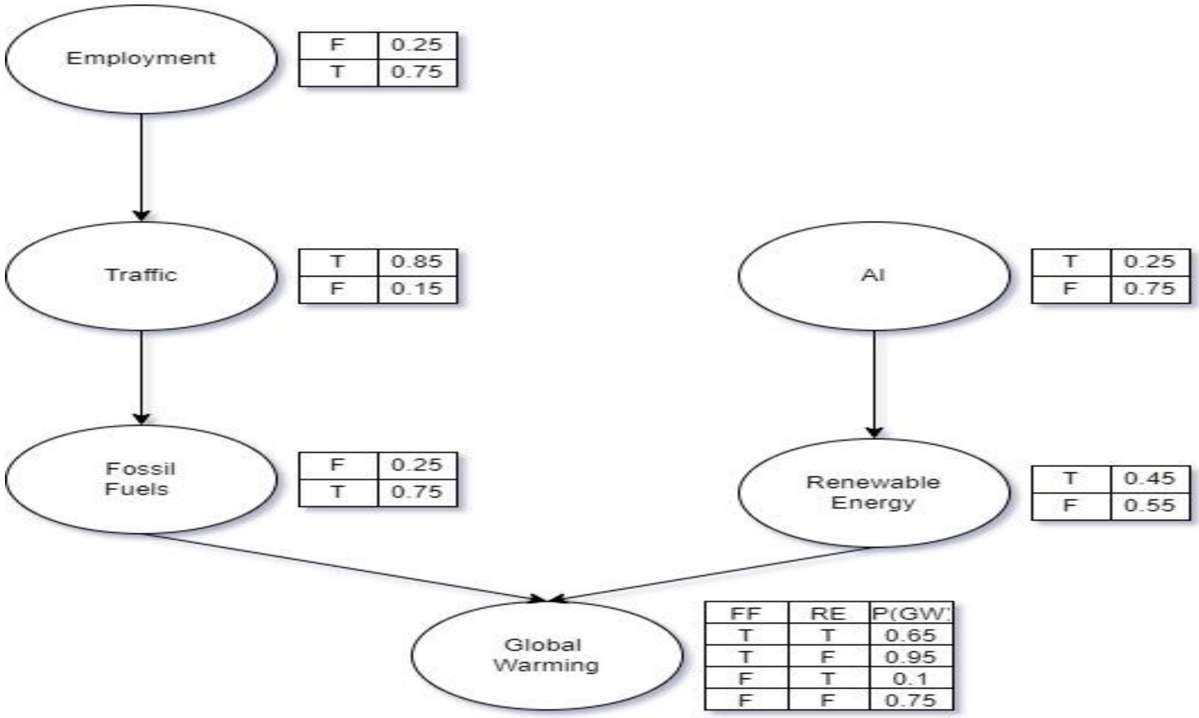

Fig. 2

The idea behind constructing the network above is as follows

1. Global warming is directly proportional to fossil fuels and inversely proportional to renewable energy, which means, when usage of fossil fuels increases and renewable energy decreases, global warming will increase, hence the probability of global warming will be close to 1
2. Burning of fossil fuels will have high impact on global warming
3. The traffic increases with more people hitting the road and most of the traffic on the road is from the people who are employed, they go to work and return home using automobiles, increase in traffic increases the consumption of fossil fuels
4. The new discovery in the field of AI will help extract renewable energy more efficiently, which helps in reducing global warming to a good extant

The Bayesian networks are also called as bayes net or decision networks. These are graphical representation of a probabilistic model, the bayes net is a directed acrylic graph[2] with some latent, unknown or numerical parameters attached to the nodes of the network, the edges of the network represent the conditional dependencies of the nodes, so when there is no edge connecting a node to another, it means that the node is conditionally independent, the bayes net implicitly encodes full joint distribution tables, the drawback in joint distribution table (more the variables, complex /unworkable the distribution) is addressed by the bayes net by using the conditional probabilities

Bayes net represent the interaction between the nodes and can calculate the probabilities of an event given some information about other dependent nodes in the network

In our Bayesian network above, we can see that all the nodes are dependent on each other because they all contribute to the final node i.e global warming, for example, even if it seems like the renewable energy is not dependent on the fossil fuels, when fossil fuels True, and renewable energy is True, the global warming is likely to be high, so when global warming is high, fossil fuels is likely to be true and renewable energy is likely to be False, so these nodes are all inter dependent on the occurrence of global warming

In the network above, the more people are employed, more the traffic on the road, the exceeding traffic then increases the consumption of fossil fuels which in turn increases global warming, on the other hand the AI field can find new ways to extract renewable energy efficiently which can reduce global warming, the probability tables are also attached beside each node in the network above, these are the conditional probabilities linked to global warming

The Bayesian nodes are implemented and queried, the output with the config is as shown below

```
Querying bayes nodes with Fossilfuel-True, Traffic-True, Renewable Energy-
False, AI-True, Employed-True
GLobal warming probabilty: 0.95

Querying bayes nodes with Fossilfuel-True, Traffic-True, Renewable Energy-False, AI-False, Employed-
True
Global warming probability: 0.95
```

Fig. 3

Question 1.3

1.3.1 DATA

The data chosen for this question is mammographic data, the information about the data is as shown below

Mammographic Mass Data Set

Download: Data Folder, Data Set Description

Abstract: Discrimination of benign and malignant mammographic masses based on BI-RADS attributes and the patient's age.

Data Set Characteristics:	Multivariate	**Number of Instances:**	961	**Area:**	Life
Attribute Characteristics:	Integer	**Number of Attributes:**	6	**Date Donated**	2007-10-29
Associated Tasks:	Classification	**Missing Values?**	Yes	**Number of Web Hits:**	197857

Fig. 4

There are 6 attributes in the dataset as shown below

6 Attributes in total (1 goal field, 1 non-predictive, 4 predictive attributes)

1. BI-RADS assessment: 1 to 5 (ordinal, non-predictive!)
2. Age: patient's age in years (integer)
3. Shape: mass shape: round=1 oval=2 lobular=3 irregular=4 (nominal)
4. Margin: mass margin: circumscribed=1 microlobulated=2 obscured=3 ill-defined=4 spiculated=5 (nominal)
5. Density: mass density high=1 iso=2 low=3 fat-containing=4 (ordinal)
6. Severity: benign=0 or malignant=1 (binominal, goal field!)

Our target feature is the attribute 6, which classifies the tumour to be benign (class 0) or malignant(class 1)

Bayes theorem is based on the probabilities as shown in the picture extracted from [2] below

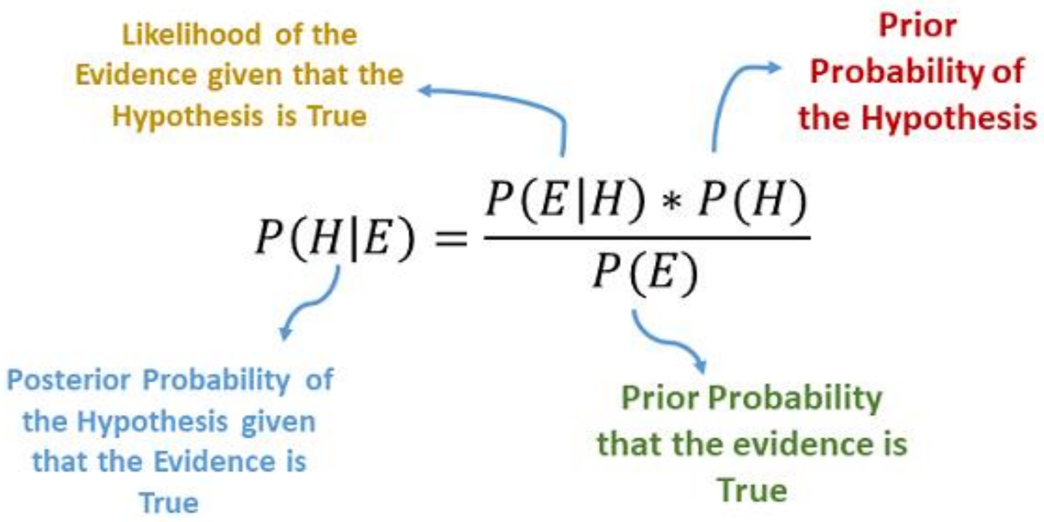

Fig. 5

Prior probability:

The frequency of occurrence of the classes/hypothesis (figure 5), in our case, the prior probability is the frequency of occurrence of class benign and class malignant. Find the frequency of benign and malignant classes in the dataset and divide the frequencies by the total number of instances, this gives us the prior probabilities P(benign) and P(malignant).

The output from the script calculating prior probs on mammographic data is

```
Prior Probabilities
Benign: 0.5150602409638554, Malignant: 0.48493975903614456
```

Probability of Evidence:

The probability of evidence is in the denominator as shown in figure 5, it is the probability of the evidence being true, it can be calculated by finding the frequency of the evidence and dividing it with the total number of instances from the dataset

P(E)=frequency of E for given column/length of the column

The script iterates through each column and finds the frequency of each unique variable in the column, the output of the script is as shown below, each dictionary in the list consists of the probabilities of evidence from each attribute, for example, the first dictionary is related to column "BI Rads"

```
Probability of evidence
Feature:  BI_Rads
Feature:  Age
Feature:  Shape
Feature:  Margin
Feature:  Density
[{5: 0.41566265060240964, 4: 0.5286144578313253, 3: 0.03162650602409638, 2: 
0.009036144578313253, 6: 0.010542168674698794}, {67: 0.0391566265060241, 58: 
0.02108433734939759, 28: 0.007530120481927711, 36: 0.01355421686746988, 60: 
0.022590361445783132, 54: 0.02710843373493976}, {3: 0.08734939759036145, 4: 
0.46686746987951805, 1: 0.23343373493975902, 2: 0.21234939759036145}, {5: 0.15963855421686746, 
1: 0.38403614457831325, 4: 0.3042168674698795, 3: 0.1280120481927711, 2: 
0.024096385542168676}, {3: 0.9156626506024096, 2: 0.061746987951807226, 1: 
0.012048192771084338, 4: 0.010542168674698794}]
```

Probability of Likelihood of Evidence:

Likelihood of the evidence given the class/hypothesis is true, in my case, I have separated the benign instances and malignant instances, the probability of likelihood of any evidence given that benign is true is the frequency of occurrence of evidence in any given column divided by the length of the column in benign class, likewise for malignant class

P(E|H) = Number of times E occurs where H occurs too / number of times H occurs

The output from the script calculating the probabilities of likelihood of evidence given the class is true is as shown below

```
Probability of likelihood (Benign)
Feature:  BI_Rads
Feature:  Age
Feature:  Shape
Feature:  Margin
Feature:  Density
[{4: 0.8304093567251462, 3: 0.05263157894736842, 5: 0.09064327485380116, 2: 
0.017543859649122806, 39: 0.02046783625730994, 81: 0.0029239766081871734}, {1: 
0.3742690058479532, 3: 0.08771929824561403, 2: 0.3333333333333333, 4: 0.2046783625730994}, {1: 
0.6549707602339181, 4: 0.1871345029239766, 5: 0.04678362573099415, 3: 0.0935672514619883, 2: 
0.017543859649122806}, {3: 0.8801169590643275, 2: 0.0935672514619883, 1: 0.014619883040935672, 
4: 0.011695906432748537}]

Probability of likelilhood (Malignant)
Feature:  BI_Rads
Feature:  Age
Feature:  Shape
Feature:  Margin
```

```
Feature:  Density
[{5: 0.7608695652173914, 3: 0.009316770186335404, 4: 0.2080745341614907, 55:
0.003105590062111801, 0: 0.003105590062111801, 6: 0.015527950310559006}, {67:
0.06521739130434782, 58: 0.024844720496894408, 57: 0.021739130434782608, 76:
0.027950310559006212}, {3: 0.08695652173913043, 4: 0.7453416149068323, 1: 0.08385093167701864,
2: 0.08385093167701864}, {5: 0.2795031055900621, 2: 0.031055900621118012}, {3:
0.953416149068323, 2: 0.027950310559006212, 4: 0.009316770186335404}]
```

NB – only few probabilities are copied for visualization of the output, the actual list is very large to be copied to the report

1.3.2 Naïve Bayes Model

In this section, we are implementing a Naïve Bayes algorithm which is used for classification in machine learning. An event is the assignment of values to the features, the probability function takes an event and returns the likelihood of that event, the probability of an event is the frequency of occurrence of that event in the given environment/dataset (how often did the event occur divided by how often it could have occurred)

There are different types of probability functions such as

- Unconditional probability – the probability of an event is not impacted by future or past events
- Joint probability – two or more events are interrelated or occurring together
- Posterior or conditional probability – Frequency of an event where other event or events are known to have occurred

Full joint distribution table is the joint distributions over all features in the domain, the drawback here is, as the number of features increase, the size of the full joint table grows exponentially, it becomes mathematical and space inefficient since it requires more calculations and storage, these issues are solved in the bayes theorem which uses conditional probabilities

- Product Rule: P(a,b) = P(a|b)P(b) = P(b|a)P(a)
- Bayes rule = P(a|b) = P(b|a)P(a) / P(b)

The Naïve Bayes model follows the process as shown below

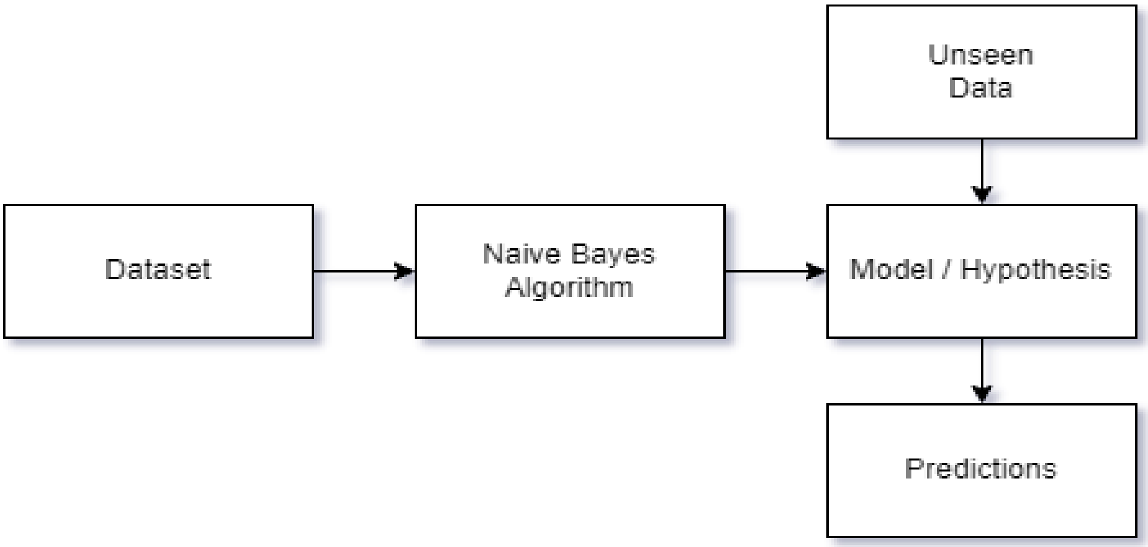

Fig. 6

The model needs to be trained with the dataset, this data has to be prepared well to select the best out of the available data without losing any important information, The data then is split as Train and Test data, this step is followed to validate our model, Once the model is trained with the train (train_X,train_Y) instances, the model is then tested with test (test_x) instances and predictions (pred_y) are obatained, this prediction is them compared with the true values (test_y) which gives the accuracy of our model

The naïve bayes model assumes that each any feature is independent of the other features, so the probability of one event won't be affected by the knowledge of other events in the dataset, hence the word NAÏVE bayes model

The steps followed by the naïve bayes model are

1. All the conditional probabilities are calculated during the training of the model
2. The prior probabilities of the classes are calculated
3. When a feature test instance is received, it will use the calculated probabilities to find the probabilities of the test instance for all available classes, whichever is greater will be our predicted class

The accuracy of my model is calculated, the output is attached below

```
Number of train instances: 664
Number of test instances: 166
The Naive Bayes model accuracy is 75.30 %
```

References:

[1] http://archive.ics.uci.edu/ml/datasets/mammographic+mass

[2] https://laptrinhx.com/naive-bayes-algorithm-4229686008/

[3] Artificial Intelligence – A modern approach, Stuart Russell and Peter Norvig